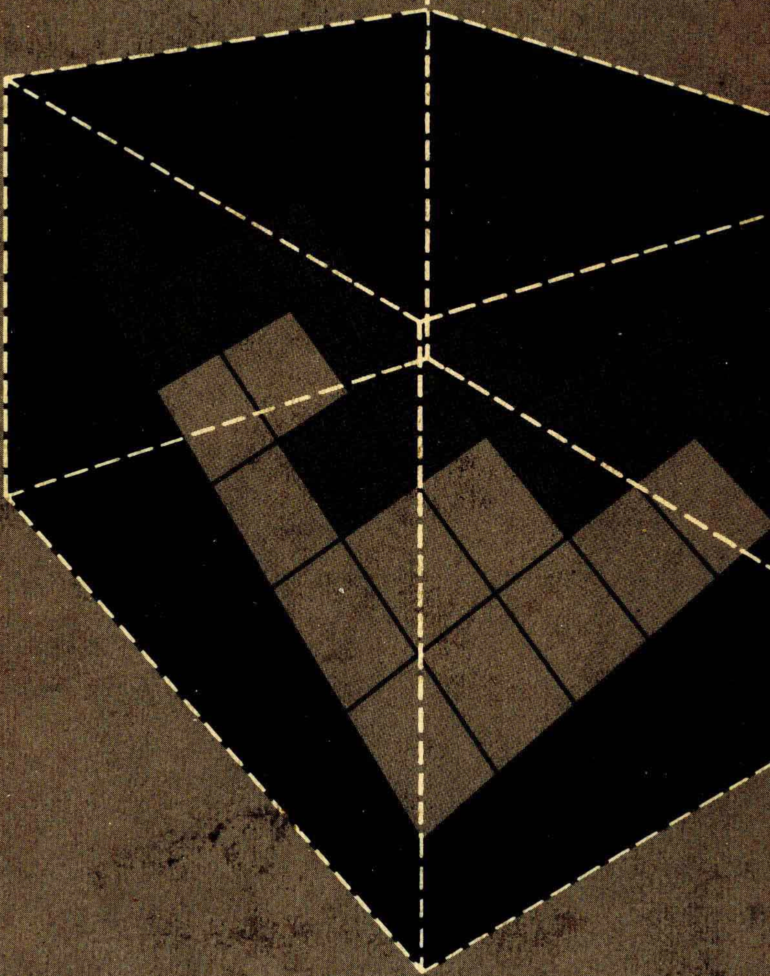
THEORY OF PSYCHOLOGICAL MEASUREMENT

EDWIN E. GHISELLI
Professor of Psychology
University of California, Berk

McGraw-Hill Series in Psychology



Theory of Psychological Measurement

Edwin E. Ghiselli
Professor of Psychology
University of California
Berkeley, California

New York
San Francisco
Toronto
London

McGRAW-HILL BOOK COMPANY

To

ROBERT C. TRYON

Friend and Teacher

Theory of Psychological Measurement

 Library of Congress Catalog Card Number 63–17952

3 4 5 6 7 8 9 - M P - 9 8 7 6

23140

McGraw-Hill Series in Psychology

Consulting Editors

Norman Garmezy

Harry F. Harlow

Lyle V. Jones

Harold W. Stevenson

Beach, Hebb, Morgan, and Nissen / *The Nepropsychology of Lashley*
Von Békésy / *Experiments in Hearing*
Berkowitz / *Aggression: A Social Psychological Analysis*
Berlyne / *Conflict, Arousal, and Curiosity*
Blum / *Psychoanalytic Theories of Personality*
Brown / *The Motivation of Behavior*
Brown and Ghiselli / *Scientific Method in Psychology*
Buckner and McGrath / *Vigilance: A Symposium*
Cofer / *Verbal Learning and Verbal Behavior*
Cofer and Musgrave / *Verbal Behavior and Learning: Problems and Processes*
Crafts, Schneirla, Robinson, and Gilbert / *Recent Experiments in Psychology*
Davitz / *The Communication of Emotional Meaning*
Deese / *The Pyschology of Learning*
Dollard and Miller / *Personality and Psychotherapy*
Ellis / *Handbook of Mental Deficiency*
Epstein / *Varieties of Perceptual Learning*
Ferguson / *Statistical Analysis in Psychology and Education*
Forgus / *Perception: The Basic Process in Cognitive Development*
Ghiselli / *Theory of Psychological Measurement*
Ghiselli and Brown / *Personnal and Industrial Psychology*
Gilmer / *Industrial Psychology*
Gray / *Psychology Applied to Human Affairs*

GUILFORD / *Fundamental Statistics in Psychology and Education*
GUILFORD / *The Nature of Intelligence*
GUILFORD / *Personality*
GUILFORD / *Psychometric Methods*
GUION / *Personnel Testing*
HAIRE / *Psychology in Management*
HIRSH / *The Measurement of Hearing*
HURLOCK / *Adolescent Development*
HURLOCK / *Child Development*
HURLOCK / *Development Psychology*
JACKSON AND MESSICK / *Problems in Human Assessment*
KARN AND GILMER / *Readings in Industrial and Business Psychology*
KRECH AND CRUTCHFIELD / *Individual in Society*
LAZARUS / *Adjustment and Personality*
LAZARUS / *Psychological Stress and the Coping Process*
LEWIN / *A Dynamic Theory of Personality*
MAHER / *Principles of Psychopathology*
MARX AND HILLIX / *Systems and Theories in Psychology*
MESSICK AND BRAYFIELD / *Decision and Choice: Contributions of Sidney Siegel*
MILLER / *Language and Communication*
MORGAN / *Physiological Psychology*
PAGE / *Abnormal Psychology*
RETHLINGSHAFER / *Motivation as Related to Personality*
ROBINSON AND ROBINSON / *The Mentally Retarded Child*
SCHERER AND WERTHEIMER / *A Psycholinguistic Experiment on Foreign Language Teaching*
SHAFFER AND LAZARUS / *Fundamental Concepts in Clinical Psychology*
SHAW AND WRIGHT / *Scales for the Measurement of Attitudes*
SIDOWSKI / *Experimental Methods and Instrumentation in Psychology*
SIEGEL / *Nonparametric Statistics for the Behavioral Sciences*
STAGNER / *Psychology of Personality*
TOWNSEND / *Introduction to Experimental Methods for Psychology and the Social Sciences*
VINACKE / *The Psychology of Thinking*
WALLEN / *Clinical Psychology: The Study of Persons*
WARREN AND AKERT / *The Frontal Granular Cortex and Behavior*
WATERS, RETHLINGSHAFER, AND CALDWELL / *Principles of Comparative Psychology*
WINER / *Statistical Principles in Experimental Design*
ZUBEK AND SOLBERG / *Human Development*

Preface

The objective of this book is to present an elementary discussion of the basic problems of psychological measurement. Since I have taught a course in theory of psychological measurement for many years, it is natural, I suppose, for me to have developed certain ideas about what topics should be included and how the material should be presented. Unquestionably others will disagree with my approaches and developments. This is as it should be, since clarification of concepts emerges from the comparison of varying points of view about them.

The topics covered in the book are few. The initial discussion of the nature and types of variables leads up to the use and development of norms and the standardizing of scores. There follows an extended discussion of correlation, because it is my opinion that an understanding of the basis and nature of the Pearsonian coefficient is absolutely necessary if the student is to grasp the concepts of reliability and validity. In view of the fact that most scores are combinations of other scores, composite variables are examined in some detail. This also provides an opportunity to present item analysis and permits the development of formulae important in reliability and validity of measurement. In connection with reliability I have presented conceptual formulations other than the classical Yule-Spearman view. Thus a variety of interpretations of reliability are possible as well as differing evaluations of ways for estimating reliability. Similarly, differing notions of validity are presented, though more attention is given to predictive validity than to either content or construct validity.

At the end of each chapter is a list of selected readings. These were chosen with the idea of providing the more interested students some historical perspective and varying points of view.

Almost all formulae are developed in full. Many will not approve of this, especially in view of the fact that the developments are in painful detail. However, it seems to me that insofar as possible students must understand the logic which underlies the formulae they will be using. Perhaps if undergraduate majors in psychology were well prepared in mathematics, much of the detail would not be necessary—but in my experience they are not. I try to teach my students to read mathematical developments in much the same manner that they read a logical argument given in verbal terms.

In my classical education the blacks and whites of Dante's "Divina

Commedia" disturbed me. But among Pirandello's subtle variations of gray I was completely comfortable. I have no difficulty tolerating ambiguity, and indeed paradoxy seems to me to be the normal state of affairs. It will bother some that I do not take a position on such matters as the normalizing of distributions of scores, linearity and homoscedasticity in correlation, reliability, and validity. And that I fail to evaluate various points of view on these issues, painting one white and all the rest black, will seem the height of maleficence to many. But with Pirandello I prefer merely to present the contrasting notions and let the reader decide for himself which, if any, is the *real* truth—"Così è se vi pare."

Various portions of this book have gone through several mimeographed editions which I have used in my course in psychological measurement. To the many students who have tried to learn from these preliminary editions and who have made numerous helpful comments I am duly grateful. I also wish to express my appreciation to my colleague Professor Read D. Tuddenham, who kindly reviewed many of the chapters. I especially wish to thank Professor Edward Minium of San Jose State College, who spent so much time reviewing the manuscript in detail and made many constructive suggestions. Finally I should like to thank Drs. M. D. Davidoff and H. W. Goheen and the editors of *Psychometrika* for permission to reproduce Table 6-2, used in estimating the tetrachoric coefficient of correlation.

In dedicating this book to Professor Robert C. Tryon I wish to acknowledge at least in some small way the stimulation, support, and above all the warm friendship he has so freely given me. His enthusiasm, ingenuity, and wisdom invariably leave their marks on all who are associated with him. It was more than three decades ago, when we were both fledglings, that as a student I was first exposed to measurement theory in his course in individual differences. For twenty-five years we have been colleagues, and each day he has given me some new insight, a new idea to steal, a bit of wit to make life sparkle. Inasmuch as he has willingly continued to be my tutor, I could feel fully justified in holding him accountable for my errors and bad logic which appear in these pages. But some slight measure of the tolerance which so characterizes this gentle man has brushed off on me and so I forgive him the shortcomings of this text.

Edwin E. Ghiselli

Contents

chapter I Introduction

When we compare the characteristics of amoebae, sharks, cats, and human beings we see vast differences. The oozing locomotion of the amoeba, the graceful undulation of the shark, the loose easy pace of the cat, and the jerky walk of the human being are clearly dissimilar. Indeed, so vast are the differences between the various kinds of organisms that the differences among individuals of any one kind seem small and insignificant.

Yet if we direct our attention toward any one kind of creature we see quite significant differences. One amoeba flows along an undeviating path whereas another wanders aimlessly. One shark thrashes through the waters with a great commotion whereas another progresses with steady rhythmic oscillations. One cat boldly pursues a direct course whereas another furtively darts from one protective shelter to another. One man walks "ten feet tall" with shoulders back whereas another shuffles along with a slouch.

The differences among individuals of any one kind, then, are very real and substantial in magnitude. The task of the psychologist is to understand individuals. He is therefore interested in the sources of the differences and in the factors that in a given situation cause one individual to behave in one way and another to behave quite differently. He is interested, too, in the factors that cause changes in the individual—improvements in performance, reductions in ability, or cyclical variations in social adjustment.

THE PURPOSE OF PSYCHOLOGICAL MEASUREMENT

Obviously, before the factors determining the psychological characteristics of individuals can be systematically examined those very characteristics must be described. This is the purpose of psychological measurement. Measurement essentially is concerned with the operations designed to provide quantitative descriptions of the extent to which individuals manifest or possess specified traits. Therefore it is important that, as students of psychology, we understand something of the theory and problems of psychological measurement and of the nature and use of quantitative descriptions as given by the operations of measurement.

Many varied operations have been devised for measuring the numerous traits that psychologists see living creatures as manifesting or possessing. There are sundry tests of knowledge and of problem solving, questionnaires wherein the individual reports his perceptions about himself, and rating procedures that require one individual to judge the traits of others. In one way or another all these devices yield descriptions of the extent to which individuals manifest or possess specified traits. These quantitative descriptions, or scores, given by the various operations of measurement differ widely in their utility and meaningfulness. Some are valuable as they stand, whereas others must be subjected to further operations, usually of a mathematical nature, so that they can be transformed to other values or scores which are more meaningful and useful.

TOPICS IN PSYCHOLOGICAL MEASUREMENT TO BE COVERED

Psychological measurement is a broad field, replete with both methodological and theoretical problems. In this book we shall not cover all the basic problems, nor shall we even be able to present a complete discussion of those we do consider. Our principal concern will be with providing an introduction to some of the fundamental aspects of measurement and of the description of the differences among individuals.

We shall begin with a general consideration of some of the basic problems in the psychological description of individuals, discussing such matters as the definition of psychological traits, the development of operations designed to measure those traits, and the ways in which individuals differ from one another. This will lead us to a consideration of the use of distributions of scores, or norms, as a basis for giving some meaning to scores and for making scores on different traits comparable. Therefore we shall have to discuss ways and means for describing and dealing with distributions of scores and for standardizing scores on distributions.

For many practical and theoretical problems, indices of the degree of relationship between traits are necessary. Such indices tell us about the similarity in the order of individuals on different traits, the accuracy with which we can distinguish one trait from another, and something of the factors that determine differences among individuals. Consequently we shall delve somewhat into the theory of correlation and shall examine ways for describing the degree of relationship between traits.

Many of the tests and other operations we use in measuring individuals are complex, being composed of a series of items, subtests, or other kinds of parts or components which are added together. Consequently we must examine the combination of components into composites, and the relationships between scores on different composites.

The instruments we use to describe individuals quantitatively in terms of various psychological traits never measure those traits with perfect precision. Some are highly reliable and others have a lesser degree of reliability. Reliability of measurement is a problem of fundamental importance, but one not easily stated. We shall consider the various theoretical approaches that have been made to the problem of reliability and shall discuss ways and means for estimating the degree to which a test gives reliable measurements.

We shall return to the problem of composite variables and consider ways in which the components forming a composite can be differentially weighted, along with the effects of such weighting. Differential weighting of variables has considerable importance with respect to the problem of prediction of an individual's standing on one trait from a composite of his scores on a series of other traits.

Finally, we shall discuss the problem of validity of measurement. Validity has to do with prediction and with the determination of the traits actually measured by a given test. As is the case with reliability of measurement, we shall see that there are differing views about the nature of validity of measurement. We shall examine these various approaches to validity in order to gain some insight into the problem.

THEORIES AND MODELS

It has so often been said that the world is complex that the import of this statement is frequently forgotten. Indeed, sometimes it seems that the complexity is so great that no one man or group of men will ever be able to understand it completely. We have a few bits of knowledge which we try to piece together to make some kind of a meaningful whole. Often, however, the facts available are not sufficient to make a total integrated picture. Furthermore, that which at one time appears to be a fact seems pure fantasy at another time. Once it was obvious that the sun revolved around the earth. Now we know that the reverse is true. Even if we do have a goodly number of facts, and they are facts which "stand still" and remain facts, the complexities of their relationships usually are so great that we are unable to grasp them.

Yet we must deal with the world and its inhabitants, and therefore some understanding of it is necessary. In order to integrate the fragments of knowledge we do have and to bring some comprehensible order into the chaos of ignorance and some understanding into the complexity, we often do a peculiar thing—we create a world in our minds and pretend it is the real world. When as scientists we create a substitute world, we term it a theoretical model.

A theoretical model, therefore, "stands in" for some phenomenon, that is, is a representation of it. In order to provide an integration of the partial knowledge we have about the phenomenon and to make it meaningful, the model must necessarily go beyond the known facts. Consequently the model describes features of the phenomenon which are not given directly by the facts. If the model is a good representation, the inferred features will be like those of the phenomenon; if it is a poor one, they will not.

Consider the ancients' view of the shape of the world. They noted, for example, that unless a surface is flat, objects placed on it slide off. Since people and things maintain their position on the surface of the earth and do not slide off, it seemed to the ancients that the surface of the earth necessarily must be flat. To be sure, there appeared to be some inconsistencies such as the fact that ships approaching from the distance appeared to rise out of the sea. But this could be explained away as being due to valleys and hills on the sea similar in nature to the valleys and hills on the land, irregular fluctuations from the basic plane of the earth.

In many instances where we are dealing with phenomena about which we have only fragments of information, we develop a theory or model into which the bits fit so as to organize them into a meaningful whole. Many of the features of our model are not given by the available facts, and we develop them by interpolations and extrapolations from those facts we do have, by inference, by plain "horseback" guess, and by nothing more nor less than our imagination. Thus the ancients' model of the flat earth had an edge, a top, and a bottom even though no one had ever seen the edge and the bottom, because they observed that all flat things like tabletops and coins have edges, tops, and bottoms. Since objects have to be supported by something or else they fall, the ancients decided that a giant must be holding up the earth.

Today it is obvious to us that the ancients' model of the world is incorrect. Yet in spite of their completely false notions they were able to accomplish some remarkable feats of navigation. In the time of the pharaohs the Egyptians were able to sail from the Red Sea to the far southern regions of Africa, and the Phoenicians were able to sail from their home ports in the Middle East to the distant land now called Spain.

Thus a model may be manifestly incorrect and still be useful as a description of a set of phenomena and useful for making predictions about them. Being more sophisticated and wiser than our ancestors, we realize that the world is round. If we wish to sail or fly from San Francisco to Singapore, we plot our course as an arc on a sphere. Nevertheless, if we were planning a short voyage or a flight—say from San Francisco to Los Angeles—we should probably not bother to use the more complicated spherical model but should be satisfied with the simpler flat model of the ancients. We should plot our course as a straight line on a plane as given by a Mercator projection rather than as an arc on a globe. For the relatively short distance involved, the error resulting from the use of the incorrect flat model rather than the more correct spherical model is, for all practical purposes, negligible. Practicality, then, may lead us to adopt a model which we know is incorrect but which nonetheless in restricted areas is useful for description and prediction.

Superior though our spherical model of the world may be compared with the flat model of the ancients, we must recognize that it, too, is not correct. Common observation indicates that the surface of the earth, pocked as it is with valleys and knobbed with mountains, is certainly not like the surface of a smooth sphere. If we use a sphere as a model it is only an approximate representation of the surface of the earth. The exterior of the earth obviously is a very complex surface, not a simple and smooth one like the surface of a true sphere. Nevertheless, we might say that it is not unreasonable to take a sphere as a representation of the earth. The differences between the surface of a sphere and the true surface of the earth are not great, and for practical purposes of planning a journey we do not introduce much error by pretending the world is spherical. As a matter of fact, recent evidence suggests that the earth is somewhat pear-shaped. The surface of a pear-shaped object is far more complex than that of a sphere and consequently more difficult to deal with. But if the "pearedness" of the earth is not too great, a spherical model may still serve with reasonable accuracy. If we believe the error is not too great, we should be willing to tolerate it for the sake of greater ease of comprehending and describing movements along the surface. Indeed, we do not know precisely what the shape of the world is. We do know that it is very complicated and that, while certainly not precisely spherical, it is nevertheless something like a sphere. Furthermore, the shape of the earth is not constant but in fact changes over time. Various celestial and terrestrial forces are at work, and even puny man changes the surface of the earth. A model might therefore be adopted to simplify a complex situation, to make it understandable and consequently "livable" even though the model is contrary to fact.

If we decide that we are going to consider the world as a sphere, then we can represent it by a physical model, perhaps by means of a ball. But the mathematical properties of a sphere are well known, so instead of a ball we could use as our model the various mathematical formulae which describe a sphere. There are formulae which permit us to calculate the total surface of a sphere or the various portions of it, the lengths of arcs on its surface, and the like. All these formulae are related to one another and provide ways for describing spheres. Knowing the characteristics of certain parts of a sphere, we can calculate the characteristics of other parts. In some instances, then, it is not necessary to have a physical model as a representation of a set of phenomena; instead, we can use a mathematical model, which gives us very precise descriptions and is more convenient.

As we consider the various problems of psychological measurement we shall find it helpful from time to time to talk about models. We shall find models not only very useful for purposes of description and prediction but also valuable for simplification of complex phenomena so that they become comprehensible. The models in many instances may be manifestly incorrect representations; nevertheless, they are useful when we are beset with incomplete facts and rendered helpless by a complexity beyond our understanding.

Suppose that, in order to describe the way in which intelligence is distributed among children, we administer an intelligence test to 50 children. We shall undoubtedly find that the distribution of scores is quite irregular. For instance, the number of children who have IQs of 99 and 101 may be greater than the number who have an IQ of 100. Since our facts are few, there being only 50 children, we are likely to ascribe the irregularity in the distribution to an insufficient number of cases. We should say that if we had a larger number of children the distribution would be quite smooth. We should therefore disregard certain of our facts and adopt as a model of the distribution of intelligence one which is smooth but has the same average and variability as our actual distribution.

When we give the same test many times to the same individuals we often find that the scores of each person change from time to time. The variations in an individual's scores are random, showing no systematic improvement or reduction. In order to understand these results we say that each person possesses a given true amount of the ability measured by the test but his score varies from time to time because of the effects of unsystematic and variable factors in the situation. On one occasion a person's score may be a little lower because he has a cold, and on another it may be a little higher because the lighting is extra bright and he can see what he is doing. We do not know that this actually occurs, but it is a reasonable explanation of the variation. So we can set up a simple mathematical model to explain and to understand test performance by saying that *the score an individual obtains on a given administration of a test is his true ability plus an error.*

So we see that it is sometimes helpful to use theoretical models to represent distributions of scores rather than the actual distribution itself, and to represent the factors that determine an individual's performance on a test. In this same way models are useful for dealing with other problems in psychological measurement when our facts are incomplete, when we have doubts about the adequacy of our measuring devices, and when the phenomena with which we deal are complex.

chapter 2 The Description of Individual Differences

As we observe the world around us we see that the individual objects in it—be they things, animals, people, or social groups—differ among themselves in a variety of ways. Trees differ among themselves in height, animals in speed of locomotion, human beings in verbal facility, and social groups in homogeneity. Trees also differ among themselves in the nature of their foliage, animals in the ease with which they can be trained, human beings in the patterns of their personality traits, and social groups in their type of leadership. Not all individuals of a given kind, then, are the same with respect to some particular property. Rather, the rule is that individuals differ among themselves.

Sometimes the variation among individuals is qualitative, being in terms of kind; and in other instances it is quantitative, being in terms of frequency, amount, or degree. When we propose to study individual differences we must specify or define the property with which we are concerned. From this definition we can develop a series of operations that will permit us to describe individuals in terms of that property. Qualitative description is termed classification, and quantitative description is termed measurement. Measurement involves the use of numbers, that is, values which provide quantitative descriptions of individuals and which can be manipulated to give us further information about those individuals.