

COMPLETE MANAGEMENT LIBRARY

VOLUME IV

Joseph Zeidner

*Chief, Support Systems Research Laboratory
U. S. Army Personnel Research Office*

SCIENTIFIC SELECTION OF PERSONNEL

National Foremen's Institute
Bureau of Business Practice
National Sales Development Institute
WATERFORD, CONNECTICUT

COMPLETE MANAGEMENT LIBRARY
VOLUME IV

© MCMLXIII
Prentice-Hall, Inc.
Waterford, Connecticut

Reproduction in whole or part is expressly prohibited
unless with the written permission of the publisher.

LIBRARY OF CONGRESS CATALOG CARD NUMBER: 63-16641

PRINTED IN THE UNITED STATES OF AMERICA

Third Printing, December, 1969

PREFACE

This book is intended to aid management in deciding when to initiate a test selection program, in determining if it is being run effectively, and in calculating the gains or losses resulting from its use. Basic principles are emphasized, along with the necessary technical content required to understand the use of tests in solving practical personnel problems.

The book is not intended to serve as a do-it-yourself guide. Test construction and validation, interpreting test results, and establishing procedures from using test scores fall within the province of the expert. However, a truly effective program cannot exist without an understanding by management of at least the main principles and procedures. Thereby, management will be able, independently, to recognize needs and to understand and evaluate proposals made by the experts.

Specific chapters discuss the various test types; the conditions when tests will work best; the steps in test development; the building of a yardstick or criterion and test evaluation; and translation of a test score into a personnel decision.

A later book in this series describes the practical steps needed to set up and run a testing program and contains critical reviews of the major types of tests currently used in industrial selection. The theoretical basis of test development, practical features, advantages and disadvantages of many of the leading tests are described. Additional sources of test information, references and test publishers are also listed.

In writing this book, I have been greatly helped by my professional association over the last 12 years with the U. S. Army Personnel Research Office. I gratefully acknowledge the assistance of my colleague Robert S. Andrews for his critical review of the entire manuscript and for his suggestions which led to a number of improvements. Special thanks are due to Joseph D. Cooper who gave helpful suggestions for revision, and to the various test authors and publishers who gave permission for the reproduction of test materials.

Finally, my wife, Dorothy, provided the necessary encouragement and support during the months in which this book was being written.

Joseph Zeidner

Washington, D. C.
March, 1963

TABLE OF CONTENTS

Chapter One	1	Psychological Tests and Their Role in Industry <i>Effective Utilization of Manpower</i> <i>Functions and Types of Tests</i> <i>Progress in Measuring Human Charac-</i> <i>teristics</i> <i>Summary</i>
Chapter Two	19	When to Use Tests in Selection <i>How Extensively Are Tests Used?</i> <i>Are Tests Being Used Correctly?</i> <i>Where Tests Have Been Found to Work</i> <i>Certain Shortcomings of Tests</i> <i>The Bonus in Testing</i> <i>When Tests Will Work Best</i> <i>Summary</i>
Chapter Three	44	How Tests Are Developed <i>Steps in Test Construction</i> <i>Preliminary Tryout</i> <i>Summary</i>
Chapter Four	59	How Tests Are Evaluated <i>The Criterion</i> <i>Reliability</i> <i>Validity: The Value of Tests</i> <i>Summary</i>

Chapter Five	88	Putting Tests to Work
		<i>Practical Problems in Setting Test Standards</i>
		<i>Methods of Meeting Job Quotas</i>
		<i>Setting a Qualifying Score</i>
		<i>Methods of Comparing Scores</i>
		<i>Interpreting Test Results</i>
		<i>Interpreting Individual Test Scores</i>
		<i>Interpreting a Selection Program's Value</i>
		<i>Summary</i>
References	116	
Index	119	

CHAPTER ONE

Psychological Tests and Their Role in Industry

Everyone will agree that a firm must match the right man to the right job. When this is done well, productivity rises, training costs fall, and waste, accidents, absenteeism, and turnovers decrease. The only *certain* way of telling whether a man is good for a job is to try him out. Obviously, this is too costly and time-consuming. Some other pre-selection device is needed.

Most firms use one or more pre-screening devices. These may include: a review of past employment, the application and pre-employment and department interviewing, reference checks and physical examinations. Many firms believe the limited gains of more elaborate screening procedures do not justify their cost.

On the other hand, management has become ever more interested in the total adjustment of the employee to his work and in his long-term value to the firm. Modern se-

lection procedures focus on the applicant as a whole being, not merely as a slice of a person looking for a specific job. The trend is to regard the applicant in terms of his potential ability to do several different jobs. The assessment takes in all his present capabilities, and what he can be trained to do. While the company expects to profit directly, it also expects the employee to take a bonus in the form of greater job satisfaction.

Psychological tests are becoming increasingly prevalent in such modern selection procedures. This chapter will consider the role of selection tests in effective utilization of manpower, the functions and types of tests, and progress in measuring human characteristics.

Effective Utilization of Manpower

The job description states what a man is expected to do and the degree of proficiency needed. Individuals, also, can be described in terms of skills and abilities. Effective utilization of manpower, then, hinges on matching the job and the man. Of course, there is a hitch in this. That is our ability to measure a person's appropriate skills and abilities, assuming we have evaluated a job properly. In turn, this depends on measuring the right characteristics, and determining how reliably or accurately we measure them.

The measure of ability may be useless if it is irrelevant to job success. Suppose, for example, that we want to measure a man's ability as a salesman. We might accurately describe his verbal, spatial, and numerical reasoning abilities. But none of these may be related to successful sales work. A first problem in measurement, then, is to identify and describe the critical job skills.

A second problem is to determine ways of measuring the desired skills. Even if we were to select the right job characteristics, this would be of little value if their measures were expressed crudely or inaccurately. This could be like using a yardstick to measure the diameter of a hair. Now, suppose we want to measure intelligence, sociability, and general activity of sales applicants, and do so by relying on ratings by several interviewers. Such ratings would most likely have little value in predicting sales performance. The attributes being rated are important to job success, but they can't be reliably and precisely described by most raters.

Typically, the rater makes an evaluation which he expresses in adjective terms (satisfactory, excellent, superior), or according to some scale of performance that he has experienced. This, however, doesn't measure with sufficient accuracy to permit ranking people on a consistent basis. This, of course, indicates the need for a quantitative scale in comparing individual differences.

ROLE OF SELECTION TESTS

All of us know men who consider themselves good judges of human abilities. If we judge them, by their own successful evaluations and predictions, we would rate them very high on any kind of scale. Why, then, use any other means of evaluation? Why use a test?

Admittedly, *if* there were enough good judges, and *if* they had time to make judgments, and *if* their judgments could be passed along without misunderstanding or distortion, then there would be no need for tests.

But all evidence indicates that not many judges can evaluate people consistently. Every evaluator has his own personal standards based on his own experiences. Consciously or unconsciously, he tends to evaluate according

to his own standards. All evidence shows inaccuracies and disparities in such judgments. Further, the problem of using judges becomes more confounded when we expect to predict future performance through limited observation of the applicant. Thus, judgments by individuals are neither practical nor economic assessments of future performance.

Any evaluation's most important role is ability to predict performance. No test is a perfect predictor, but a good test does provide objective scores of a man's standing in comparison with others, on the abilities measured. When a test is itself evaluated, you can know its own limits of accuracy. These tell how much you can rely on its predictions. A well-designed test selection program should identify differences among individuals on relevant abilities, measure them accurately, and use them in behalf of both the individual and the firm in making personnel decisions.

Functions and Types of Tests

PURPOSE OF TESTING

The purpose of every psychological test is to measure differences among individuals. Such measures can serve a number of functions: prediction, diagnosis, or research. Since this book's major concern is with the use of tests for selection purposes, the focus is directed at the predictive function of tests.

A selection test's primary purpose is to predict actual or potential job proficiency. The fact that two individuals perform differently on tests is of little real value, unless such differences can be shown to relate to some future activity of interest to the firm. A test's predictive value

depends on how well it can serve as an indicator of a significant area of behavior. It is of no great interest to know that an applicant can get 30 of 35 answers correct on an automotive information test, unless a close relationship can be shown between the applicant's score on the automotive information test and his performance on the job as a mechanic. If there is such a close correspondence, that is, if high scorers on the test turn out to be good mechanics and low scorers poor mechanics, the test is serving its purpose.

A test need not necessarily resemble closely the job it is designed to predict. What's necessary is that a relationship be demonstrated between the test and job performance. The apparent similarity between behavior sampled in the test and the predicted job behavior may vary widely. For an applicant for a stenographic position, a small dictation task given in a standard way may serve as an effective test. (This is actually a type of job sample test.) In selecting a student pilot, a pattern analysis test may be used effectively. Here, there is little seeming correspondence to piloting an aircraft. Yet such tests are known to be effective, because they tap a spatial visualization ability which repeatedly has been found to be related to flying skill.

On the other hand, a test which seems ideally suited for a job may be of no value. One firm used an eye-hand coordination test to select assemblers of electronic equipment. The movements in the test were very close to those required on the job. Yet a comparison of test results with later job success showed no relationship.

WHAT IS A PSYCHOLOGICAL TEST?

A psychological test may be defined as an objective and standardized measure for comparing the behavior of in-

dividuals. "Objectivity" of tests means that the administration, scoring, and interpretation of scores are objective insofar as they are not influenced by the actions of the test administrator or the judgments of the test scorer. The goal here is that an individual will achieve the same score on a test, no matter who gives him the test or who scores it. This is not always attained, because complete objectivity depends on perfect control over all factors. In practice, this can't be fully accomplished. Every psychological test, however, is designed to obtain objectivity and generally attains it to a significant degree, especially in comparison with other methods used to evaluate personal attributes.

Standardization refers to the uniformity of testing procedures and conditions. If the scores obtained by different persons are to be compared to one another, and are to be interpreted in the same way, testing conditions must be the same for all. The test developer spends a great deal of effort in specifying the controlled conditions needed for standardized test administration.

Another aspect of standardization deals with obtaining *norms*. Norms provide the distribution of test scores which define "normal" performance on the test. Without such norms it is impossible to interpret test performance—what standard or level represents good, average, or poor test performance. Norms are obtained by administering the test to a large, representative sample of people similar to the kind of people for whom the test was designed.

The broad definition of a test provided here allows us to consider as tests such varied measures as: reaction times obtained by use of apparatus; stress and anxiety scores based upon systematic, observational procedures; personality descriptions obtained by means of a biographical

information questionnaire; and work output indices derived from standard productivity records.

Despite all the apparent differences in these measures, all consist of samples of an individual's behavior during a standard, objective performance. The practical value of any of these measures must be demonstrated in an empirical study, relating test performance to some other meaningful situation.

The definition of a test provided here eliminates some important procedures from consideration as tests. For example, most interviews cannot be considered as tests; each individual cannot be treated exactly the same way. There is usually an important individual interaction between the applicant and the interviewer.

It may be possible to develop a highly structured, standardized interview consisting of a uniform set of questions asked of all persons. Such an interview procedure might possibly be considered as a test. Other informal assessment procedures, such as observations, talks, or even a set of oral questions, not phrased in exactly the same manner or sequence to all applicants, rarely fall within the definition of a test.

TYPES OF PSYCHOLOGICAL TESTS

Broad Classification. There are many ways of classifying tests according to the behavior they measure. One possible method is to divide tests into those that describe what people *can* do and what they *will* do. Tests in the first group may be called *ability measures*; tests in the second group, *personality measures*.

Ability measures can be subdivided into measures of potential or capacity, called aptitude tests, and into measures of present knowledge, called achievement tests. The difference between aptitude and achievement tests is not

very clear-cut; it is really based upon the use we are going to make of the test. For example, an achievement test can be used as an aptitude test when it is used to predict future performance. Thus, we may give an individual an automotive information test before he actually has any automotive training, in order to determine his potential for this kind of work.

Personality tests, providing descriptions of what people will do, constitute a rather broad and loose category of measures. Objective personality tests focus on what a person does or how he responds to a given situation; they do not directly attempt to measure feelings or aspirations, since these are not observable. Such attributes are used as concepts in the development of test questions. The answers to these test questions can be objectively scored. The purpose of personality tests is to predict typical behavior. Interest is not directed at defining peak performance, but at describing most likely behavior.

Specific Test Types. Tests are more commonly classified into more refined groupings, according to the aspect of behavior they measure.

Those of general ability are among the oldest and most widely known. They measure general intellectual level or learning ability. Intelligence (or IQ) tests are examples. (These tests are now more frequently described as general classification tests.) Measures of mental ability have been found to be very useful in applicant screening. They provide a quick means of determining whether a person has sufficient mental ability (or perhaps too much) to perform well in a job.

Differential aptitude tests measure specific abilities (such as mechanical knowledge, artistic and sensory acuity). Such tests have great value for differential job placement of employees, and in job counseling. For such uses,

they are replacing general ability tests. Aptitude tests do not indicate the present skill of an applicant, but rather his potential for a given job. A major problem in the use of aptitude tests is determining the right ones to use. The name of a test can't be used to indicate its value for a job.

Achievement tests measure how much a man knows about a particular job, or the level of skill acquired. In industry, they are called trade or proficiency tests. They are generally used for rapid screening or as part of a larger selection battery. In using such tests for screening, it is assumed that the person tested has had training or experience in the specific area covered by the test questions. Some proficiency tests, such as typing, can be considered performance tests. The applicant actually shows his skill by a work sample. Many trade tests measure proficiency indirectly, by asking for specific information known only by people in the trade.

Personality tests are generally divided into a number of special types. A variety of types are used in selection programs.

Interest tests measure preferences for certain kinds of activities. Knowing the interest of an applicant or an employee may help in predicting job satisfaction. Such tests are used because it is generally impossible to adequately determine a person's interests by directly questioning him. Interest tests should never be used alone in selection, since interest in a job area does not necessarily mean competency in it.

Attitude tests measure how favorably an individual feels toward institutions, individuals, groups, and to a wide range of objects and concepts. Attitude scales are sometimes used in industry to measure the employee's attitude toward the job, morale, and effectiveness of a training program.

Temperament measures describe such individual attributes as energy level, sociability, and aggressiveness. Such tests generally ask a person to indicate which actions and feelings are characteristic of him.

Adjustment tests describe a person's behavioral pattern, to determine if he is reasonably comfortable with himself and with his social setting. Critics of adjustment and temperament tests do not believe that scores on each attribute represent distinct elements which are mutually exclusive. The major value of such tests is that they rank persons in a group in respect to specific behavioral areas. Such rankings may be valuable in prediction, or as a basis of further psychological investigation.

Character tests measure attributes to which society has assigned values, such as honesty, cooperativeness, and thriftiness. People taking such tests know what constitutes desirable and undesirable character traits. Thus self-assessment, even if made in good faith, may not correspond to actual behavior in all instances. Indeed, there may be no general factors of character excellence; adequate descriptions of character may require extensive testing in a variety of situations.

Interest, temperament, adjustment, and character tests are sometimes quite effective in screening and selecting employees. In many cases, however, they have been demonstrated to be of no value. For all practical purposes, they must be used with extreme care, since an applicant knows the information obtained from the tests may be used against him. The critical problem in the use of such tests is how to obtain an honest evaluation from the individual, rather than the most favorable picture (assuming that the person can provide a detached, impartial observation of himself). Personality tests which have been