

*Fourth Edition*

# Psychological Testing

*Principles and Applications*



Kevin R. Murphy  
Charles O. Davidshofer



# Psychological Testing

## Principles and Applications



Fourth Edition



Kevin R. Murphy  
Colorado State University

Charles O. Davidshofer  
Colorado State University



Prentice Hall, Upper Saddle River, New Jersey 07458

*Library of Congress Cataloging-in-Publication Data*

MURPHY, KEVIN R., (date)

Psychological testing : principles and applications / Kevin R.

Murphy, Charles O. Davidshofer. — 4th ed.

p. cm.

Includes bibliographical references and indexes.

ISBN 0-13-263815-0

1. Psychological test. 2. Psychometrics. I. Davidshofer, Charles O. II. Title.

BF176.M87 1998

150'.28'7—dc21

97-9966

CIP

Editor in Chief: *Nancy Roberts*

Executive Editor: *Bill Webber*

Editorial Assistant: *Emsel Hasan*

Production Liaison: *Fran Russello*

Editorial/Production Supervision: *Kim Gueterman*

Buyer: *Lynn Pearlman*

Cover Designer: *Bruce Kenselaar*

Illustrator: *Asterisk Group*

Director, Image Resource Center: *Lori Morris-Nantz*

Photo Research Supervisor: *Melinda Lee Reo*

Image Permission Supervisor: *Kay Dellosa*

Photo Researcher: *Rona Tuccillo*

Copyeditor: *William Thomas*

Marketing Manager: *Mike Alread*

This book was set in 10/12 Palatino by The Composing Room of Michigan, Inc.  
and was printed and bound by Hamilton Printing Company.  
The cover was printed by Phoenix Color Corp.



© 1998, 1994, 1991, 1988 by Prentice-Hall, Inc.  
Simon & Schuster / A Viacom Company  
Upper Saddle River, New Jersey 07458

All rights reserved. No part of this book may be  
reproduced, in any form or by any means,  
without permission in writing from the publisher,

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

ISBN 0-13-263815-0

Prentice-Hall International (UK) Limited, *London*

Prentice-Hall of Australia Pty. Limited, *Sydney*

Prentice-Hall Canada, Inc., *Toronto*

Prentice-Hall Hispanoamericana, S. A., *Mexico*

Prentice-Hall of India Private Limited, *New Delhi*

Prentice-Hall of Japan, Inc., *Tokyo*

Simon & Schuster Asia Pte. Ltd., *Singapore*

Editora Prentice-Hall do Brasil, Ltda., *Rio de Janeiro*

# **Psychological Testing**



## Preface



*Tests are used to make decisions.* This simple sentence describes both the theme and the rationale of this book. Too often, psychological testing is presented as a dry, technical, abstract subject. It is not. Psychological tests have a substantial impact on a variety of important decisions. In some settings, such as the armed services, tests represent the *only* feasible method of making selection and classification decisions. The point is that tests affect people's lives, for good or for ill, and a firm understanding of psychological testing is necessary in many of the settings where important decisions (e.g., college admissions, job placement, clinical assessments) must be made about individuals.

Most students in a testing course have taken many tests and will probably take many more in the future. These students have a real, practical interest in testing. They are also likely to be skeptical about the accuracy and the value of psychological tests. A good part of your job when teaching a testing course is to present a full and fair evaluation of the advantages and drawbacks of psychological testing. We believe that this can best be done by focusing on the impact of tests on decisions. Students generally show little interest in learning a "laundry list" of test names, but they are very interested in knowing how tests are used and whether the use of tests leads to better or worse decisions than would be reached without tests.

Our text is divided into four sections. In Part I (Chapters 1 through 3), we introduce the concepts of psychological testing and discuss the impact of testing on society. Part II (Chapters 4 through 10) discusses the principles of psychological measurement and the techniques used to analyze tests. Part III (Chapters 11 through 17) discusses the development of tests, with particular attention to the domains of cognitive ability, interests,

and personality. Part IV (Chapters 18 through 22) discusses the use of psychological tests to make important decisions about individuals.

This book does *not* have several features found in other testing books. First, we do not cover dozens of obscure tests. Rather, we have focused on widely used or technically superior exemplars of the major classes of tests. You will not encounter in this book many tests that are never heard of outside a psychological measurement class, and, frankly, we see no good reason why you should. Second, we do not cover several topics that, while important, have little relevance for decision making. Thus, we do not discuss at length topics such as attitude measurement or the assessment of values.

Each chapter contains one or more brief sections entitled *Critical Discussion*. The critical discussion sections present a variety of issues that are controversial (e.g., Should IQ Scores of Black Examinees be Based on White Norms? in Chapter 5), or that illustrate applications of concepts discussed in the chapters (e.g., Using Item Response Theory to Detect Test Bias in Chapter 10), or that provide a different perspective of familiar material (e.g., Personnel Selection from the Applicant's Point of View in Chapter 19). These critical discussions may provide ideas for further classroom discussions, term papers, or projects and are designed to help your students integrate material from this course with their other interests and knowledge bases.

As you'll note when reading through the text, we are, on the whole, optimistic about psychological testing. There are areas where the technology has failed to keep up with the theory, where the basic theories are flawed, or where the applications of testing have done more harm than good. On the whole, however, psychological tests often provide the fairest and most accurate method of making important decisions. In part, this reflects the strength of tests and, in part, it reflects the weakness of the competition (e.g., interviews, letters of recommendation, clinical intuition). In any case, we think that psychological tests do make a contribution and are likely to be with us for a long time to come. We hope that this text will contribute to your students' understanding of the advantages and drawbacks of psychological testing.

We have made many changes in preparing this revision. First, we have rearranged several chapters in a way that we think makes the book easier to follow and more useful to students. Second, there have been many important changes in the field of testing since our last edition, most notably a major overhaul of the SAT (Scholastic Assessment Tests), and a new edition of the Sixteen Personality Factors inventory (16pf). Recent developments in a number of areas are reviewed in this edition.

Finally, at the end of every chapter, we identify and define several *key terms*, which represent important concepts introduced or discussed in that chapter. These key terms provide capsule summaries of important points introduced in each chapter.

## ACKNOWLEDGMENTS

This book could not have been produced without the help and encouragement of many of our colleagues. We would like to particularly thank Frank Landy, who encouraged us to write the first edition and who suggested the overall structure for the book. We also

thank Dennis Roberts, who collaborated in the early development of the book and who contributed substantially to the overall structure and focus of the book. Finally, we would like to thank colleagues who have helped us in formulating our ideas and presentations or alerted us to errors and omissions in our first three editions. These include Jeanette Cleveland, George Thornton, Kurt Geisinger, and numerous instructors and students who have provided useful suggestions. We also appreciate feedback from several of our colleagues who have suggested ways of improving our previous editions.

Kevin R. Murphy  
Charles O. Davidshofer



# Contents



## **PREFACE   xiii**

## **Part I   Introduction to Psychological Testing**

### **1   TESTS AND MEASUREMENTS   1**

- Psychological Tests—A Definition   3
- Types of Tests   6
- Tests and Decisions—Uses of Psychological Tests   9
- Testing Activities of Psychologists   12
- Information about Tests   13
- Standards for Testing   15
- Critical Discussion: Alternatives to Psychological Tests*   17
- Summary   18
- Key Terms   18

### **2   DEFINING AND MEASURING PSYCHOLOGICAL ATTRIBUTES: ABILITY, INTERESTS, AND PERSONALITY   19**

- Psychological Attributes and Decisions   19
- Intelligence—General Mental Ability   20



<i>Critical Discussion: IQ or EQ?</i>	34
Interests	35
Personality	39
<i>Critical Discussion: The Relationship among Abilities, Interests, and Personality Characteristics</i>	46
Summary	47
Key Terms	48

### **3 TESTING AND SOCIETY 49**

Types of Decisions	50
Societal Concerns	52
<i>Critical Discussion: Science and Politics in the Ability-testing Debate—The Bell Curve</i>	56
<i>Critical Discussion: The Strange Case of Sir Cyril Burt</i>	62
Analyzing the Societal Consequences of Tests	62
<i>Critical Discussion: Are There Technical Solutions to Societal Problems?</i>	65
Summary	65
Key Terms	66

## **Part II Principles of Psychological Measurement**

### **4 BASIC CONCEPTS IN MEASUREMENT AND STATISTICS 67**

Psychological Measurement	67
Evaluating Psychological Tests	73
Statistical Concepts	75
<i>Critical Discussion: Measurement Scales and Statistics</i>	83
Summary	84
Key Terms	85

### **5 SCALES, TRANSFORMATIONS, AND NORMS 86**

Transformations	87
Equating Scale Scores	91
Norms	95
Expectancy Tables	102

*Critical Discussion: Should IQ Scores of Black Examinees Be Based on White Norms?* 106

Summary 107

Key Terms 107

## **6 RELIABILITY: THE CONSISTENCY OF TEST SCORES 109**

Sources of Consistency and Inconsistency in Test Scores 110

General Model of Reliability 112

Simple Methods of Estimating Reliability 115

Reliability Estimates and Error 120

The Generalizability of Test Scores 121

*Critical Discussion: Generalizability and Reliability Theory* 124

Summary 125

Key Terms 126

## **7 USING AND INTERPRETING INFORMATION ABOUT TEST RELIABILITY 127**

Using the Reliability Coefficient 127

Factors Affecting Reliability 132

Special Issues in Reliability 136

How Reliable Should Tests Be? 142

*Critical Discussion: Precise but Narrow—The Bandwidth-Fidelity Dilemma* 143

Summary 144

Key Terms 144

## **8 VALIDITY OF MEASUREMENT: CONTENT AND CONSTRUCT-ORIENTED VALIDATION STRATEGIES 146**

Validation Strategies 147

Assessing the Validity of Measurement 148

Content-oriented Validation Strategies 149

*Critical Discussion: Face Validity* 155

Construct-oriented Validation Strategies 155

Content and Construct Validity 166

*Critical Discussion: How Do We Know Whether Intelligence Tests Really Measure Intelligence?* 167

Summary 168

Key Terms 169

## **9 VALIDITY FOR DECISIONS: CRITERION-RELATED VALIDITY 170**

Decision and Prediction 171

Criteria 172

Criterion-related Validation Strategies 173

Construct Validity and Criterion-related Validity 179

Interpreting Validity Coefficients 181

*Critical Discussion: Values and Validation* 190

Summary 191

Key Terms 192

## **10 ITEM ANALYSIS 193**

Purpose of Item Analysis 194

Distractor Analysis 195

Item Difficulty 196

Item Discrimination 199

Interactions among Item Characteristics 202

The Item Characteristic Curve 204

*Critical Discussion: Using Item Response Theory to Detect  
Test Bias* 213

Summary 214

Key Terms 215

## **Part III Developing Measures of Ability, Interests, and Personality**

## **11 THE PROCESS OF TEST DEVELOPMENT 216**

Constructing the Test 216

*Critical Discussion: Do Different Test Formats Give Unique  
Interpretative Information?* 226

Norming and Standardizing Tests 227

Test Publication and Revision 230

*Critical Discussion: Should Tests Be Normed Nationally, Regionally,  
or Locally?* 231

Summary	232
Key Terms	234

## **12 COMPUTERIZED TEST ADMINISTRATION AND INTERPRETATION 236**

Use of Computers in Testing: A Taxonomy	237
<i>Critical Discussion: Test Information on the Internet</i>	238
Computerized Test Administration	239
Computerized Adaptive Testing	241
Computer-based Test Interpretation	244
Detecting Invalid or Unusual Response Patterns	248
<i>Critical Discussion: The Barnum Effect and Computerized Test Interpretation</i>	249
Summary	250
Key Terms	250

## **13 ABILITY TESTING: INDIVIDUAL TESTS 251**

The Role of the Examiner	252
Individual Tests of General Mental Ability	254
Individual Tests of Specific Abilities	271
<i>Critical Discussion: Testing Examinees from Different Racial and Ethnic Groups</i>	272
Summary	273
Key Terms	274

## **14 ABILITY TESTING: GROUP TESTS 275**

Advantages and Disadvantages of Group Tests	276
Measures of General Cognitive Ability	277
Scholastic Tests	281
Multiple-aptitude Batteries	289
<i>Critical Discussion: The Case against Large-scale Testing</i>	296
Summary	297
Key Terms	298

## **15 ISSUES IN ABILITY TESTING 299**

Bias and Fairness in Mental Testing	300
<i>Critical Discussion: Are Judgments of Item Bias Useful?</i>	312

Culture-reduced Testing	313
<i>Critical Discussion: Intelligence and Adaptive Behavior</i>	320
Heritability, Consistency, and Modifiability of IQ	321
<i>Critical Discussion: Intelligence and Social Class</i>	324
<i>Critical Discussion: Project Head Start</i>	329
Summary	330
Key Terms	331

## **16 INTEREST TESTING 332**

The Strong Interest Inventory	333
The Kuder Occupational Interest Survey	348
Career Assessment Inventory	355
Jackson Vocational Interest Survey	360
<i>Critical Discussion: Should Interest Tests Be Used in Personnel Selection?</i>	365
Summary	368
Key Terms	368

## **17 PERSONALITY TESTING 369**

Development of Personality Testing	369
Objective Measures of Personality	370
Projective Tests of Personality	382
<i>Critical Discussion: Are Scales on Personality Tests Interpreted Independently?</i>	389
<i>Critical Discussion: Reliability and Validity of Projective Tests, Personality Inventories, and Intelligence Tests</i>	390
Summary	391
Key Terms	392

## **Part IV Using Tests to Make Decisions**

### **18 TESTS AND EDUCATIONAL DECISIONS 393**

Applications in Institutional Decision Making	394
<i>Critical Discussion: The National Assessment of Educational Progress</i>	400
Applications in Individual Decision Making	401

*Critical Discussion: What Is Portfolio Assessment? Should It be Used for Educational Assessment?* 404

*Critical Discussion: What Is Criterion-referenced Testing?* 405

Summary 406

Key Terms 407

## **19 PSYCHOLOGICAL MEASUREMENT IN INDUSTRY: ISSUES IN PREDICTION 408**

Extent and Impact of Personnel Testing 408

Predicting Performance 410

Comparative Assessment of Personnel Tests 425

*Critical Discussion: Can Graphology Predict Occupational Success?* 426

Assessing Integrity 427

The Validity of Personnel Tests 431

The Social and Legal Context of Employment Testing 434

*Critical Discussion: Personnel Selection from the Applicant's Point of View* 437

Summary 438

Key Terms 439

## **20 PSYCHOLOGICAL MEASUREMENT IN INDUSTRY: CRITERION MEASUREMENT 441**

Criterion Measures 441

Judgmental Measures 446

Can We Measure the Value of Performance? 460

*Critical Discussion: The Ethics of Evaluation—Telling More Than You Know* 462

Summary 463

Key Terms 464

## **21 DIAGNOSTIC TESTING: CLINICAL APPLICATIONS 465**

The Minnesota Multiphasic Personality Inventory 466

MMPI-2 472

The Bender–Gestalt 473

Scatter Analysis 478

Diagnostic Classification System 480

*Critical Discussion: Professional Ethics  
and Clinical Testing* 485

Summary 485

Key Terms 486

## **22 CLINICAL ASSESSMENT 487**

The Nature of Clinical Assessment 488

Studying Clinical Judgment 492

Structured Assessment Programs 501

*Critical Discussion: Psychological Assessment  
in the Courtroom* 511

Summary 512

Key Terms 513

### **APPENDIX A: FORTY REPRESENTATIVE TESTS 515**

### **APPENDIX B: ETHICAL PRINCIPLES OF PSYCHOLOGISTS AND CODE OF CONDUCT 517**

### **APPENDIX C: CODE OF FAIR TESTING PRACTICES IN EDUCATION 536**

### **APPENDIX D: REVIEW OF BASIC STATISTICS 540**

### **REFERENCES 547**

### **AUTHOR INDEX 587**

### **SUBJECT INDEX 595**



# Tests and Measurements



The term *psychological test* brings to mind a number of conflicting images. On the one hand, the term might make one think of the type of test so often described in television, movies, and the popular literature, wherein a patient answers questions like, “How long have you hated your mother?” and in doing so reveals hidden facets of his or her personality to the clinician. On the other hand, the psychological test might refer to a long series of multiple-choice questions such as those answered by hundreds of high school students taking college entrance examinations. Another type of “psychological test” is the self-scored type published in the *Reader’s Digest*, which purports to tell you whether your marriage is on the rocks, whether you are as anxious as the next fellow, or whether you should change your job or your life-style.

In general, psychological tests are neither mysterious, as our first example might suggest, nor frivolous, as our last example might suggest. Rather, psychological tests represent systematic applications of a few relatively simple principles in an attempt to measure personal attributes thought to be important in describing or understanding individual behavior. The aim of this book is to describe the basic principles of psychological measurement and to describe the major types of tests and their applications. We will not present test theory in all its technical detail, nor will we describe (or even mention) all the different psychological tests currently available. Rather, our goal is to provide the information needed to make sensible evaluations of psychological tests and their uses within education, industry, and clinical practice.

The first question that should be addressed in a psychological testing text is, “Why is psychological testing important?” There are several possible answers to this question,



but we believe that the best answer lies in the simple statement that forms the central theme of this book: *Tests are used to make important decisions about individuals*. College admissions officers consult test scores before deciding whether to admit or reject applicants. Clinical psychologists use a variety of objective and projective tests in the process of choosing a course of treatment for individual clients. The military uses test scores as aids in deciding which jobs an individual soldier might be qualified to fill. Tests are used in the world of work, both in personnel selection and in professional certification and licensure. Almost everyone reading this book has taken at least one standardized psychological test. Scores on such a test may have had some impact on an important decision that has affected your life. The area of psychological testing is therefore one of considerable practical importance.

Psychological tests are used to measure a wide variety of attributes—intelligence, motivation, mastery of seventh-grade mathematics, vocational preferences, spatial ability, anxiety, form perception, and countless others. Unfortunately, one feature that all psychological tests share in common is their limited precision. They rarely, if ever, provide exact, definitive measures of variables that are believed to have important effects on human behavior. Thus, psychological tests do not provide a basis for making *completely* accurate decisions about individuals. In reality, no method guarantees complete accuracy. Thus, although psychological tests are known to be imperfect measures, a special panel of the National Academy of Sciences concluded that psychological tests generally represent the best, fairest, and most economical method of obtaining the information necessary to make sensible decisions about individuals (Wigdor & Garner, 1982a, 1982b). The conclusions reached by the National Academy panel form another important theme that runs through this book. Although psychological tests are far from perfect, *they represent the best, fairest, and most accurate technology available for making many important decisions about individuals*.

Psychological testing is highly controversial. Public debate over the use of tests, particularly standardized tests of intelligence, has raged since at least the 1920s (Cronbach, 1975; Haney, 1981; Scarr, 1989).<sup>1</sup> There is presently extensive literature, both popular and technical, dealing with issues such as test bias and test fairness. Federal and state laws have been passed calling for minimum competency testing and for truth in testing, terms that refer to a variety of efforts to regulate testing and to increase public access to information on test development and use. Tests and testing programs have been challenged in the courts, often successfully.

Psychological testing is not only important and controversial, it is a highly specialized and somewhat technical enterprise. In many of the natural sciences, measurement is a relatively straightforward process that involves assessing the physical properties of objects, such as height, weight, or velocity.<sup>2</sup> However, for the most part,

<sup>1</sup>Special issues of *American Psychologist* in November 1965 and October 1981 provide excellent summaries of many of the issues in this debate.

<sup>2</sup>Note, however, that physical measurement is neither static nor simple. Proposals to redefine the basic unit of length, the meter, in terms of the time light takes to travel from point to point (Robinson, 1983) provide an example of continuing progress in redefining the bases of physical measurement.