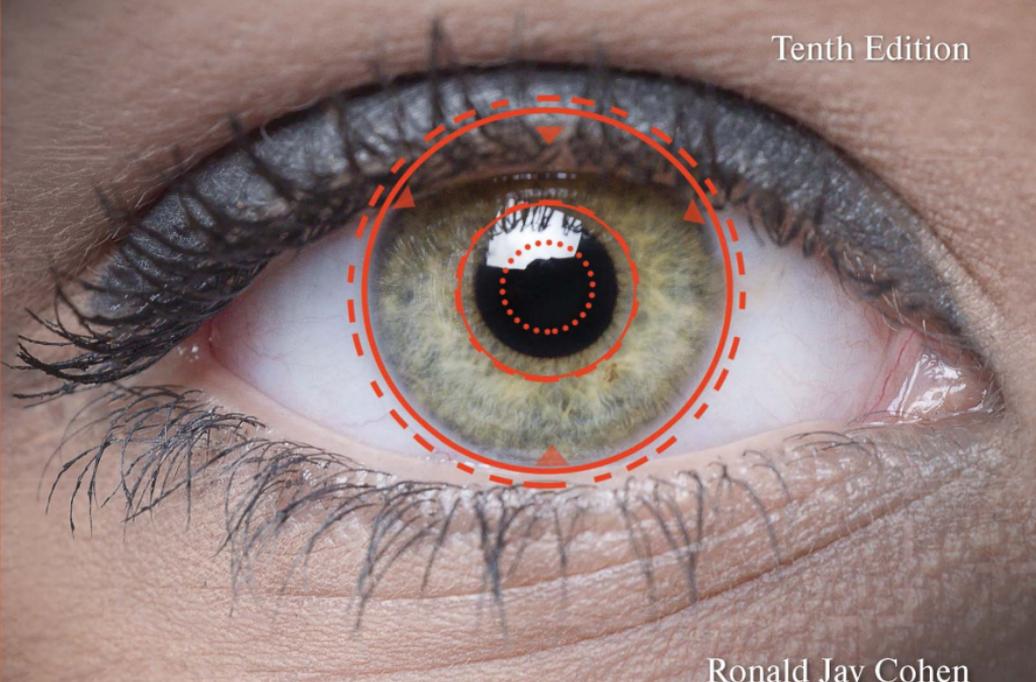


This International Student Edition is for use outside of the U.S.

# Psychological Testing and Assessment

*An Introduction to Tests and Measurement*

Tenth Edition



**Mc  
Graw  
Hill**

Ronald Jay Cohen

W. Joel Schneider

Renée M. Tobin

# Psychological Testing and Assessment

*An Introduction to Tests and Measurement*

**TENTH EDITION**

**Ronald Jay Cohen**

RJ COHEN CONSULTING

**W. Joel Schneider**

TEMPLE UNIVERSITY

**Renée M. Tobin**

TEMPLE UNIVERSITY





## PSYCHOLOGICAL TESTING AND ASSESSMENT

Published by McGraw Hill LLC, 1325 Avenue of the Americas, New York, NY 10121.  
Copyright ©2022 by McGraw Hill LLC. All rights reserved. Printed in the United States of America. No part of this publication may be reproduced or distributed in any form or by any means, or stored in a database or retrieval system, without the prior written consent of McGraw Hill LLC, including, but not limited to, in any network or other electronic storage or transmission, or broadcast for distance learning.

Some ancillaries, including electronic and print components, may not be available to customers outside the United States.

This book is printed on acid-free paper.

1 2 3 4 5 6 7 8 9 LWI 26 25 24 23 22 21

ISBN 978-1-265-79973-1

MHID 1-265-79973-3

Cover Image: *rimom/Shutterstock*

All credits appearing on page or at the end of the book are considered to be an extension of the copyright page.

The Internet addresses listed in the text were accurate at the time of publication. The inclusion of a website does not indicate an endorsement by the authors or McGraw Hill LLC, and McGraw Hill LLC does not guarantee the accuracy of the information presented at these sites.

*This book is dedicated with love to the memory of Edith and Harold Cohen.*



© 2017 Ronald Jay Cohen. All rights reserved.

# Contents

Preface xiii

## **PART |** *An Overview*

### **1 Psychological Testing and Assessment 1**

TESTING AND ASSESSMENT 1

Psychological Testing and Assessment Defined 2

THE TOOLS OF PSYCHOLOGICAL ASSESSMENT 8

The Test 8

The Interview 10

The Portfolio 12

Case History Data 13

Behavioral Observation 13

Role-Play Tests 14

Computers as Tools 15

Other Tools 18

WHO, WHAT, WHY, HOW, AND WHERE? 18

Who Are the Parties? 19

In What Types of Settings Are Assessments Conducted, and Why? 21

How Are Assessments Conducted? 27

Where to Go for Authoritative Information: Reference Sources 33

CLOSE-UP Behavioral Assessment Using Smartphones 5

MEET AN ASSESSMENT PROFESSIONAL Meet Dr. Alan Ogle 25

EVERYDAY PSYCHOMETRICS Everyday Accommodations 32

SELF-ASSESSMENT 36

REFERENCES 36

### **2 Historical, Cultural, and Legal/Ethical Considerations 41**

A HISTORICAL PERSPECTIVE 41

Antiquity to the Nineteenth Century 41

The Twentieth Century 44

CULTURE AND ASSESSMENT 47

Evolving Interest in Culture-Related Issues 47

Some Issues Regarding Culture and Assessment 52

Tests and Group Membership 58

LEGAL AND ETHICAL CONSIDERATIONS 60

The Concerns of the Public 60

The Concerns of the Profession 68

The Rights of Testtakers 74

CLOSE-UP The Controversial Career of Henry Herbert Goddard	49
MEET AN ASSESSMENT PROFESSIONAL Meet Dr. Neil Krishan Aggarwal	56
EVERYDAY PSYCHOMETRICS Life-or-Death Psychological Assessment	71
SELF-ASSESSMENT	79
REFERENCES	80

## **PART II** *The Science of Psychological Measurement*

### **3 A Statistics Refresher 85**

SCALES OF MEASUREMENT	86
Nominal Scales	88
Ordinal Scales	89
Interval Scales	90
Ratio Scales	91
Measurement Scales in Psychology	91
DESCRIBING DATA	93
Frequency Distributions	93
Measures of Central Tendency	98
Measures of Variability	101
Skewness	105
Kurtosis	105
THE NORMAL CURVE	106
The Area Under the Normal Curve	107
STANDARD SCORES	110
z Scores	110
T Scores	111
Other Standard Scores	111
CORRELATION AND INFERENCE	113
The Concept of Correlation	114
The Pearson $r$	116
The Spearman Rho	118
Graphic Representations of Correlation	119
Meta-Analysis	123
EVERYDAY PSYCHOMETRICS Consumer (of Graphed Data), Beware!	97
CLOSE-UP The Normal Curve and Psychological Tests	108
MEET AN ASSESSMENT PROFESSIONAL Meet Dr. Joni L. Mihura	124
SELF-ASSESSMENT	126
REFERENCES	127

### **4 Of Tests and Testing 129**

SOME ASSUMPTIONS ABOUT PSYCHOLOGICAL TESTING AND ASSESSMENT	130
Assumption 1: Psychological Traits and States Exist	130

Assumption 2: Psychological Traits and States Can Be Quantified and Measured	132
Assumption 3: Test-Related Behavior Predicts Non-Test-Related Behavior	133
Assumption 4: All Tests Have Limits and Imperfections	133
Assumption 5: Various Sources of Error Are Part of the Assessment Process	134
Assumption 6: Unfair and Biased Assessment Procedures Can Be Identified and Reformed	134
Assumption 7: Testing and Assessment Offer Powerful Benefits to Society	135
WHAT'S A "GOOD TEST"?	136
Reliability	136
Validity	137
Other Considerations	137
NORMS	140
Sampling to Develop Norms	140
Types of Norms	146
Fixed Reference Group Scoring Systems	149
Norm-Referenced versus Criterion-Referenced Evaluation	150
Culture and Inference	153
EVERYDAY PSYCHOMETRICS	Putting Tests to the Test 138
CLOSE-UP	How "Standard" Is <i>Standard</i> in Measurement? 141
MEET AN ASSESSMENT PROFESSIONAL	Meet Dr. Steve Julius and Dr. Howard W. Atlas 152
SELF-ASSESSMENT	154
REFERENCES	155
<b>5 Reliability</b>	<b>157</b>
MEASUREMENT ERROR	157
TRUE SCORES VERSUS CONSTRUCT SCORES	158
THE CONCEPT OF RELIABILITY	159
Sources of Error Variance	160
RELIABILITY ESTIMATES	163
Test-Retest Reliability Estimates	163
Parallel-Forms and Alternate-Forms Reliability Estimates	164
Split-Half Reliability Estimates	167
Other Methods of Estimating Internal Consistency	170
Measures of Inter-Scorer Reliability	172
USING AND INTERPRETING A COEFFICIENT OF RELIABILITY	174
The Purpose of the Reliability Coefficient	175
The Nature of the Test	176
The True Score Model of Measurement and Alternatives to It	179
RELIABILITY AND INDIVIDUAL SCORES	183
The Standard Error of Measurement	183
The Standard Error of the Difference Between Two Scores	187
CLOSE-UP	Psychology's Replicability Crisis 165
EVERYDAY PSYCHOMETRICS	The Importance of the Method Used for Estimating Reliability 173
MEET AN ASSESSMENT PROFESSIONAL	Meet Dr. Bryce B. Reeve 184

SELF-ASSESSMENT 189

REFERENCES 190

## **6 Validity 193**

THE CONCEPT OF VALIDITY 193

Face Validity 195

Content Validity 196

CRITERION-RELATED VALIDITY 200

What Is a Criterion? 200

Concurrent Validity 202

Predictive Validity 202

CONSTRUCT VALIDITY 205

Evidence of Construct Validity 206

VALIDITY, BIAS, AND FAIRNESS 211

Test Bias 211

Test Fairness 214

MEET AN ASSESSMENT PROFESSIONAL Meet Dr. Adam Shoemaker 197

CLOSE-UP The Preliminary Validation of a Measure of Individual Differences in Constructive versus Unconstructive Worry 212

EVERYDAY PSYCHOMETRICS Adjustment of Test Scores by Group Membership: Fairness in Testing or Foul Play? 216

SELF-ASSESSMENT 218

REFERENCES 218

## **7 Utility 221**

WHAT IS TEST UTILITY? 222

Factors That Affect a Test's Utility 222

UTILITY ANALYSIS 227

What Is a Utility Analysis? 227

How Is a Utility Analysis Conducted? 228

Some Practical Considerations 242

METHODS FOR SETTING CUT SCORES 245

The Angoff Method 246

The Known Groups Method 246

IRT-Based Methods 247

Other Methods 248

MEET AN ASSESSMENT PROFESSIONAL Meet Dr. Delphine Courvoisier 225

CLOSE-UP Utility Analysis: An Illustration 229

EVERYDAY PSYCHOMETRICS The Utility of Police Use of Body Cameras 239

SELF-ASSESSMENT 248

REFERENCES 249

## **8 Test Development 251**

TEST CONCEPTUALIZATION 252

Some Preliminary Questions 254

Pilot Work	256
TEST CONSTRUCTION	256
Scaling	256
Writing Items	261
Scoring Items	268
TEST TRYOUT	268
What Is a Good Item?	269
ITEM ANALYSIS	270
The Item-Difficulty Index	270
The Item-Reliability Index	271
The Item-Validity Index	272
The Item-Discrimination Index	272
Item-Characteristic Curves	275
Other Considerations in Item Analysis	278
Qualitative Item Analysis	280
TEST REVISION	282
Test Revision as a Stage in New Test Development	282
Test Revision in the Life Cycle of an Existing Test	284
The Use of IRT in Building and Revising Tests	288
INSTRUCTOR-MADE TESTS FOR IN-CLASS USE	291
Addressing Concerns About Classroom Tests	291
CLOSE-UP Creating and Validating a Test of Asexuality	253
MEET AN ASSESSMENT PROFESSIONAL Meet Dr. Scott Birkeland	276
EVERYDAY PSYCHOMETRICS Adapting Tools of Assessment for Use with Specific Cultural Groups	283
SELF-ASSESSMENT	293
REFERENCES	294

## **PART** *The Assessment of Intelligence*

### **9 Intelligence and Its Measurement 297**

WHAT IS INTELLIGENCE?	297
Perspectives on Intelligence	299
MEASURING INTELLIGENCE	312
Some Tasks Used to Measure Intelligence	312
Some Tests Used to Measure Intelligence	314
ISSUES IN THE ASSESSMENT OF INTELLIGENCE	334
Culture and Measured Intelligence	335
The Flynn Effect	340
The Construct Validity of Tests of Intelligence	341
A PERSPECTIVE	341
CLOSE-UP Factor Analysis	302
MEET AN ASSESSMENT PROFESSIONAL Meet Dr. Rebecca Anderson	315
EVERYDAY PSYCHOMETRICS The Armed Services Vocational Aptitude Battery (ASVAB): A Test You Can Take	330

SELF-ASSESSMENT 342

REFERENCES 343

## 10 Assessment for Education 349

THE ROLE OF TESTING AND ASSESSMENT IN EDUCATION 349

THE CASE FOR AND AGAINST EDUCATIONAL TESTING IN THE SCHOOLS 350

THE COMMON CORE STATE STANDARDS 351

Response to Intervention (RtI) 352

Dynamic Assessment 358

ACHIEVEMENT TESTS 360

Measures of General Achievement 360

Measures of Achievement in Specific Subject Areas 361

APTITUDE TESTS 363

The Preschool Level 365

The Elementary-School Level 370

The Secondary-School Level 372

The College Level and Beyond 373

DIAGNOSTIC TESTS 376

Reading Tests 377

Math Tests 378

PSYCHOEDUCATIONAL TEST BATTERIES 378

The Kaufman Assessment Battery for Children, Second Edition Normative Update  
(KABC-II NU) 378

The Woodcock-Johnson IV (WJ IV) 380

OTHER TOOLS OF ASSESSMENT IN EDUCATIONAL SETTINGS 381

Performance, Portfolio, and Authentic Assessment 381

Peer Appraisal Techniques 383

Measuring Study Habits, Interests, and Attitudes 384

EVERYDAY PSYCHOMETRICS The *Common Core* Controversy 353

MEET AN ASSESSMENT PROFESSIONAL Meet Eliane Keyes, M.A. 357

CLOSE-UP Educational Assessment: An Eastern Perspective 371

SELF-ASSESSMENT 385

REFERENCES 385

## PART IV *The Assessment of Personality*

### 11 Personality Assessment: An Overview 390

PERSONALITY AND PERSONALITY ASSESSMENT 390

Personality 390

Personality Assessment 391

Traits, Types, and States 391

PERSONALITY ASSESSMENT: SOME BASIC QUESTIONS 395

Who? 396

What?	402
Where?	404
How?	404
DEVELOPING INSTRUMENTS TO ASSESS PERSONALITY	413
Logic and Reason	413
Theory	416
Data Reduction Methods	416
Criterion Groups	419
PERSONALITY ASSESSMENT AND CULTURE	431
Acculturation and Related Considerations	431
CLOSE-UP The Personality of Gorillas	397
EVERYDAY PSYCHOMETRICS Some Common Item Formats	408
MEET AN ASSESSMENT PROFESSIONAL Meet Dr. Rick Malone	414
SELF-ASSESSMENT	435
REFERENCES	435
<b>12 Personality Assessment Methods</b>	<b>444</b>
OBJECTIVE METHODS	444
How Objective Are Objective Methods of Personality Assessment?	445
PROJECTIVE METHODS	445
Inkblots as Projective Stimuli	447
Pictures as Projective Stimuli	453
Words as Projective Stimuli	461
Sounds as Projective Stimuli	464
The Production of Figure Drawings	465
Projective Methods in Perspective	468
BEHAVIORAL ASSESSMENT METHODS	472
The Who, What, When, Where, Why, and How of It	474
Varieties of Behavioral Assessment	478
Issues in Behavioral Assessment	485
A PERSPECTIVE	487
MEET AN ASSESSMENT PROFESSIONAL Meet Dr. Monica Webb Hooper	476
EVERYDAY PSYCHOMETRICS Confessions of a Behavior Rater	479
CLOSE-UP General (g) and Specific (s) Factors in the Diagnosis of Personality Disorders	488
SELF-ASSESSMENT	490
REFERENCES	490

## **PART V** *Testing and Assessment in Action*

### **13 Clinical and Counseling Assessment** 499

AN OVERVIEW	499
The Diagnosis of Mental Disorders	501

The Interview in Clinical Assessment	504
Case History Data	511
Psychological Tests	511
CULTURALLY INFORMED PSYCHOLOGICAL ASSESSMENT	513
Cultural Aspects of the Interview	515
SPECIAL APPLICATIONS OF CLINICAL MEASURES	518
The Assessment of Addiction and Substance Abuse	518
Forensic Psychological Assessment	520
Diagnosis and evaluation of emotional injury	526
Profiling	526
Custody Evaluations	527
CHILD ABUSE AND NEGLECT	530
Elder Abuse and Neglect	532
Suicide Assessment	534
THE PSYCHOLOGICAL REPORT	535
The Barnum Effect	535
Clinical Versus Mechanical Prediction	537
MEET AN ASSESSMENT PROFESSIONAL Meet Dr. Stephen Finn	507
CLOSE-UP PTSD in Returning Veterans and Military Culture	516
EVERYDAY PSYCHOMETRICS Measuring Financial Competency	524
SELF-ASSESSMENT	539
REFERENCES	540
<b>14 Neuropsychological Assessment</b>	<b>550</b>
THE NERVOUS SYSTEM AND BEHAVIOR	550
Neurological Damage and the Concept of Organicity	551
THE NEUROPSYCHOLOGICAL EVALUATION	554
When a Neuropsychological Evaluation Is Indicated	554
General Elements of a Neuropsychological Evaluation	556
The Physical Examination	559
NEUROPSYCHOLOGICAL TESTS	565
Tests of General Intellectual Ability	565
Tests to Measure the Ability to Abstract	567
Tests of Executive Function	568
Tests of Perceptual, Motor, and Perceptual-Motor Function	572
Tests of Verbal Functioning	573
Tests of Memory	573
Neuropsychological Test Batteries	576
OTHER TOOLS OF NEUROPSYCHOLOGICAL ASSESSMENT	580
MEET AN ASSESSMENT PROFESSIONAL Meet Dr. Jeanne P. Ryan	566
EVERYDAY PSYCHOMETRICS Medical Diagnostic Aids and Neuropsychological Assessment	581
CLOSE-UP A Typical In-Office Dementia Evaluation	583
SELF-ASSESSMENT	584
REFERENCES	584

## 15 Assessment, Careers, and Business 590

### CAREER CHOICE AND CAREER TRANSITION 590

The Structure of Vocational Interests 590

Measures of Interest 592

Measures of Ability and Aptitude 594

Measures of Personality 596

Other Measures 599

### SCREENING, SELECTION, CLASSIFICATION, AND PLACEMENT 601

The Résumé and the Letter of Application 602

The Application Form 602

Letters of Recommendation 602

Interviews 603

Portfolio Assessment 604

Performance Tests 604

Physical Tests 609

### COGNITIVE ABILITY, PRODUCTIVITY, AND MOTIVATION MEASURES 611

Measures of Cognitive Ability 611

Productivity 612

Motivation 613

### JOB SATISFACTION, ORGANIZATIONAL COMMITMENT, AND ORGANIZATIONAL CULTURE 617

Job Satisfaction 617

Organizational Commitment 618

Organizational Culture 619

### OTHER TOOLS OF ASSESSMENT FOR BUSINESS APPLICATIONS 619

Consumer Psychology 620

The Measurement of Attitudes 621

Surveys 623

Motivation Research Methods 625

### EVERYDAY PSYCHOMETRICS The Selection of Personnel for the Office of Strategic Services (OSS): Assessment and Psychometrics in Action 606

### MEET AN ASSESSMENT PROFESSIONAL Meet Dr. Jed Yalof 620

### SELF-ASSESSMENT 629

### REFERENCES 629

Name Index I-1

Glossary/Index I-22

Timeline T-1

# Preface

W

e are proud to welcome instructors of a measurement course in psychology to this tenth edition of *Psychological Testing and Assessment*. Thank you for the privilege of assisting in the exciting task of introducing the world of tests and measurement to your students. In this preface, we impart our vision for a measurement textbook, as well as the philosophy that has driven, and that continues to drive, the organization, content, writing style, and pedagogy of this book. We'll briefly look back at this book's heritage and discuss what is new and distinctive about this tenth edition. Of particular interest to instructors, this preface will overview the authors' general approach to the course content and distinguish how that approach differs from other measurement textbooks. For students who happen to be curious enough to read this preface (or ambitious enough to read it despite the fact that it was not assigned), we hope that your takeaway from it has to do with the authors' genuine dedication to making this book the far-and-away best available textbook for your measurement course.

## Our Vision for a Textbook on Psychological Testing and Assessment

First and foremost, let's get out there that the subject matter of this course is *psychological testing and assessment*—a fact that is contrary to the message conveyed by an array of would-be competitor books, all distinguished by their anachronistic “psychological testing” title. Of course we cover tests and testing, and no available textbook does it better or more comprehensively. But it behooves us to observe that we are now well into the twenty-first century and it has long been recognized that tests are only one tool of assessment. Psychological testing is a process that can be—perhaps reminiscent of those books with the same title—impersonal, noncreative, uninspired, routine, and even robotic in nature. By contrast, psychological assessment is a human, dynamic, custom, creative, and collaborative enterprise. These aspects of the distinction between *psychological testing* and *psychological assessment* are not trivial.

Paralleling important differences between our book's title and that of other books in this area are key differences in the way that the subject matter of the course is approached. In routine writing and through a variety of pedagogical tools, we attempt to draw students into the world of testing and assessment by *humanizing* the material. Our approach to the course material stands in stark contrast to the “by-the-numbers” approach of some of our competitors; the latter approach can easily alienate readers, prompting them to “tune out.” Let's briefly elaborate on this critical point.

Although most of our competitors begin by organizing their books with an outline that for the most part mimics our own—right down to the inclusion of the Statistics Refresher that we innovated some 30 years ago—the way that they cover that subject matter, and the pedagogical tools they rely on to assist student learning, bear only cosmetic resemblance to our approach. We take every opportunity to illustrate the course material by putting a human face to it, and by providing practical, “every day” examples of the principles and procedures at work. This approach differs in key ways from the approach of other books in the area, in which a “practical approach” may instead be equated with the intermingling of statistical or other exercises within every chapter of the book. Presumably, according to the latter vision, a textbook is a simultaneous delivery system for both course-related information and course-related exercises. Students are expected to read their textbooks until such time that their reading is interrupted by an exercise. After the completion of the exercise, students are expected to go back to the reading, but only until they happen upon another exercise. It is thus the norm to interrupt absorption in assigned reading on a relatively random (variable ratio) schedule in order to have students complete

general, one-size-fits-all exercises. Students using such a book are not encouraged to concentrate on assigned reading; they may even be tacitly encouraged to do the opposite. The emphasis given to students having to complete exercises scattered within readings seems especially misplaced when, as is often the case with such one-size-fits-all tasks, some of the exercises will be way too easy for students in some classes and way too difficult for students in others. This situation brings to mind our own experience with testing-related exercises being assigned to varied groups of introductory students.

For several years and through several editions, our textbook was published with a supplementary exercises workbook. After extensive feedback from many instructors, some of whom used our book in their classes and some of whom did not, we determined that matters related to the choice, content, and level of supplementary exercises were better left to individual instructors as opposed to textbook authors. In general, instructors preferred to assign their own supplementary exercises, which could be custom-designed for the needs of their particular students and the goals of their particular course. A workbook of exercises, complete with detailed, step-by-step, illustrated solutions of statistical and psychometric problems, was determined by us to add little value to our textbook and it is therefore no longer offered. What we learned, and what we now believe, is that there is great value to supplementary, ancillary exercises for students taking an introductory course in measurement. However, these exercises are of optimal use to the student when they are custom-designed (or selected) by the instructor based on factors such as the level and interest of the students in the class, and the students' in-class and out-of-class study schedule. To be clear, supplemental exercises randomly embedded in a textbook work, in our view, not to facilitate students' immersion and concentration in assigned reading, but to obliterate it.<sup>1</sup>

Given that decisions regarding supplementary exercises are best left to individual instructors, the difference between our own approach to the subject matter of the course and that of other approaches are even more profound. In this tenth edition, we have concentrated our attention and effort to crafting a textbook that will immerse and involve students in assigned readings and motivate them to engage in critical and generative thinking about what they have read. Contrast that vision with one in which author effort is divided between writing text and writing nonsupplementary exercises. Could the net result of the latter approach be a textbook that divides student attention between assigned readings and assigned (or unassigned) exercises? Seasoned instructors may concur with our view that most students will skip the intrusive and distracting exercises when they are not specifically assigned for completion by the instructor. In the case where the exercises *are* assigned, students may well skim the reading to complete the exercises.

No available textbook is more focused on being practical, timely, and “real-life” oriented than our book is. Further, no other textbook provides students in an introductory course with a more readable or more comprehensive account of how psychological tests and assessment-related procedures are used in practice. That has been the case for some 30 years and it most certainly is the case today. With that as background, let's briefly sum up some of our concerns with regard to certain members of the current community of “psychological testing” books.

Especially with regard to a textbook at the introductory level, what is critical is the breadth and depth of coverage of how tests and other tools of assessment are actually used in practice. Practice-level proficiency and hands-on experience are always nice, but may in some cases be too ambitious. For example, a practical approach to factor analysis in a textbook for an introductory measurement course need not equip the student to conduct a factor analysis.

---

1. We urge any instructors curious about this assertion to informally evaluate it by asking a student or two how they feel about the prospect of scattering statistical exercises in their assigned reading. If the assigned reading is at all immersive, the modal response may be something like “maddening.”

Rather, the coverage ideally provides the student with a sound grounding in what this widely used set of techniques are, as well as how and why they are used. Similarly a practical approach to test utility, as exemplified in Chapter 7, provides students with a sound grounding in what that construct is, as well as how and why it is applied in practice.

Of course when it comes to breadth and depth of coverage of how tests and other tools of assessment are actually used in practice, we have long been the standard by which other books are measured. Consider in this context a small sampling of what is new, timely, and relevant in this tenth edition. The subject of our Chapter 1 *Close-Up* is behavioral assessment using smart phones. The subject of our *Everyday Psychometrics* in Chapter 7 on utility is the utility of police use of body cameras.<sup>2</sup> Terrorism is a matter of worldwide concern and in Chapter 11, the professional profiled in our *Meet an Assessment Professional* feature is Colonel Rick Malone of the United States Army's Criminal Investigation Command. Dr. Malone shares some intriguing insights regarding his area of expertise: threat assessment. Much more about our vision for this textbook and its supplements, as well as more previews of what is new and exciting in this tenth edition, is presented in what follows.

## Organization

From the first edition of our book forward, we have organized the information to be presented into five major sections. Part I, *An Overview*, contains two chapters that do just that. Chapter 1 provides a comprehensive overview of the field, including some important definitional issues, a general description of tools of assessment, and related important information couched as answers to questions regarding the *who*, *what*, *why*, *how*, and *where* of the enterprise.

The foundation for the material to come continues to be laid in the second chapter of the overview, which deals with historical, cultural, and legal/ethical issues. The material presented in Chapter 2 clearly sets a context for everything that will follow. To relegate such material to the back of the book (as a kind of elective topic, much like the way that legal/ethical issues are treated in some books), or to ignore presentation of such material altogether (as most other books have done with regard to cultural issues in assessment), is, in our estimation, a grave error. “Back page infrequency” (to borrow an MMPI-2 term) is too often the norm, and relegation of this critically important information to the back pages of a textbook too often translates to a potential shortchanging of students with regard to key cultural, historical, and legal/ethical information. The importance of exposure early on to relevant historical, cultural, and legal/ethical issues cannot be overemphasized. This exposure sets a context for succeeding coverage of psychometrics and creates an essential lens through which to view and process such material.

Part II, *The Science of Psychological Measurement*, contains Chapters 3 through 8. These six chapters were designed to build—logically and sequentially—on the student's knowledge of psychometric principles. Part II begins with a chapter reviewing basic statistical principles and ends with a chapter on test construction. In between, there is extensive discussion of assumptions inherent in the enterprise, the elements of good test construction, as well as the concepts of norms, correlation, inference, reliability, and validity. All of the measurement

---

2. This essay is an informative and timely discussion of the utility of police-worn body cameras in reducing use-of-force complaints. Parenthetically, let's share our view that the concept of *utility* seems lost in, or at least given inadequate coverage in other measurement books. It seems that we may have caught many of those “psychological testing” books off-guard by devoting a chapter to this construct beginning with our seventh edition—this at a time when *utility* was not even an indexed term in most of them. Attempts to compensate have ranged from doing nothing at all to doing near nothing at all by equating “utility” with “validity.” For the record, although utility is related to validity, much as reliability is related to validity, we believe it is misleading to even intimate that “utility” and “validity” are synonymous.

textbooks that came before us were written based on the assumption that every student taking the course was up to speed on all of the statistical concepts that would be necessary to build on learning about psychometrics. In theory, at least, there was no reason not to assume this previous knowledge; statistics was a prerequisite to taking the course. In practice, a different picture emerged. It was simply not the case that all students were adequately and equally prepared to begin learning statistics-based measurement concepts. Our remedy for this problem, some 30 years ago, was to include a “Statistics Refresher” chapter early on, just prior to building on students’ statistics-based knowledge. The rest, as they say, is history...

Our book forever changed for the better the way the measurement course was taught and the way all subsequent textbooks for the course would be written. Our unique coverage of the assessment of intelligence and personality, as well as our coverage of assessment for various applications (ranging from neuropsychological to business and organizational applications), made relics of the typical “psychological testing” course outline as it existed prior to the publication of our first edition in 1988.

In our seventh edition, in response to increasing general interest in test utility, we added a chapter on this important construct right after our chapters on the constructs of reliability and validity. Let’s note here that topics such as utility and utility analysis can get extremely complicated. However, we have never shied away from the presentation of complicated subject matter. For example, we were the first introductory textbook to present detailed information related to factor analysis. As more commercial publishers and other test users have adopted the use of item response theory (IRT) in test construction, our coverage of IRT has kept pace. As more test reviews have begun to evaluate tests not only in terms of variables such as reliability and validity but in terms of *utility*, we saw a need for the inclusion of a chapter on that topic.

Of course, no matter how “difficult” the concepts we present are, we never for a moment lose sight of the appropriate level of presentation. This book is designed for students taking a first course in psychological testing and assessment. Our objective in presenting material on methods such as IRT and utility analysis is simply to acquaint the introductory student with these techniques. The depth of the presentation in these and other areas has always been guided and informed by extensive reviews from a geographically diverse sampling of instructors who teach measurement courses. For users of this textbook, what currently tends to be required is a conceptual understanding of commonly used IRT methods. We believe our presentation of this material effectively conveys such an understanding. Moreover, it does so without unnecessarily burdening students with level-inappropriate formulas and calculations.

Part III of this book, *The Assessment of Abilities and Aptitudes*, contains two chapters, one on intelligence and its assessment, and the other on assessment in schools and other educational settings. In past editions of this book, two chapters were devoted to the assessment of intelligence. To understand why, it is instructive to consider what the coverage of intelligence testing looked like in the then available introductory measurement textbooks three decades ago. While the books all covered tests of intelligence, they devoted little or no attention to defining and discussing the construct of intelligence. We called attention to this problem and attempted to remedy it by differentiating our book with a chapter devoted to imparting a conceptual understanding of intelligence. Although revolutionary at the time, the logic of our approach had widespread appeal. Before long, the typical “psychological testing” course of the 1980s was being restructured to include conceptual discussions of concepts such as “intelligence” and “personality” before proceeding to discuss their measurement. The “psychological testing” textbooks of the day also followed our lead. And so, to the present day, two-chapter-coverage of the assessment of intelligence (with the first chapter providing a discussion of the construct of intelligence) has become the norm.

In retrospect, it seems reasonable to conclude that our addition of a chapter on the nature of intelligence, much like our addition of a statistics refresher, did more than remedy a serious drawback in existing measurement textbooks; it forever revolutionized the way that the

measurement course was taught in classrooms around the world. It did this first of all by making the teaching of the course more logical. This is so because the logic of our guiding principle—fully define and discuss the psychological construct being measured before discussing its measurement—had wide appeal. In our first edition, we also extended that logic to the discussion of the measurement of other psychological constructs such as personality. Another benefit we saw in adding the conceptual coverage was that such coverage would serve to “humanize” the content. After all, “Binet” was more than just the name of a psychological test; it was the name of a living, breathing person.

Also, since our first edition, we have revolutionized textbook coverage of psychological tests—this by a philosophy of “less is more” when it comes to such coverage. Back in the 1980s, the “psychological testing” books of the day had elements reminiscent of *Tests in Print*. They provided reliability, validity, and related psychometric data on dozens of psychological tests. But we raised the question, “Why duplicate in a textbook information about dozens of tests that is readily available from reference sources?” We further resolved to limit detailed coverage of psychological tests to a handful of representative tests. Once again, the simple logic of our approach had widespread appeal, and other textbooks in the area—both then, and to the present day—all followed suit.

There is another trend in textbook coverage of the measurement course that also figured prominently in our decision to cover the assessment of intelligence in a single chapter. This trend has to do with the widespread availability of online resources to supplement coverage of a specific topic. We have long taken advantage of this fact by making available various supplementary materials online to our readers, or by supplying links to such materials.

Some three decades after we revolutionized the organization of textbook coverage of the measurement course in so many significant ways, it was time to re-evaluate whether two chapters to cover the subject of intelligence assessment was still necessary. We gave thoughtful consideration to this question and sought-out the opinion of trusted colleagues. In the end, we determined that coverage of the construct and assessment of intelligence could be accomplished in a single chapter. And so, in the interest of streamlining this book in length, Chapter 9 in the ninth edition incorporated text formerly in Chapters 9 and 10 of the eighth edition. This combined chapter was maintained in the tenth edition of the textbook.

Part IV, *The Assessment of Personality*, contains two chapters, which respectively overview how personality assessments are conducted, and the various methods used.

Part V, *Testing and Assessment in Action*, is designed to convey to students a sense of how a sampling of tests and other tools of assessment are actually used in clinical, counseling, business, and other settings.

## Content

In addition to a logical organization that sequentially builds on student learning, we view *content* selection as another key element of our appeal. The multifaceted nature and complexity of the discipline affords textbook authors wide latitude in terms of what material to elaborate on, what material to ignore, and what material to highlight, exemplify, or illustrate. In selecting content to be covered for chapters, the primary question for us was most typically “What do students need to know?” So, for example, since the publication of previous editions of this book, the field of educational evaluation has been greatly influenced by the widespread implementation of the *Common Core Standards*. Accordingly, we take cognizance of these changes in the K-through-12 education landscape and their implications for evaluation in education. Students of educational assessment *need* to know about the *Common Core Standards* and relevant coverage of these standards can be found in this tenth edition in our chapter on educational assessment.

While due consideration is given to creating content that students need to know, consideration is also given to relevant topics that will engage interest and serve as stimuli for

critical or generative thinking. In the area of neuropsychological assessment, for example, the topic of Alzheimer's disease is one that generates a great deal of interest. Most students have seen articles or feature stories in the popular media that review the signs and symptoms of this disease. However, while students are aware that such patients are typically referred to a neurologist for formal diagnosis, many questions remain about how a diagnosis of Alzheimer's disease is clinically made. The *Close-Up* in our chapter on neuropsychological assessment addresses those frequently asked questions. It was guest-authored by an experienced neurologist and written especially for students of psychological assessment reading this textbook.

Let's note here that in this tenth edition, more than in any previous edition of this textbook, we have drawn on the firsthand knowledge of psychological assessment experts from around the world. Specifically, we have asked these experts to guest-author brief essays in the form of *Close-Up*, *Everyday Psychometrics*, or *Meet an Assessment Professional* features. For example, in one of our chapters that deal with personality assessment, two experts on primate behavior (including one who is currently working at Dian Fossey's research center in Karisoke, in Rwanda) prepared an essay on evaluating the personality of gorillas. Written especially for us, this *Close-Up* makes an informative contribution to the literature on cross-species personality assessment. In our chapter on test construction, an Australian team of behavioral scientists guest-authored a *Close-Up* entitled "Adapting Tools of Assessment for Use with Specific Cultural Groups." This essay recounts some of the intriguing culture-related challenges inherent in the psychological assessment of clients from the Aboriginal community.

Sensitivity to cultural issues in psychological testing and assessment is essential, and this textbook has long set the standard for coverage of such issues. Coverage of cultural issues begins in earnest in Chapter 2, where we define culture and overview the importance of cultural considerations in everything from test development to standards of evaluation. Then, much like an identifiable musical theme that recurs throughout a symphony, echoes of the importance of culture repeat in various chapters throughout this book. For example, the echo is heard in Chapter 4 where, among other things, we continue a long tradition of acquainting students with the "do's and don'ts" of culturally informed assessment. In Chapter 13, our chapter on assessment in clinical and counseling settings, there is a discussion of acculturation and culture as these issues pertain to clinical assessment. Also in that chapter, students will find a thought-provoking *Close-Up* entitled, "PTSD in Veterans and the Idealized Culture of Warrior Masculinity." Guest-authored especially for us by Duncan M. Shields, this timely contribution to the clinical literature sheds light on the diagnosis and treatment of post-traumatic stress disorder (PTSD) from a new and novel, cultural perspective.

In addition to standard-setting content related to cultural issues, mention must also be made of our leadership role with respect to coverage of historical and legal/ethical aspects of measurement in psychology. Our own appreciation for the importance of history is emphasized by the listing of noteworthy historical events that is set within the front and back covers of this textbook. As such, readers may be greeted with some aspect of the history of the enterprise on every occasion that they open the book. Although historical vignettes are distributed throughout the book to help set a context or advance understanding, formal coverage begins in Chapter 2. Important historical aspects of testing and assessment may also be found in *Close-Ups*. See, for example, the fascinating account of the controversial career of Henry Goddard found in Chapter 2. In a *Close-Up* in Chapter 15, students will discover what contemporary assessment professionals can learn from World War II-vintage assessment data collected by the Office of Strategic Services (OSS). In this engrossing essay, iconic data meets contemporary data analytic methods with brilliant new insights as a result. This *Close-Up* was guest-authored by Mark F. Lenzenweger, who is a State University of New York (SUNY) Distinguished Professor in the Department of Psychology at the State University of New York at Binghamton.

Much like content pertaining to relevant historical and culture-related material, our discussion of legal-ethical issues, from our first edition through to the present day, has been standard-setting.

Discussion of legal and ethical issues as they apply to psychological testing and assessment provides students not only with context essential for understanding psychometric principles and practice, but another lens through which to filter understanding of tests and measurement. In the first edition, while we got the addition of this pioneering content right, we could have done a better job in terms of placement. In retrospect, the first edition would have benefitted from the discussion of such issues much earlier than the last chapter. But in response to the many compelling arguments reviewers and users of that book, discussion of legal/ethical issues was prioritized in Chapter 2 by the time that our second edition was published. The move helped ensure that students were properly equipped to appreciate the role of legal and ethical issues in the many varied settings in which psychological testing and assessment takes place.

Another element of our vision for the content of this book has to do with the art program; that is, the photos, drawings, and other types of illustrations used in a textbook. Before the publication of our ground-breaking first edition, what passed for an art program in the available “psychological testing” textbooks were some number-intensive graphs and tables, as well as photos of test kits or test materials. In general, photos and other illustrations seemed to be inserted more to break up text than to complement it. For us, the art program is an important element of a textbook, not a device for pacing. Illustrations can help draw students into the narrative, and then reinforce learning by solidifying meaningful visual associations to the written words. Our figures and graphics bring concepts to life. Photos can be powerful tools to stir the imagination. See, for example, the photo of Army recruits being tested in Chapter 1, or the photo of Ellis Island immigrants being tested in Chapter 2. Photos can bring to life and “humanize” the findings of measurement-related research. See, for example, the photo in Chapter 3 regarding the study that examined the relationship between grades and cell phone use in class. Photos of many past and present luminaries in the field (such as John Exner, Jr. and Ralph Reitan), and photos accompanying the persons featured in our *Meet an Assessment Professional* boxes all serve to breathe life into their respective accounts and descriptions.

In the world of textbooks, photos such as the sampling of the ones described here may not seem very revolutionary. However, in the world of *measurement* textbooks, our innovative art program has been and remains quite revolutionary. One factor that has always distinguished us from other books in this area is the extent to which we have tried to “humanize” the course subject matter; the art program is just another element of this textbook pressed into the service of that objective.

**“Humanization” of Content** This tenth edition was conceived with a commitment to continuing our three-decade tradition of exemplary organization, exceptional writing, timely content, and solid pedagogy. Equally important was our desire to spare no effort in making this book as readable and as involving for students as it could possibly be. Our “secret sauce” in accomplishing this is, at this point, not much of a secret. We have the highest respect for the students for whom this book is written. We try to show that respect by never underestimating their capacity to become immersed in course-relevant narratives that are presented clearly and straightforwardly. With the goal of further drawing the student into the subject matter, we make every effort possible to “humanize” the presentation of topics covered. So, what does “humanization” in this context actually mean?

While other authors in this discipline impress us as blindly intent on viewing the field as Greek letters to be understood and formulas to be memorized, we view an introduction to the field to be about *people* as much as anything else. Students are more motivated to learn this material when they can place it in a human context. Many psychology students simply do not respond well to endless presentations of psychometric concepts and formulas. In our opinion, to *not* bring a human face to the field of psychological testing and assessment, is to risk perpetuating all of those unpleasant (and now unfair) rumors about the course that first began circulating long before the time that the senior author himself was an undergraduate.

Our effort to humanize the material is evident in the various ways we have tried to bring a face (if not a helping voice) to the material. The inclusion of *Meet an Assessment Professional* is a means toward that end, as it quite literally “brings a face” to the enterprise. Our inclusion of interesting biographical facts on historical figures in assessment is also representative of efforts to humanize the material. Consider in this context the photo and brief biographical statement of MMPI-2 senior author James Butcher in Chapter 11 (p. 426). Whether through such images of historical personages or by other means, our objective has been made to truly involve students via intriguing, real-life illustrations of the material being discussed. See, for example, the discussion of life-or-death psychological assessment and the ethical issues involved in the *Close-Up* feature of Chapter 2. Or check out the candid “confessions” of a behavior rater in the *Everyday Psychometrics* feature in Chapter 12.

So how has our “humanization” of the material in this discipline been received by some of its more “hard core” and “old school” practitioners? Very well, thank you—at least from all that we have heard, and the dozens of reviews that we have read over the years. What stands out prominently in the mind of the senior author (RJC) was the reaction of one particular psychometrician whom I happened to meet at an APA convention not long after the first edition of this text was published. Lee J. Cronbach was quite animated as he shared with me his delight with the book, and how refreshingly different he thought that it was from anything comparable that had been published. I was so grateful to Lee for his encouragement, and felt so uplifted by that meeting, that I subsequently requested a photo from Lee for use in the second edition. The photo he sent was indeed published in the second edition of this book—this despite the fact that at that time, Lee had a measurement book that could be viewed as a direct competitor to ours. Regardless, I felt it was important not only to acknowledge Lee’s esteemed place in measurement history, but to express my sincere gratitude in this way for his kind, inspiring, and motivating words, as well as for what I perceived as his most valued “seal of approval.”

### **Pedagogical Tools**

The objective of incorporating timely, relevant, and intriguing illustrations of assessment-related material is furthered by several *pedagogical tools* built into the text. One pedagogical tool we created several editions ago is *Everyday Psychometrics*. In each chapter of the book, relevant, practical, and “everyday” examples of the material being discussed are highlighted in an *Everyday Psychometrics* box. For example, in the *Everyday Psychometrics* presented in Chapter 1 (“Everyday Accommodations”), students will be introduced to accommodations made in the testing of persons with handicapping conditions. In Chapter 4, the *Everyday Psychometrics* feature (“Putting Tests to the Test”) equips students with a working overview of the variables they need to be thinking about when reading about a test and evaluating how satisfactory the test really is for a particular purpose. In Chapter 5, the subject of the *Everyday Psychometrics* is how the method used to estimate diagnostic reliability may affect the obtained estimate of reliability.

A pedagogical tool called *Meet an Assessment Professional* was first introduced in the seventh edition. This feature provides a forum through which everyday users of psychological tests from various fields can share insights, experiences, and advice with students. The result is that in each chapter of this book, students are introduced to a different test user and provided with an intriguing glimpse of their professional life—this in the form of a *Meet an Assessment Professional (MAP)* essay. For example, in Chapter 4, students will meet a team of test users, Drs. Steve Julius and Howard Atlas, who have pressed psychometric knowledge into the service of professional sports. They provide a unique and fascinating account of how application of their knowledge of was used to improve the on-court of achievement of the Chicago Bulls. A MAP essay from Stephen Finn, the well-known proponent of therapeutic assessment is presented in Chapter 13. Among the many MAP essays in this edition are essays from two mental-health professionals serving in the military.

Dr. Alan Ogle introduces readers to aspects of the work of an Air Force psychologist in Chapter 1. In Chapter 11, army psychiatrist Dr. Rick Malone shares his expertise in the area of threat assessment. The senior author of an oft-cited meta-analysis that was published in *Psychological Bulletin* shares her insights on meta-analytic methods in Chapter 3, while a psychiatrist who specializes in cultural issues introduces himself to students in Chapter 2.

Our use of the pedagogical tool referred to as a “*Close-Up*” is reserved for more in-depth and detailed consideration of specific topics related to those under discussion. The *Close-Up* in our chapter on test construction, for example, acquaints readers with the trials and tribulations of test developers working to create a test to measure asexuality. The *Close-Up* in one of our chapters on personality assessment raises the intriguing question of whether it is meaningful to speak of general (*g*) and specific (*s*) factors in the diagnosis of personality disorders.

There are other pedagogical tools that readers (as well as other textbook authors) may take for granted—but we do not. Consider, in this context, the various tables and figures found in every chapter. In addition to their more traditional use, we view tables as space-saving devices in which a lot of information may be presented. For example, in the first chapter alone, tables are used to provide succinct but meaningful comparisons between the terms *testing* and *assessment*, the *pros* and *cons* of computer-assisted psychological assessment, and the *pros* and *cons* of using various sources of information about tests.

*Critical thinking* may be defined as “the active employment of judgment capabilities and evaluative skills in the thought process” (Cohen, 1994, p. 12). *Generative thinking* may be defined as “the goal-oriented intellectual production of new or creative ideas” (Cohen, 1994, p. 13). The exercise of both of these processes, we believe, helps optimize one’s chances for success in the academic world as well as in more applied pursuits. In the early editions of this textbook, questions designed to stimulate critical and generative thinking were raised “the old-fashioned way.” That is, they were right in the text, and usually part of a paragraph. Acting on the advice of reviewers, we made this special feature of our writing even more special beginning with the sixth edition of this book; we raised these critical thinking questions in the margins with a *Just Think* heading. Perhaps with some encouragement from their instructors, motivated students will, in fact, give thoughtful consideration to these (critical and generative thought-provoking) *Just Think* questions.

In addition to critical thinking and generative thinking questions called out in the text, other pedagogical aids in this book include original cartoons created by the authors, original illustrations created by the authors (including the model of memory in Chapter 14), and original acronyms created by the authors.<sup>3</sup> Each chapter ends with a *Self-Assessment* feature that students may use to test themselves with respect to key terms and concepts presented in the text.



connect

The tenth edition of *Psychological Testing and Assessment* is now available online with Connect, McGraw-Hill Education’s integrated assignment and assessment platform. Connect also offers SmartBook for the new edition, which is the first adaptive reading experience proven to improve grades and help students study more effectively. All of the title’s website and ancillary content is also available through Connect, including:

- An Instructor’s Manual for each chapter.
- A full Test Bank of multiple choice questions that test students on central concepts and ideas in each chapter.
- Lecture Slides for instructor use in class.

---

3. By the way, our use of the French word for black (*noir*) as an acronym for levels of measurement (nominal, ordinal, interval, and ratio) now appears in other textbooks.

Cohen, R. J. (1994). *Psychology & adjustment: Values, culture, and change*. Allyn & Bacon.



# connect<sup>®</sup>

## Instructors: Student Success Starts with You

### Tools to enhance your unique voice

Want to build your own course? No problem. Prefer to use our turnkey, prebuilt course? Easy. Want to make changes throughout the semester? Sure. And you'll save time with Connect's auto-grading too.

# 65%

Less Time Grading



Laptop: McGraw Hill; Woman/dog: George Doyle/Getty Images

### Study made personal

Incorporate adaptive study resources like SmartBook<sup>®</sup> 2.0 into your course and help your students be better prepared in less time. Learn more about the powerful personalized learning experience available in SmartBook 2.0 at [www.mheducation.com/highered/connect/smartbook](http://www.mheducation.com/highered/connect/smartbook)

### Affordable solutions, added value

Make technology work for you with LMS integration for single sign-on access, mobile access to the digital textbook, and reports to quickly show you how each of your students is doing. And with our Inclusive Access program you can provide all these tools at a discount to your students. Ask your McGraw Hill representative for more information.

Padlock: Jobalou/Getty Images

### Solutions for your challenges

A product isn't a solution. Real solutions are affordable, reliable, and come with training and ongoing support when you need it and how you want it. Visit [www.supportateverystep.com](http://www.supportateverystep.com) for videos and resources both you and your students can use throughout the semester.

Checkmark: Jobalou/Getty Images

## Students: Get Learning that Fits You

### Effective tools for efficient studying

Connect is designed to make you more productive with simple, flexible, intuitive tools that maximize your study time and meet your individual learning needs. Get learning that works for you with Connect.

### Study anytime, anywhere

Download the free ReadAnywhere app and access your online eBook or SmartBook 2.0 assignments when it's convenient, even if you're offline. And since the app automatically syncs with your eBook and SmartBook 2.0 assignments in Connect, all of your work is available every time you open it. Find out more at [www.mheducation.com/readanywhere](http://www.mheducation.com/readanywhere)

*"I really liked this app—it made it easy to study when you don't have your textbook in front of you."*

- Jordan Cunningham,  
Eastern Washington University



Calendar: owattaphotos/Getty Images

### Everything you need in one place

Your Connect course has everything you need—whether reading on your digital eBook or completing assignments for class, Connect makes it easy to get your work done.

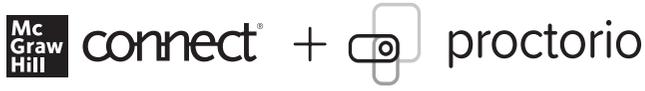
### Learning for everyone

McGraw Hill works directly with Accessibility Services Departments and faculty to meet the learning needs of all students. Please contact your Accessibility Services Office and ask them to email [accessibility@mheducation.com](mailto:accessibility@mheducation.com), or visit [www.mheducation.com/about/accessibility](http://www.mheducation.com/about/accessibility) for more information.

Top: Jenner Images/Getty Images, Left: Hero Images/Getty Images, Right: Hero Images/Getty Images



## Remote Proctoring & Browser-Locking Capabilities



New remote proctoring and browser-locking capabilities, hosted by Proctorio within Connect, provide control of the assessment environment by enabling security options and verifying the identity of the student.

Seamlessly integrated within Connect, these services allow instructors to control students' assessment experience by restricting browser activity, recording students' activity, and verifying students are doing their own work.

Instant and detailed reporting gives instructors an at-a-glance view of potential academic integrity concerns, thereby avoiding personal bias and supporting evidence-based claims.

### *Writing Assignment*

Available within McGraw-Hill Connect<sup>®</sup> and McGraw-Hill Connect<sup>®</sup> Master, the Writing Assignment tool delivers a learning experience to help students improve their written communication skills and conceptual understanding. As an instructor you can assign, monitor, grade, and provide feedback on writing more efficiently and effectively.

### *Writing Style*

What type of *writing style* or author *voice* works best with students being introduced to the field of psychological testing and assessment? Instructors familiar with the many measurement books that have come (and gone) may agree with us that the “voice” of too many authors in this area might best be characterized as humorless and academic to the point of arrogance or pomposity. Students do not tend to respond well to textbooks written in such styles, and their eagerness and willingness to spend study time with these authors (and even their satisfaction with the course as a whole) may easily suffer as a consequence.

In a writing style that could be characterized as somewhat informal and—to the extent possible, given the medium and particular subject being covered—“conversational,” we have made every effort to convey the material to be presented as clearly as humanly possible. In practice, this means:

- keeping the vocabulary of the presentation appropriate (without ever “dumbing-down” or trivializing the material);
- presenting so-called difficult material in step-by-step fashion where appropriate, and always preparing students for its presentation by placing it in an understandable context;
- italicizing the first use of a key word or phrase and then bolding it when a formal definition is given;
- providing a relatively large glossary of terms to which students can refer;
- supplementing material where appropriate with visual aids, tables, or other illustrations.
- supplementing material where appropriate with intriguing historical facts (as in the Chapter 12 material on projectives and the projective test created by B. F. Skinner);
- incorporating timely, relevant, and intriguing illustrations of assessment-related material in the text as well as in the online materials.

In addition, we have interspersed some elements of humor in various forms (original cartoons, illustrations, and vignettes) throughout the text. The judicious use of humor to engage and maintain student interest is something of a novelty among measurement textbooks. Where else would one turn for pedagogy that employs an example involving a bimodal distribution of test scores from a new trade school called *The Home Study School of Elvis Presley Impersonators*? As readers learn about face validity, they discover why it “gets no respect” and how it has been characterized as “the Rodney Dangerfield of psychometric variables.” Numerous other illustrations could be cited here. But let’s reserve those smiles as a pleasant surprise when readers happen to come upon them.

## Acknowledgments

Thanks to the members of the academic community who have wholeheartedly placed their confidence in this book through all or part of its tenth-edition life-cycle to date. Your trust in our ability to help your students navigate the complex world of measurement in psychology is a source of inspiration to us. We appreciate the privilege of assisting you in the education and professional growth of your students, and we will never take that privilege for granted.

Every edition of this book has begun with blueprinting designed with the singular objective of making this book far-and-away best in the field of available textbooks in terms of organization, content, pedagogy, and writing. Helping the authors to meet that objective were developmental editor Erin Guendelsberger and project supervisor Jamie Laferrera along with a number of guest contributors who graciously gave of their time, talent, and expertise. To be the all-around best textbook in a particular subject area takes, as they say, “a village.” On behalf of the authors, a hearty “thank you” is due to many “villagers” in the academic and professional community who wrote or reviewed something for this book, or otherwise contributed to it. First and foremost, thank you to all of the following people who wrote essays designed to enhance and enrich the student experience of the course work. In order of appearance of the tenth edition chapter that their essay appeared in, we say thanks to the following contributors of guest-authored *Meet an Assessment Professional*, *Everyday Psychometrics*, or *Close-Up*:

Alan, D. Ogle of the 559th Medical Group, Military Training Consult Service of the United States Air Force;

Dror Ben-Zeev of the Department of Psychiatry of the Geisel School of Medicine at Dartmouth;  
Neil Krishan Aggarwal of the New York State Psychiatric Institute;

Joni L. Mihura of the Department of Psychology of the University of Toledo;

Michael Chmielewski of the Department of Psychology of Southern Methodist University;

Jason M. Chin of the University of Toronto Faculty of Law;

Ilona M. McNeill of the University of Melbourne (Australia);

Patrick D. Dunlop of the University of Western Australia;

Delphine Courvoisier of Beau-Séjour Hospital, Geneva, Switzerland;

Alex Sutherland of RAND Europe, Cambridge, United Kingdom;

Barak Ariel of the Institute of Criminology of the University of Cambridge (United Kingdom);

Lori A. Brotto of the Department of Gynaecology of the University of British Columbia;

Morag Yule of the Department of Gynaecology of the University of British Columbia;

Sivasankaran Balaratnasingam of the School of Psychiatry and Clinical Neurosciences of the University of Western Australia;

Zaza Lyons of the School of Psychiatry and Clinical Neurosciences of the University of Western Australia;

Aleksander Janca of the School of Psychiatry and Clinical Neurosciences of the University of Western Australia;

Yuanbo Gu of the School of Psychology of Shaanxi Normal University (China);

Ning He of the School of Psychology of Shaanxi Normal University (China);

Xuqun You of the School of Psychology of Shaanxi Normal University (China);

Chengting Ju of the School of Psychology of Shaanxi Normal University (China);

Rick Malone of the U.S. Army Criminal Investigation Command, Quantico, VA;

Winnie Eckardt of The Dian Fossey Gorilla Fund International, Atlanta, GA;

Alexander Weiss of the Department of Psychology of the University of Edinburgh (UK);

Monica Webb Hooper of the Case Comprehensive Cancer Center at Case Western Reserve University;

Carla Sharp of the Department of Psychology at the University of Houston (TX);

Liliana B. Sousa of the Faculty of Psychology and Educational Sciences of the University of Coimbra (Portugal);

Duncan M. Shields of the Faculty of Medicine of the University of British Columbia;

Eric Kramer of Medical Specialists of the Palm Beaches, (Neurology), Atlantis, Florida;

Jed Yalof of the Department of Graduate Psychology of Immaculata University;

Mark F. Lenzenweger of the Department of Psychology of the State University of New York at Binghamton;

Jessica Klein of the Department of Psychology of the University of Florida (Gainesville);

Anna Taylor of the Department of Psychology of Illinois State University;

Suzanne Swagerman of the Department of Biological Psychology of Vrije Universiteit (VU), Amsterdam, The Netherlands;

Eco J.C. de Geus of the Department of Biological Psychology of Vrije Universiteit (VU), Amsterdam, The Netherlands;

Kees-Jan Kan of the Department of Biological Psychology of Vrije Universiteit (VU), Amsterdam, The Netherlands;

Dorret I. Boomsma of the Department of Biological Psychology of Vrije Universiteit (VU), Amsterdam, The Netherlands;

Faith Miller of the Department of Educational Psychology of the University of Minnesota; and, Daniel Teichman formerly of the Department of Computer and Information Science and Engineering of the University of Florida (Gainesville).

For their enduring contribution to this and previous editions of this book, we thank Dr. Jennifer Kisamore for her work on the original version of our chapter on test utility, and Dr. Bryce Reeve who wrote a Meet an Assessment Professional essay. Thanks to the many assessment professionals who, whether in a past or the current edition, took the time to introduce students to what they do. For being a potential source of inspiration to the students who they “met” in these pages, we thank the following assessment professionals: Dr. Rebecca Anderson, Dr. Howard W. Atlas, Dr. Scott Birkeland, Dr. Anthony Bram, Dr. Stephen Finn, Dr. Chris Gee, Dr. Joel Goldberg, Ms. Eliane Hack, Dr. Steve Julius, Dr. Nathaniel V. Mohatt, Dr. Barbara C. Pavlo, Dr. Jeanne P. Ryan, Dr. Adam Shoemaker, Dr. Benoit Verdon, Dr. Erik Viirre, and Dr. Eric A. Zillmer. Thanks also to Dr. John Garruto for his informative contribution to Chapter 10.

While thanking all who contributed in many varied ways, we remind readers that the present authorship team takes sole responsibility for any possible errors that may have somehow found their way into this tenth edition.

## Meet the Authors

Ronald Jay Cohen, Ph.D., ABPP, ABAP, is a Diplomate of the American Board of Professional Psychology in Clinical Psychology, and a Diplomate of the American Board of Assessment Psychology. He is licensed to practice psychology in New York and Florida, and a “scientist-practitioner” and “scholar-professional” in the finest traditions of each of those terms. During a long and gratifying professional career in which he has published numerous journal articles and books, Dr. Cohen has had the privilege of personally working alongside some of the luminaries in the field of psychological assessment, including David Wechsler (while Cohen was a clinical psychology intern at Bellevue Psychiatric Hospital in New York City) and Doug Bray (while working as an assessor for AT&T in its Management Progress Study). After serving his clinical psychology internship at Bellevue, Dr. Cohen was appointed Senior Psychologist there, and his clinical duties entailed not only psychological assessment but the supervision and training of others in this enterprise. Subsequently, as an independent practitioner in the New York City area, Dr. Cohen taught various courses at local universities on an adjunct basis, including undergraduate and graduate courses in psychological assessment. Asked by a colleague to conduct a qualitative research study for an advertising agency, Dr. Cohen would quickly become a sought-after qualitative research consultant with a client list of major companies and organizations—among them Paramount Pictures, Columbia Pictures, NBC Television, the Campbell Soup Company, Educational Testing Service, and the College Board. Dr. Cohen’s approach to qualitative research, referred to by him as *dimensional qualitative research*, has been emulated and written about by qualitative researchers around the world. Dr. Cohen is a sought-after speaker and has delivered invited addresses at the Sorbonne in Paris, Peking University in Beijing, and numerous other universities throughout the world. It was Dr. Cohen’s work in the area of qualitative assessment that led him to found the scholarly journal *Psychology & Marketing*. Since the publication of the journal’s first issue in 1984, Dr. Cohen has served as its Editor-in-Chief.

W. Joel Schneider, Ph.D., is Associate Professor of Counseling and School Psychology in the Department of Psychological Studies in Education at Temple University in Philadelphia. He completed his doctorate in clinical psychology at Texas A&M University. Dr. Schneider spent 15 years on the faculty at Illinois State University before joining the faculty at Temple. He is lead author of the second edition of the 2018 book, *Essentials of Assessment Report Writing*. He regularly teaches graduate-level courses in assessment and oversees clinic-based practicum students. Dr. Schneider has participated as an examiner in the standardization of several psychological tests, including the RIAS, CASE, and CASE-R. His primary research interests are psychological assessment of cognitive abilities and personality, psychometrics, statistics, and research methods, and psychotherapy with individuals, groups, couples, and families. Dr. Schneider has been involved in a number of research projects funded by the Pennsylvania Department of Education and the U.S. Department of Education. His work may be found in the *Archives of Clinical Neuropsychology*, *Psychological Methods*, *Journal of Psychoeducational Assessment*, *Applied Neuropsychology*, and *Best Practices in School Psychology*. He served as a test reviewer for the *Mental Measurements Yearbook* and currently serves as an Associate Editor for *Journal of Psychoeducational Assessment*. He is also an editorial board member for *Journal of Intelligence* and *Journal of School Psychology*.

Renée M. Tobin, Ph.D., is Professor and Chair of the Department of Psychological Studies in Education at Temple University in Philadelphia. She completed her master's degree in social psychology and her doctorate in school psychology at Texas A&M University. Dr. Tobin spent 15 years on the faculty at Illinois State University before joining the faculty at Temple. She is co-author of the 2015 book, *DSM-5 Diagnosis in the Schools*. She regularly teaches graduate-level courses in assessment, counseling, and consultation. She served as an examiner for the standardization of several psychological tests, including the RIAS, CASE, and CASE-R. Her primary research interests center broadly on personality and social development. Dr. Tobin has extensive experience conducting mixed-methods research with children, adolescents, young adults, and their families, particularly among diverse populations in various school contexts. She has been involved in a number of program evaluation projects since 2010, which include serving as co-leader of the evaluation team for the Livingston County Children's Network (funded by the Illinois Children's Healthcare Foundation) and as coordinator of the continuous quality improvement team for the Champaign Area Relationship Education for Youth (CARE4U) grant program (funded by the U.S. Department of Health and Human Services). Her work may be found in the *Journal of Personality and Social Psychology*, *Journal of Personality*, *Psychological Science*, *School Psychology Quarterly*, and *Best Practices in School Psychology*. She served as a test reviewer for the *Mental Measurements Yearbook* and served as an Associate Editor for *Journal of Psychoeducational Assessment* for over 10 years. She is currently an editorial board member for *Journal of School Psychology*.

### And on a Personal Note . . .

I think back to the time when we were just wrapping up work on the sixth edition of this book. At that time, I received the unexpected and most painful news that my mother had suffered a massive and fatal stroke. It is impossible to express the sense of sadness and loss experienced by myself, my brother, and my sister, as well as the countless other people who knew this gentle, loving, and much-loved person. To this day, we continue to miss her counsel, her sense of humor, and just knowing that she's there for us. We continue to miss her genuine exhilaration, which in turn exhilarated us, and the image of her welcoming, outstretched arms whenever we came to visit. Her children were her life, and the memory of her smiling face, making each of us feel so special, survives as a private source of peace and comfort for us all. She always kept a copy of this book proudly displayed on her coffee table, and I am very sorry that a copy of more recent editions did not make it to that most special place. My dedication of this book is one small way I can meaningfully acknowledge her contribution, as well as that of my beloved, deceased father, to my personal growth. As in the sixth edition, I am using my parents' wedding photo in the dedication. They were so good together in life. And so there Mom is, reunited with Dad. Now, that is something that would make her very happy.

As the reader might imagine, given the depth and breadth of the material covered in this textbook, it requires great diligence and effort to create and periodically re-create an instructional tool such as this that is timely, informative, and readable. Thank you, again, to all of the people who have helped through the years. Of course, I could not do it myself were it not for the fact that even through ten editions, this truly Herculean undertaking remains a labor of love.

Ronald Jay Cohen, Ph.D., ABPP, ABAP  
*Diplomate, American Board of Professional Psychology (Clinical)*  
*Diplomate, American Board of Assessment Psychology*

## *Psychological Testing and Assessment*

All fields of human endeavor use measurement in some form, and each field has its own set of measuring tools and measuring units. For example, you become aware of unique measurement units when making major purchases. When buying a new smartphone or computer, measurements of speed (e.g., gigahertz), screen resolution (e.g., 12 megapixels), and storage (e.g., 512 gigabytes) are salient, whereas the 4 Cs (i.e., cut, color, clarity, and carat) become relevant measurement terms when considering a marriage proposal. You also witnessed the worldwide importance of developing faster measurement tools to identify asymptomatic virus carriers during the COVID-19 pandemic. As a student of psychological measurement, you need a working familiarity with some of the commonly used units of measure in psychology as well as knowledge of some of the many measuring tools employed. In the pages that follow, you will gain that knowledge as well as an acquaintance with the history of measurement in psychology and an understanding of its theoretical basis.

Good helpers take time to understand the situation before helping a person. Great helpers *make time* to understand the person who needs help. Psychological assessment applies scientific rigor to the gentle art of understanding people before helping them. Psychological assessment encompasses a wide variety of methods, including direct observation, interviews, questionnaires, tests, and case file reviews.

Tests have been used by educators since ancient times, but psychological tests were developed only after psychology emerged as a formal scientific discipline in the late 1800s. Whereas educational testing tells us how much a person has learned, psychological assessment tells us what can be learned about a person. The experience of being closely listened to and deeply understood is itself a great comfort to many individuals who have sought the help of psychological assessment providers.

### Testing and Assessment

The roots of contemporary psychological testing and assessment can be found in early twentieth-century France. In 1905, Alfred Binet and a colleague published a test designed to help place Paris schoolchildren in appropriate classes. The first society-wide application of psychological testing resulted from an attempt by Parisian educators and lawmakers to live up to the ideals inscribed on public buildings all over France: *liberté, égalité, fraternité* (liberty, equality, fraternity). In a series of sweeping educational reforms in the 1870s–1890s, France became one of the first countries to mandate free public education for all its children. Of course, mandating high-quality education for everyone is not the same as educating everyone equally well. Not long after the laws went into effect, French educational institutions were confronted with the full

magnitude of human diversity. Children with intellectual disabilities need higher levels of support. In previous generations, children with intellectual disabilities were given intensive education only if their families could pay for such services. No longer.

How does one meet the complex educational needs of students with the severest of disabilities while also treating students equally? French educational administrators wanted an efficient, accurate, and fair method of deciding which children were best served by learning in separate, special classes with slower, more intensive instruction. The Minister of Public Instruction commissioned a study of the matter, and the committee asked Alfred Binet and his colleague Theodore Simon to create a test that would help school personnel make placement decisions. Binet and Simon warned that without objective scientific rigor, decisions are made haphazardly, “which are subjective, and consequently uncontrolled. [...] Some errors are excusable in the beginning, but if they become too frequent, they may ruin the reputations of these new [public school] institutions” (Binet & Simon, 1905, pp. 11–12).

Binet and Simon created a series of tests designed to forecast which students would likely fall ever further behind their peers without additional support. Although the Binet–Simon test became known as an “intelligence test,” its designers specifically warned that the test did not measure intelligence in its totality. Rather, the test was designed for the narrow purpose of identifying intellectually disabled children who needed additional help. Subsequent research found that the tests achieved their stated design goals reasonably well. Binet’s test would have consequences well beyond the Paris school district. Within a decade an English-language version of Binet’s test was prepared for use in schools in the United States. When the United States declared war on Germany and entered World War I in 1917, the military needed a way to screen large numbers of recruits quickly for intellectual and emotional problems. Psychological testing provided this methodology. During World War II, the military would depend even more on psychological tests to screen recruits for service. Following the war, more and more tests purporting to measure an ever-widening array of psychological variables were developed and used. There were tests to measure not only intelligence but also personality, brain functioning, performance at work, and many other aspects of psychological and social functioning.

William Stern, who developed a refined method of scoring Binet’s test—the Intelligence Quotient (IQ)—was horrified when Binet’s tests were later used by many institutions as tools of oppression rather than for their original purpose of liberation. He wrote movingly about how IQ tests should not be used to degrade individuals (Stern, 1933, as translated by Lamiell, 2003):

Under all conditions, human beings are and remain the centers of their own psychological life and their own worth. In other words, they remain persons, even when they are studied and treated from an external perspective with respect to others’ goals. ... Working “on” a human being must always entail working “for” a human being. (pp. 54–55)

We adopt Stern’s ideals and share his vision that with proper ethical safeguards, psychological tests can fulfill their original purpose—helping individuals and creating a more just society for everyone.

### *Psychological Testing and Assessment Defined*

The world’s receptivity to Binet’s test in the early twentieth century spawned not only more tests but more test developers, more test publishers, more test users, and the emergence of what, logically enough, has become known as a testing enterprise. “Testing” was the term used to refer to everything from the administration of a test (as in “Testing in progress”) to the interpretation of a test score (“The testing indicated that . . .”). During World War I, the term “testing” aptly described the group screening of thousands of military recruits. We suspect that it was then that the term gained a powerful foothold in the vocabulary of

professionals and laypeople. The use of “testing” to denote everything from test administration to test interpretation can be found in postwar textbooks (such as Chapman, 1921; Hull, 1922; Spearman, 1927) as well as in various test-related writings for decades thereafter. However, by World War II a semantic distinction between testing and a more inclusive term, “assessment,” began to emerge.

Military, clinical, educational, and business settings are but a few of the many contexts that entail behavioral observation and active integration by assessors of test scores and other data. In such situations, the term *assessment* may be preferable to *testing*. In contrast to testing, assessment acknowledges that tests are only one type of tool used by professional assessors (along with other tools, such as the interview), and that the value of a test, or of any other tool of assessment, is intimately linked to the knowledge, skill, and experience of the assessor.

The semantic distinction between psychological testing and psychological assessment is blurred in everyday conversation. Somewhat surprisingly, the distinction between the two terms still remains blurred in some published “psychological testing” textbooks. Yet the distinction is important. Society at large is best served by a clear definition of and differentiation between these two terms as well as related terms such as *psychological test user* and *psychological assessor*. Clear distinctions between such terms may also help avoid the turf wars now brewing between psychology professionals and members of other professions seeking to use various psychological tests. In many psychological evaluation contexts, conducting an assessment requires greater education, training, and skill than simply administering a test.

#### JUST THINK . . .

Describe a situation in which testing is more appropriate than assessment. By contrast, describe a situation in which assessment is more appropriate than testing.

We define **psychological assessment** as the gathering and integration of psychology-related data for the purpose of making a psychological evaluation that is accomplished through the use of tools such as tests, interviews, case studies, behavioral observation, and specially designed apparatuses and measurement procedures. We define **psychological testing** as the process of measuring psychology-related variables by means of devices or procedures designed to obtain a sample of behavior. Some of the differences between these two processes are presented in Table 1–1.<sup>1</sup>

**Varieties of assessment** The term *assessment* may be modified in a seemingly endless number of ways, each such modification referring to a particular variety or area of assessment. Sometimes the meaning of the specialty area can be readily discerned just from the word or term that modifies “assessment.” For example, the term “therapeutic psychological assessment” refers to assessment that helps individuals understand and solve their problems. Also intuitively obvious, the term **educational assessment** refers to, broadly speaking, the use of tests and other tools to evaluate abilities and skills relevant to success or failure in a school or pre-school context. Intelligence tests, achievement tests, and reading comprehension tests are some of the evaluative tools that may spring to mind with the mention of the term “educational assessment.” But what springs to mind with the mention of other, less common assessment terminology? Consider, for example, terms like *retrospective assessment*, *remote assessment*, and *ecological momentary assessment*.

---

1. Especially when discussing general principles related to the creation of measurement procedures, as well as the creation, manipulation, or interpretation of data generated from such procedures, the word *test* (as well as related terms, such as *test score*) may be used in the broadest and most generic sense; that is, “test” may be used in shorthand fashion to apply to almost any procedure that entails measurement (including, e.g., situational performance measures). Accordingly, when we speak of “test development” in Chapter 8, many of the principles set forth will apply to the development of other measurements that are not, strictly speaking, “tests” (such as situational performance measures, as well as other tools of assessment). Having said that, let’s reemphasize that a real and meaningful distinction exists between the terms *psychological testing* and *psychological assessment*, and that effort should continually be made not to confuse the meaning of these two terms.

**Table 1–1**  
**Testing in Contrast to Assessment**

*In contrast to the process of administering, scoring, and interpreting psychological tests (psychological testing), psychological assessment is a problem-solving process that can take many different forms. How psychological assessment proceeds depends on many factors, not the least of which is the reason for assessing. Different tools of evaluation—psychological tests among them—might be marshaled in the process of assessment, depending on the particular objectives, people, and circumstances involved as well as on other variables unique to the particular situation.*

*Admittedly, the line between what constitutes testing and what constitutes assessment is not always as clear as we might like it to be. However, by acknowledging that such ambiguity exists, we can work to sharpen our definition and use of these terms. It seems useful to distinguish the differences between testing and assessment in terms of the typical objective, process, and outcome of an evaluation and also in terms of the role and skill of the evaluator. Keep in mind that, although these are useful distinctions to consider, exceptions can always be found.*

Testing	Assessment
<i>Objective</i>	
To obtain some gauge, usually numerical in nature, with regard to an ability or attribute.	To answer a referral question, solve a problem, or arrive at a decision through the use of tools of evaluation.
<i>Process</i>	
Testing may be conducted individually or in groups. After test administration, the tester adds up “the number of correct answers or the number of certain types of responses . . . with little if any regard for the how or mechanics of such content” (Maloney & Ward, 1976, p. 39).	Assessment is individualized. In contrast to testing, assessment focuses on how an individual processes rather than simply the results of that processing.
<i>Role of Evaluator</i>	
The tester is not key to the process; one tester may be substituted for another tester without appreciably affecting the evaluation.	The assessor is key to the process of selecting tests and/or other tools of evaluation as well as in drawing conclusions from the entire evaluation.
<i>Skill of Evaluator</i>	
Testing requires technician-like skills in administering and scoring a test as well as in interpreting a test result.	Assessment requires an educated selection of tools of evaluation, skill in evaluation, and thoughtful organization and integration of data.
<i>Outcome</i>	
Testing yields a test score or series of test scores.	Assessment entails a logical problem-solving approach that brings to bear many sources of data designed to shed light on a referral question.

For the record, the term **retrospective assessment** is defined as *the use of evaluative tools to draw conclusions about psychological aspects of a person as they existed at some point in time prior to the assessment*. There are unique challenges and hurdles to be overcome when conducting retrospective assessments regardless if the subject of the evaluation is alive (Teel et al., 2016) or is deceased (Reyman & Shankar, 2015). **Remote assessment** refers to *the use of tools of psychological evaluation to gather data and draw conclusions about a subject who is not in physical proximity to the person or people conducting the evaluation*. One example of how psychological assessments may be conducted remotely was provided in this chapter’s *Close-Up* feature. In each chapter of this book, we will spotlight one topic for “a closer look.”

## Behavioral Assessment Using Smartphones\*

Much like the state of one's physical health, the state of one's mental health and functioning is changing and fluid. Varied internal factors (such as neurochemistry and hormonal shifts), external factors (such as marital discord and job pressures), or combinations thereof may affect mental health and functioning. This fluctuation is as true for people with no diagnosis of mental disorder as it is for patients suffering from chronic psychiatric illnesses.

Changes in people's mental health status rarely come "out of the blue" (or, without warning). Behavioral signs that someone is experiencing increased stress and mental health difficulties may include changes in sleep and eating patterns, social engagement, and physical activity. Because these changes may emerge gradually over time, they can go unnoticed by family members, close friends, or even the affected individuals themselves. By the time most people seek support or professional care, their mental health and functioning may have deteriorated substantially. Identifying behavioral patterns that are associated with increased risk for underlying mental health difficulties is a first step toward more efficient treatment, perhaps even prevention.

Dr. Dror Ben-Zeev and his colleagues have begun to identify problematic behavioral patterns using a device that is already in the hands of billions: the smartphone. The smartphone (or, a mobile phone that features computational capacity) comes equipped with multiple embedded sensors that measure variables such as acoustics, location, and movement. Ben-Zeev's team uses sophisticated smartphone software that enables them to repurpose these sensors and capture an abundance of information about the smartphone user's environment and behavior. Their program activates the smartphone's microphone every few minutes to capture ambient sound. If the software detects human conversation, it remains active for the duration of the conversation. To protect user's privacy, the speech detection system does not record raw audio. It processes the data in real-time to extract and store conversation-related data while actual conversations cannot be reconstructed. The software calculates both the number of conversations and the average length of a conversation engaged in during a 24-hour period.

In addition to re-purposing the microphone in a cell phone, Ben-Zeev's system repurposes the smartphone's global



GaudiLab/Shutterstock

positioning system (GPS). When the user is outdoors, the GPS generates geospatial coordinates helpful in determining the daily distance covered, as well as the amount of time spent at specific locations. When the research team conducts studies with individuals who do not move from one location to another, such as hospitalized patients in closed psychiatric units, they place microbluetooth beacons in different rooms throughout the venue. As the subject moves from one room to another, the smartphone's bluetooth sensor receives signals sent by the beacons, and records the subject's precise position in the unit.

A typical smartphone also comes equipped with *accelerometers*; these devices are designed to detect motion. Ben-Zeev's monitoring system collects the accelerometer data to determine whether the individual is or is not active.

The smartphone system collects and stores all of the sensor data and transmits it periodically to a secure study server. There, the information is processed and displayed on a digital dashboard. By means of this system, multidimensional data from faraway places can be viewed online to help clinicians and researchers better understand experiences that cause changes in stress level and general mental health. One smartphone-sensing study conducted with college undergraduate and graduate student subjects over a 10-week period included pre- and post-measures of depression. The data suggested that social engagement (as measured by the speech detection software) and daily geospatial activity (as measured by GPS) were significantly related to changes in level of depression (Ben-Zeev et al., 2015a).

(continued)

\*This *Close-Up* was written by Dror Ben-Zeev of the Department of Psychiatry of the Geisel School of Medicine at Dartmouth.

## Behavioral Assessment Using Smartphones *(continued)*

Of course, tracking someone via their smartphone without their awareness and consent would be unethical. However, for people who may be at risk for mental health problems, or for those who already struggle with psychiatric conditions and need support, this unobtrusive approach may have value. Explaining to patients (or their representatives) what the technology is, how it works, and how data from it may be used for patient benefit, may well allay any privacy concerns. Preliminary research has suggested that even patients with severe mental illness can understand and appreciate the potential benefits of remote assessment by means of the smartphone tracking system (Ben-Zeev et al., 2012). Most of the subjects studied stated that they would have no objection to using a system that could not only passively detect when they were not doing well, but offer them helpful and timely suggestions for improving their mental state

(Ben-Zeev et al., 2015b). Patients and mental health professionals alike appreciate the promise of this potentially useful method for detecting emerging high-risk patterns that require preventative or immediate treatment.

As technology evolves, one can imagine a future in which at-risk individuals derive benefit from smartphones repurposed to serve as objectively scalable measures of behavior (Ben-Zeev, 2017). Used in a clinically skilled fashion and with appropriate protections of patient privacy, these ubiquitous devices, now repurposed to yield behavioral data, may be instrumental in creating meaningful diagnostic insights and profiles. In turn, such minute-to-minute assessment data may yield highly personalized and effective treatment protocols.

Used with permission of Dror Ben-Zeev.

In this chapter, the *Close-Up* box explored how the smartphone revolution in communication may also signal a revolution in the way that psychological assessments are conducted.

Psychological assessment by means of smartphones also serves as an example of an approach to assessment called **ecological momentary assessment (EMA)**. EMA refers to the “in the moment” evaluation of specific problems and related cognitive and behavioral variables at the exact time and place that they occur. Using various tools of assessment, EMA has been used to help tackle diverse clinical problems including post-traumatic stress disorder (Black et al., 2016), problematic smoking (Ruscio et al., 2016), chronic abdominal pain in children (Schurman & Friesen, 2015), and attention-deficit/hyperactivity symptoms (Li & Lansford, 2018).

**The process of assessment** In general, the process of assessment begins with a referral for assessment from a source such as a teacher, parent, school psychologist, counselor, judge, clinician, or corporate human resources specialist. Typically one or more referral questions are put to the assessor about the assessee. Some examples of referral questions are: “Can this child function in a general education environment?,” “Is this defendant competent to stand trial?,” and “How well can this employee be expected to perform if promoted to an executive position?”

The assessor may meet with the assessee or others before the formal assessment in order to clarify aspects of the reason for referral. The assessor prepares for the assessment by selecting the tools of assessment to be used. For example, if the assessment occurs in a corporate or military setting and the referral question concerns the assessee’s leadership ability, the assessor may wish to employ a measure (or two) of leadership. Typically, the assessor’s own past experience, education, and training play a key role in the specific tests or other tools to be employed in the assessment. Sometimes an institution

### JUST THINK . . .

What qualities makes a good leader? How might these qualities be measured?

in which the assessment is taking place has prescribed guidelines for which instruments can and cannot be used. In almost every assessment situation, particularly situations that are relatively novel to the assessor, the tool selection process is informed by some research in preparation for the assessment. For example, in the assessment of leadership, the tool selection procedure might be informed by reviewing publications dealing with behavioral studies of leadership (Derue et al., 2011), psychological studies of leaders (Kouzes & Posner, 2007), cultural issues in leadership (Byrne & Bradley, 2007), or whatever aspect of leadership on which the assessment will be focused (Carnevale et al., 2011; Elliott, 2011; Rosenman et al., 2015).

Subsequent to the selection of the instruments or procedures to be employed, the formal assessment will begin. After the assessment, the assessor writes a report of the findings that is designed to answer the referral question. More feedback sessions with the assessee and/or interested third parties (such as the assessee's parents and the referring professional) may also be scheduled.

Different assessors may approach the assessment task in different ways. Some assessors approach the assessment with minimal input from assessee themselves. Other assessors view the process of assessment as more of a collaboration between the assessor and the assessee. For example, in one approach to assessment, referred to (logically enough) as **collaborative psychological assessment**, the assessor and assessee may work as "partners" from initial contact through final feedback (Finello, 2011; Fischer, 1978, 2004, 2006). The assessment provider encourages collaboration by asking questions like, "After this assessment is finished, what would you like to know that you do not know already?" One variety of collaborative assessment includes an element of therapy as part of the process. Stephen Finn and his colleagues (Finn, 2003, 2009, 2011; Finn & Martin, 1997; Finn & Tonsager, 2002; Fischer & Finn, 2014) have described a collaborative approach to assessment called **therapeutic psychological assessment**. In traditional psychological evaluations, the assessment is designed to have its intended benefits at the end of the process: The examiner explains the results, summarizes the case conceptualization, and shares a list of recommendations designed to help the examinee.

In contrast, therapeutic psychological assessment aims to be helpful throughout the assessment process. The results are not revealed at the end, but shared immediately so that both the assessor and the assessee can co-develop an interpretation of the results and decide what questions require further assessment. In this way, therapeutic self-discovery and new understandings are encouraged throughout the assessment process.

Another approach to assessment that seems to have picked up momentum in recent years, most notably in educational settings, is referred to as *dynamic assessment* (Poehner & van Compernelle, 2011). The term *dynamic* may suggest that a psychodynamic or psychoanalytic approach to assessment is being applied, but that is not the case. As used in the present context, *dynamic* is used to describe the interactive, changing, or varying nature of the assessment. In general, **dynamic assessment** refers to an interactive approach to psychological assessment that usually follows a model of (1) evaluation, (2) intervention of some sort, and (3) evaluation. Dynamic assessment is most typically employed in educational settings, although it may be employed in correctional, corporate, neuropsychological, clinical, and most any other setting as well.

Intervention between evaluations, sometimes even between individual questions posed or tasks given, might take many different forms, depending upon the purpose of the dynamic assessment (Haywood & Lidz, 2007). For example, an assessor may intervene in the course of an evaluation of an assessee's abilities with increasingly more explicit feedback or hints. The purpose of the intervention may be to provide assistance with mastering the task at hand. Progress in mastering the same or similar tasks is then measured. In essence, dynamic assessment provides a means for evaluating how the assessee processes or benefits from some type of intervention (feedback, hints, instruction, therapy, and so forth) during the course of evaluation. In some educational contexts, dynamic assessment may be viewed as a way of measuring not just learning but "learning potential," or "learning how to learn" skills. Computers are one tool used to help meet the objectives of dynamic assessment (Wang, 2011). There are others . . .

# The Tools of Psychological Assessment

## The Test

A **test** is defined simply as a measuring device or procedure. When the word *test* is prefaced with a modifier, it refers to a device or procedure designed to measure a variable related to that modifier. Consider, for example, the term *medical test*, which refers to a device or procedure designed to measure some variable related to the practice of medicine (including a wide range of tools and procedures, such as X-rays, blood tests, and testing of reflexes). In a like manner, the term **psychological test** refers to a device or procedure designed to measure variables related to psychology (such as intelligence, personality, aptitude, interests, attitudes, or values). Whereas a medical test might involve analysis of a sample of blood, tissue, or the like, a psychological test almost always involves analysis of a sample of behavior. The behavior sample could range from responses to a pencil-and-paper questionnaire, to verbal responses to questions related to the performance of some task. The behavior sample could be elicited by the stimulus of the test itself, or it could be naturally occurring behavior (observed by the assessor in real time as it occurs, or it can be recorded and observed at a later time).

Psychological tests and other tools of assessment may differ with respect to a number of variables, such as content, format, administration procedures, scoring and interpretation procedures, and technical quality. The *content* (subject matter) of the test will, of course, vary with the focus of the particular test. But even two psychological tests purporting to measure the same thing—for example, personality—may differ widely in item content. This difference is, in part, because two test developers might have entirely different views regarding what is important in measuring “personality”; different test developers employ different definitions of “personality.” Additionally, different test developers come to the test development process with different theoretical orientations. For example, items on a psychoanalytically oriented personality test may have little resemblance to those on a behaviorally oriented personality test, yet both are personality tests. A psychoanalytically oriented personality test might be chosen for use by a psychoanalytically oriented assessor, and an existentially oriented personality test might be chosen for use by an existentially oriented assessor.

### JUST THINK . . .

Imagine you wanted to develop a test for a personality trait you termed “goth.” How would you define this trait? What kinds of items would you include in the test? Why would you include those kinds of items? How would you distinguish this personality trait from others?

The term **format** pertains to the form, plan, structure, arrangement, and layout of test items as well as to related considerations such as time limits. *Format* is also used to refer to the form in which a test is administered: computerized, pencil-and-paper, or some other form. When making specific reference to a computerized test, the format may also involve the form of the software: local or online/cloud-based software and storage. The term *format* is not confined to tests. *Format* is also used to denote the form or structure of other evaluative tools and processes, such as the guidelines for creating a portfolio work sample.

Tests differ in their *administration procedures*. Some tests, particularly those designed for administration on a one-to-one basis, may require an active and knowledgeable test administrator. The test administration may involve demonstration of various kinds of tasks demanded of the assessee, as well as trained observation of an assessee’s performance. Alternatively, some tests, particularly those designed for administration to groups, may not even require the test administrator to be present while the testtakers independently complete the required tasks.

Tests differ in their *scoring and interpretation procedures*. To better understand how and why, let’s define *score* and *scoring*. Sports enthusiasts are no strangers to these terms. For them, these terms refer to the number of points accumulated by competitors and the process of accumulating those points. In testing and assessment, we formally define **score** as a code

or summary statement, usually but not necessarily numerical in nature, that reflects an evaluation of performance on a test, task, interview, or some other sample of behavior. **Scoring** is the process of assigning such evaluative codes or statements to performance on tests, tasks, interviews, or other behavior samples. In the world of psychological assessment, many different types of scores exist. Some scores result from the simple summing of responses (such as the summing of correct/incorrect or agree/disagree responses), and some scores are derived from more elaborate procedures.

Scores themselves can be described and categorized in many different ways. For example, one type of score is the *cut score*. A **cut score** (also referred to as a *cutoff score* or simply a *cutoff*) is a reference point, usually numerical, derived by judgment and used to divide a set of data into two or more classifications. Some action will be taken or some inference will be made on the basis of these classifications. Cut scores on tests, usually in combination with other data, are used in schools in many contexts. For example, they may be used in grading, and in making decisions about the class or program to which children will be assigned. Cut scores are used by employers as aids to decision making about personnel hiring, placement, and advancement. State agencies use cut scores as aids in licensing decisions. There are probably more than a dozen different methods that can be used to formally derive cut scores (Dwyer, 1996). If you're curious about what some of those different methods are, stay tuned; we cover that in an upcoming chapter.

Sometimes no formal method is used to arrive at a cut score. Some teachers use an informal “eyeball” method to proclaim, for example, that a score of 65 or more on a test means “pass” and a score of 64 or below means “fail.” Whether formally or informally derived, cut scores typically take into account, at least to some degree, the values of those who set them. Consider, for example, two professors who teach the same course at the same college. One professor might set a cut score for passing the course that is significantly higher (and more difficult for students to attain) than the other professor. There is also another side to the human equation as it relates to cut scores, one that is seldom written about in measurement texts. This phenomenon concerns the emotional consequences of “not making the cut” and “just making the cut” (see Figure 1–1).

Tests differ widely in terms of their guidelines for scoring and interpretation. Some tests are self-scored by the testtakers themselves, others are scored by computer, and others require scoring by trained examiners. Some tests, such as most tests of intelligence, come with test manuals that are explicit not only about scoring criteria but also about the nature of the interpretations that can be made from the scores. Other tests, such as the Rorschach Inkblot Test, are sold with no manual at all. The (presumably qualified) purchaser buys the stimulus materials and then selects and uses one of many available guides for administration, scoring, and interpretation.

Tests differ with respect to their **psychometric soundness** or technical quality. Synonymous with the antiquated term *psychometry*, **psychometrics** is defined as the science of psychological measurement. Variants of these words include the adjective *psychometric* (which refers to measurement that is psychological in nature) and the nouns **psychometrist** and **psychometrician** (both terms referring to a professional who uses, analyzes, and interprets psychological test data). One speaks of the psychometric soundness of a test when referring to how consistently and how accurately a psychological test measures what it purports to measure. Assessment professionals also speak of the psychometric *utility* of a particular test or assessment method. In this context, **utility** refers to the usefulness or practical value that a test or other tool of assessment has for a particular purpose. These concepts are elaborated on in subsequent chapters. Now, returning to our discussion of tools of assessment, meet one well-known tool that, as they say, “needs no introduction.”

#### JUST THINK . . .

How might one test of intelligence have more utility than another test of intelligence in the same school setting?