

COMPREHENSIVE HANDBOOK OF PSYCHOLOGICAL ASSESSMENT

VOLUME 2 PERSONALITY ASSESSMENT

Mark J. Hilsenroth

Daniel L. Segal

Volume Editors

Michel Hersen

Editor-in-Chief



John Wiley & Sons, Inc.

**COMPREHENSIVE HANDBOOK
OF
PSYCHOLOGICAL ASSESSMENT**

COMPREHENSIVE HANDBOOK OF PSYCHOLOGICAL ASSESSMENT

VOLUME 2 PERSONALITY ASSESSMENT

Mark J. Hilsenroth

Daniel L. Segal

Volume Editors

Michel Hersen

Editor-in-Chief



John Wiley & Sons, Inc.

This book is printed on acid-free paper. ☺

Copyright © 2004 by John Wiley & Sons, Inc. All rights reserved.

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.

Published simultaneously in Canada.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4470, or on the web at www.copyright.com. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, e-mail: permcoordinator@wiley.com.

Limit of Liability/Disclaimer of Warranty: While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties or merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold with the understanding that the publisher is not engaged in rendering professional services. If legal, accounting, medical, psychological, or any other expert assistance is required, the services of a competent professional person should be sought.

Designations used by companies to distinguish their products are often claimed as trademarks. In all instances where John Wiley & Sons, Inc., is aware of a claim, the product names appear in initial capital or all capital letters. Readers, however, should contact the appropriate companies for more complete information regarding trademarks and registration.

For general information on our products and services please contact our Customer Care Department within the United States at (800) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books. For more information about Wiley products, visit our web site at www.Wiley.com.

Library of Congress Cataloging-in-Publication Data:

Comprehensive handbook of psychological assessment / editor-in-chief, Michel Hersen.

p. cm.

Includes bibliographical references and index.

Contents: v. 1 Intellectual and neuropsychological assessment / editors, Gerald Goldstein and Sue R. Beers — v. 2. Personality assessment / editors, Mark J. Hilsenroth and Daniel L. Segal — v. 3. Behavioral assessment / editors, Stephen N. Haynes and Elaine M. Heiby — v. 4. Industrial and organizational assessment / editor, Jay C. Thomas.

ISBN 0-471-41610-X (set : hardcover : alk. paper) — ISBN 0-471-41611-8 (v. 1 : hardcover : alk. paper) — ISBN 0-471-41612-6 (v. 2 : hardcover : alk. paper) — ISBN 0-471-41613-4 (v. 3 : hardcover : alk. paper) — ISBN 0-471-41614-2 (v. 4 : hardcover : alk. paper)

1. Psychological tests. I. Hersen, Michel.

BF176 .C654 2003

150'.28'7—dc21

2002193381

Printed in the United States of America.

10 9 8 7 6 5 4 3 2 1

To my wife, Jessica Hilsenroth,
for her love, support, and encouragement

To my wife, Cindy Kamilar,
and daughter, Shaynie—
with love

Contents

Handbook Preface xi

Contributors xiii

SECTION ONE OBJECTIVE ASSESSMENT OF PERSONALITY AND PSYCHOPATHOLOGY

- 1 OBJECTIVE ASSESSMENT OF PERSONALITY AND PSYCHOPATHOLOGY: AN OVERVIEW** 3
Daniel L. Segal and Frederick L. Coolidge

Part One: Adult Assessment Instruments

- 2 THE PERSONALITY ASSESSMENT INVENTORY (PAI)** 15
Leslie C. Morey and Christina D. Boggs
- 3 THE MINNESOTA MULTIPHASIC PERSONALITY INVENTORY-2 (MMPI-2)** 30
James N. Butcher, Mera M. Atlis, and Jungwon Hahn
- 4 THE SIXTEEN PERSONALITY FACTOR (16PF) QUESTIONNAIRE** 39
Heather E.P. Cattell
- 5 THE BECK DEPRESSION INVENTORY-II (BDI-II), BECK HOPELESSNESS SCALE (BHS), AND BECK SCALE FOR SUICIDE IDEATION (BSS)** 50
David J.A. Dozois and Roger Covin
- 6 MEASURING ANXIETY, ANGER, DEPRESSION, AND CURIOSITY AS EMOTIONAL STATES AND PERSONALITY TRAITS WITH THE STAI, STAXI, AND STPI** 70
Charles D. Spielberger and Eric C. Reheiser
- 7 THE HAMILTON DEPRESSION RATING SCALE (HAMD)** 87
Kenneth A. Kobak
- 8 THE EYSENCK PERSONALITY SCALES: THE EYSENCK PERSONALITY QUESTIONNAIRE-REVISED (EPQ-R) AND THE EYSENCK PERSONALITY PROFILER (EPP)** 99
Jeremy Miles and Susanne Hempel
- 9 THE MILLON CLINICAL MULTIAXIAL INVENTORY-III (MCMI-III)** 108
Theodore Millon and Sarah E. Meagher

- 10 THE PERSONALITY DIAGNOSTIC QUESTIONNAIRE-4 (PDQ-4) 122**
R. Michael Bagby and Peter Farvolden
- 11 THE STRUCTURED CLINICAL INTERVIEW FOR *DSM-IV* AXIS I DISORDERS (SCID-I) AND THE
STRUCTURED CLINICAL INTERVIEW FOR *DSM-IV* AXIS II DISORDERS (SCID-II) 134**
Michael B. First and Miriam Gibbon
- 12 THE SCHEDULE FOR AFFECTIVE DISORDERS AND SCHIZOPHRENIA (SADS) 144**
Richard Rogers, Rebecca L. Jackson, and MaryLouise Cashel
- 13 THE DIAGNOSTIC INTERVIEW SCHEDULE (DIS) 153**
Wilson M. Compton and Linda B. Cottler
- 14 THE ANXIETY DISORDERS INTERVIEW SCHEDULE FOR *DSM-IV* (ADIS-IV) 163**
Jessica R. Grisham, Timothy A. Brown, and Laura A. Campbell

Part Two: Childhood and Adolescent Assessment Instruments

- 15 THE CHILD BEHAVIOR CHECKLIST/4–18 (CBCL/4–18) 179**
Paul E. Greenbaum, Robert F. Dedrick, and Lodi Lipien
- 16 THE PERSONALITY INVENTORY FOR CHILDREN, SECOND EDITION (PIC-2), PERSONALITY
INVENTORY FOR YOUTH (PIY), AND STUDENT BEHAVIOR SURVEY (SBS) 192**
David Lachar
- 17 THE MINNESOTA MULTIPHASIC PERSONALITY INVENTORY-ADOLESCENT (MMPI-A) 213**
Ruth A. Baer and Jason C. Rinaldo
- 18 THE REYNOLDS ADOLESCENT DEPRESSION SCALE-SECOND EDITION (RADS-2) 224**
William M. Reynolds
- 19 THE CHILD ABUSE POTENTIAL (CAP) INVENTORY 237**
Joel S. Milner
- 20 THE SCHEDULE FOR AFFECTIVE DISORDERS AND SCHIZOPHRENIA FOR SCHOOL-AGE
CHILDREN: PRESENT AND LIFETIME VERSION (K-SADS-PL) 247**
Joan Kaufman and Amanda E. Schweder
- 21 THE DIAGNOSTIC INTERVIEW SCHEDULE FOR CHILDREN (DISC) 256**
David Shaffer, Prudence Fisher, and Christopher Lucas
- 22 THE DIAGNOSTIC INTERVIEW FOR CHILDREN AND ADOLESCENTS (DICA) 271**
Kathryn M. Rourke and Wendy Reich

SECTION TWO

PROJECTIVE ASSESSMENT OF PERSONALITY AND PSYCHOPATHOLOGY

Part Three: Overview, Conceptual, and Empirical Foundations

- 23 PROJECTIVE ASSESSMENT OF PERSONALITY AND PSYCHOPATHOLOGY: AN OVERVIEW 283**
Mark J. Hilsenroth
- 24 PROJECTIVE TESTS: THE NATURE OF THE TASK 297**
Martin Leichtman
- 25 THE RELIABILITY AND VALIDITY OF THE RORSCHACH AND THEMATIC APPERCEPTION TEST (TAT) COMPARED TO OTHER PSYCHOLOGICAL AND MEDICAL PROCEDURES: AN ANALYSIS OF SYSTEMATICALLY GATHERED EVIDENCE 315**
Gregory J. Meyer

Part Four: Specific Instruments

- 26 RORSCHACH ASSESSMENT: CURRENT STATUS 343**
Irving B. Weiner
- 27 THE THEMATIC APPERCEPTION TEST (TAT) 356**
Robert J. Moretti and Edward D. Rossini
- 28 THE USE OF SENTENCE COMPLETION TESTS WITH ADULTS 372**
Alissa Sherry, Eric Dahlen, and Margot Holaday
- 29 USE OF GRAPHIC TECHNIQUES IN PERSONALITY ASSESSMENT: RELIABILITY, VALIDITY, AND CLINICAL UTILITY 387**
Leonard Handler, Ashley Campbell, and Betty Martin
- 30 THE HAND TEST: ASSESSING PROTOTYPICAL ATTITUDES AND ACTION TENDENCIES 405**
Harry J. Sivec, Charles A. Waehler, and Paul E. Panek
- 31 EARLY MEMORIES AND PERSONALITY ASSESSMENT 421**
J. Christopher Fowler
- 32 THE ADULT ATTACHMENT PROJECTIVE: MEASURING INDIVIDUAL DIFFERENCES IN ATTACHMENT SECURITY USING PROJECTIVE METHODOLOGY 431**
Carol George and Malcolm West

Part Five: Specific Content Areas

- 33 PROJECTIVE ASSESSMENT OF OBJECT RELATIONS 449**
George Stricker and Jane Gooen-Piels
- 34 PROJECTIVE ASSESSMENT OF DEFENSE MECHANISMS 466**
John H. Porcerelli and Stephen Hibbard

- 35 **PROJECTIVE ASSESSMENT OF INTERPERSONAL DEPENDENCY** 476
Robert F. Bornstein
- 36 **PROJECTIVE ASSESSMENT OF BORDERLINE PSYCHOPATHOLOGY** 485
Mark A. Blais and Kimberly Bistis
- 37 **PROJECTIVE ASSESSMENT OF PSYCHOLOGICAL TRAUMA** 500
Judith Armstrong and Nancy Kaser-Boyd
- 38 **PROJECTIVE ASSESSMENT OF SUICIDAL IDEATION** 513
Daniel J. Holdwick Jr. and Leah Brzusiewicz
- 39 **PROJECTIVE ASSESSMENT OF DISORDERED THINKING** 526
James H. Kleiger

Part Six: Special Populations and Settings

- 40 **THE USE OF PROJECTIVE TESTS IN ASSESSING NEUROLOGICALLY IMPAIRED POPULATIONS** 539
Arpi Minassian and William Perry
- 41 **PROJECTIVE ASSESSMENT OF MALINGERING** 533
Jon D. Elhai, Bill N. Kinder, and B. Christopher Frueh
- 42 **PROJECTIVE ASSESSMENT OF PERSONALITY IN FORENSIC SETTINGS** 562
Joseph T. McCann
- 43 **CULTURAL APPLICATIONS OF THE RORSCHACH, APPERCEPTION TESTS, AND FIGURE DRAWINGS** 573
Barry Ritzler
- 44 **COLLABORATIVE EXPLORATION WITH PROJECTIVE TECHNIQUES: A LIFE-WORLD APPROACH** 586
Constance T. Fischer, Emilija Georgievska, and Michael Melczak

Part Seven: Applications for Children and Adolescents

- 45 **SENTENCE COMPLETION MEASUREMENT OF PSYCHOSOCIAL MATURITY** 595
P. Michiel Westenberg, Stuart T. Hauser, and Lawrence D. Cohn
- 46 **ASSESSMENT OF OBJECT REPRESENTATION IN CHILDREN AND ADOLESCENTS: CURRENT TRENDS AND FUTURE DIRECTIONS** 617
Francis D. Kelly
- 47 **PROJECTIVE ASSESSMENT OF AFFECT IN CHILDREN'S PLAY** 628
Sandra W. Russ

Author Index 643

Subject Index 661

Handbook Preface

Over the last century the scope of activity of clinical psychologists has increased exponentially. In earlier times psychologists had a much more restricted range of responsibilities. Today psychologists not only provide assessments but treat a wide variety of disorders in an equally wide variety of settings, consult, teach, conduct research, help to establish ethical policies, deal with human engineering factors, have a strong media presence, work with law enforcement in profiling criminals, and have had increasing influence in the business world and in the realm of advertising, to identify just a few of the major activities in which they are engaged. Nonetheless, the hallmark of psychologists has always been assessment and it continues to be a mainstay of their practices in the twenty-first century. Indeed, in each of the activities just described, psychologists and their assistants are performing assessments of some sort.

In the nineteenth century our predecessors in Germany began to study individual differences and abilities in what then was the most scientific way. In the more than 120 years that have elapsed since these early efforts were carried out, the field of psychological assessment has seen many developments and permutations, ranging from educational needs to identify individuals with subnormal intelligence to attempts to measure unconscious dynamics with unstructured stimuli, wide-range governmental efforts to measure intelligence and other capabilities to screen out undesirable military recruits during wartime, development of evaluative tools to ensure successful personnel selection, the advent of behavioral and physiological assessments, the increased reliance on computerized assessments, and, most recently, the spectacular innovation of virtual reality assessments using the latest electronic technologies.

Thousands of specific assessment strategies and tests that are carried out on both an individual and group basis have been devised for almost every conceivable type of human endeavor. Many of these strategies have been carefully developed, tested, and refined, with norms available for many populations and excellent reliability and validity data reported. To keep abreast of all new developments in the field of assessment is a near impossibility, although scores of journals, books, and yearly publications are available that catalog such developments.

In considering how the field of psychological assessment has evolved over the last century with the resulting explosion of new technologies and new assessment devices, it seemed to us imperative to create a resource (*Comprehensive Handbook of Psychological Assessment: CHOPA*) that distilled this vast reservoir of data in a more manageable format for researchers, clinicians, educators, and students alike. Therefore, Tracey Belmont, our editor at John Wiley & Sons, the volume editors (Gerald Goldstein, Sue R. Beers, Mark J. Hilsenroth, Daniel L. Segal, Stephen N. Haynes, Elaine M. Heiby, and Jay C. Thomas), and I as editor-in-chief developed this four-volume format. This decision was both conceptual, in order to best capture the scope of the field, and pragmatic, so that individuals wishing to purchase a single volume (as a consequence of their unique interest) would be able to do so.

CHOPA includes four volumes with a total of 121 chapters written by renowned experts in their respective areas of expertise. In order the volumes are: 1, Intellectual and Neuropsychological Assessment; 2, Personality Assessment; 3, Behavioral Assessment; and 4, Industrial and Organizational Assessment. Each volume has an introductory chapter by the editor. In the case of Volume 2, there is an introductory chapter for objective tests and an introductory chapter for projective tests. In general, introductory chapters are concerned with a historical review, range of tests, theoretical considerations, psychometric concerns, range of populations for which the tests are appropriate, cross-cultural factors, accommodation for persons with disabilities, legal and ethical issues, computerization, and future perspectives. Chapters on individual tests or approaches cover many of the same areas but in much more specific detail, in addition, of course, to the test description and development. Other chapters are more conceptual and theoretical in nature and articulate an approach to evaluation, such as the chapters on clinical interviewing and program evaluation in Volume 3.

In developing the CHOPA concept and selecting chapters and contributors, our objective has been to be comprehensive in a global sense but not encyclopedic (i.e., detailing every conceivable and extant assessment strategy or test). However, we believe that we are sufficiently comprehensive so that the interested reader can move to greater specificity, if needed,

on the basis of the very current list of references for each chapter.

An endeavor as complicated as CHOPA has required the efforts of many people, and here we would like to acknowledge their various contributions. First, I personally would like to thank Tracey Belmont and her superb staff at John Wiley & Sons for recognizing the value of this project and for helping to bring the pieces together. Second, I thank the volume editors for their Herculean efforts in monitoring, reviewing, and reworking the contributions of their colleagues. Next, we

owe a debt of gratitude to our eminent contributors, who so graciously have shared their high levels of expertise with us. And finally, I would like to thank all of our staff here at Pacific University who contributed technical assistance to bringing this four-volume set to publication: Carole Londeree, Kay Waldron, Angelina Marchand, and Alex Duncan.

Michel Hersen
Forest Grove, Oregon

Contributors

Judith Armstrong, PhD

University of Southern California
Los Angeles, CA

Mera M. Atlis, PhD

University of Minnesota
Minneapolis, MN

Ruth A. Baer, PhD

University of Kentucky
Lexington, KY

R. Michael Bagby, PhD, CPsych

Clinical Research Department, Section on Personality and
Psychopathology
Centre for Addiction and Mental Health
Department of Psychiatry
University of Toronto
Toronto, Ontario, Canada

Kimberly Bistis, PsyD

Erich Lindemann Mental Health Center
Harvard Medical School
Boston, MA

Mark A. Blais, PsyD

Massachusetts General Hospital
Harvard Medical School
Boston, MA

Christina D. Boggs, MS

Texas A & M University
College Station, TX

Robert F. Bornstein, PhD

Gettysburg College
Gettysburg, PA

Timothy A. Brown, PsyD

Center for Anxiety and Related Disorders
Boston University
Boston, MA

Leah Brzuskiwicz, BS

Syracuse University
Syracuse, NY

James N. Butcher, PhD

University of Minnesota
Minneapolis, MN

Ashley Campbell, MA

University of Tennessee
Knoxville, TN

Laura A. Campbell, MA

Center for Anxiety and Related Disorders
Boston University
Boston, MA

MaryLouise Cashel, PhD

University of North Texas
Denton, TX

Heather E.P. Cattell, PhD

The Institute for Personality and Ability Testing
Walnut Creek, CA

Lawrence D. Cohn, PhD

University of Texas, El Paso
El Paso, TX

Wilson M. Compton, MD, MPE

Division of Epidemiology, Services and Prevention
Research
National Institute on Drug Abuse
Bethesda, MD

Frederick L. Coolidge, PhD

University of Colorado
Colorado Springs, CO

Linda B. Cottler, PhD

Washington University School of Medicine
Saint Louis, MO

Roger Covin, MA

University of Western Ontario
London, Ontario, Canada

Eric Dahlen, PhD

University of Southern Mississippi
Hattiesburg, MS

Robert F. Dedrick, PhD

University of South Florida
Tampa, FL

David J.A. Dozois, PhD

University of Western Ontario
London, Ontario, Canada

Jon D. Elhai, PhD

Disaster Mental Health Institute
The University of South Dakota
Vermillion, SD

Peter Farvolden, PhD, CPsych

Clinical Research Department, Section on Personality and
Psychopathology
Centre for Addiction and Mental Health
Department of Psychiatry
University of Toronto
Toronto, Ontario, Canada

Michael B. First, MD

Biometrics Research Department
New York State Psychiatric Institute
Columbia University
New York, NY

Constance T. Fischer, PhD

Duquesne University
Pittsburgh, PA

Prudence Fisher, PhD

Columbia University
New York State Psychiatric Institute
New York, NY

J. Christopher Fowler, PhD

The Austen Riggs Center
Stockbridge, MA

B. Christopher Frueh, PhD

Medical University of South Carolina
Veterans Affairs Medical Center
Charleston, SC

Carol George, PhD

Mills College
Oakland, CA

Emilija Georgievska, MA

Duquesne University
Pittsburgh, PA

Miriam Gibbon, MSW

Biometrics Research Department
New York State Psychiatric Institute
Columbia University
New York, NY

Jane Gooen-Piels, PhD

Department of Psychiatry and Behavioral Sciences
Memorial Sloan-Kettering Cancer Center
New York, NY

Paul E. Greenbaum, PhD

University of South Florida
Tampa, FL

Jessica R. Grisham, MA

Center for Anxiety and Related Disorders
Boston University
Boston, MA

Jungwon Hahn, PhD

University of Minnesota
Minneapolis, MN

Leonard Handler, PhD

University of Tennessee
Knoxville, TN

Stuart T. Hauser, MD, PhD

Judge Baker Children's Center
Harvard Medical School
Boston, MA

Susanne Hempel, PhD

University of Derby
United Kingdom

Stephen Hibbard, PhD

University of Windsor
Windsor, Ontario, Canada

Mark J. Hilsenroth, PhD

The Derner Institute of Advanced Psychological Studies
Adelphi University
Garden City, NY

Margot Holaday, PhD

University of Southern Mississippi
Hattiesburg, MS

Daniel J. Holdwick Jr., PhD

Wyoming State Hospital
Evanston, WY

Rebecca L. Jackson, MS

University of North Texas
Denton, TX

Nancy Kaser-Boyd, PhD

School of Medicine
University of California, Los Angeles
Los Angeles, CA

Joan Kaufman, PhD

Yale University
New Haven, CT

Francis D. Kelly, EdD

Brightside Treatment Center for Families and Children
West Springfield, MA

Bill N. Kinder, PhD

University of South Florida
Tampa, FL

James H. Kleiger, PsyD

Private Practice
Bethesda, MD

Kenneth A. Kobak, PhD

Dean Foundation for Health Research and Education
Healthcare Technology Systems
Research Training Associates
Madison, WI

David Lachar, PhD

University of Texas
Houston Medical School
Houston, TX

Martin Leichtman, PhD

Psychiatric Associates
Overland Park, KS

Lodi Lipien, MSPH

University of South Florida
Tampa, FL

Christopher Lucas, MD

Columbia University
New York State Psychiatric Institute
New York, NY

Betty Martin, MA

University of Tennessee
Knoxville, TN

Joseph T. McCann, PsyD, JD

United Health Services Hospitals
State University of New York Upstate Medical University
Binghamton, NY

Sarah E. Meagher, PhD

Institute for Advanced Studies in Personology and
Psychopathology
Coral Gables, FL
University of Miami
Miami, FL

Michael Melczak, MA

Duquesne University
Pittsburgh, PA

Gregory J. Meyer, PhD

University of Toledo
Toledo, OH

Jeremy Miles, PhD

University of York
United Kingdom

Theodore Millon, PhD

Institute for Advanced Studies in Personology and
Psychopathology
Coral Gables, FL

Joel S. Milner, PhD

Center for the Study of Family Violence and Sexual
Assault
Northern Illinois University
DeKalb, IL

Arpi Minassian, PhD

Department of Psychiatry
University of California, San Diego
San Diego, CA

Leslie C. Morey, PhD

Texas A & M University
College Station, TX

Robert J. Moretti, PhD

Northwestern University Medical School
C.G. Jung Institute of Chicago
Chicago, IL

Paul E. Panek, PhD

The Ohio State University at Newark
Newark, OH

William Perry, PhD

Department of Psychiatry
University of California, San Diego
San Diego, CA

John H. Porcerelli, PhD

Department of Family Medicine
Wayne State University School of Medicine
Detroit, MI

Eric C. Reheiser, BA

Center for Research in Behavioral Medicine and Health
Psychology
University of South Florida
Tampa, FL

Wendy Reich, PhD

Washington University School of Medicine
St. Louis, MO

William M. Reynolds, PhD

Humboldt State University
Arcata, CA

Jason C. Rinaldo, MA

University of Kentucky
Lexington, KY

Barry Ritzler, PhD

Long Island University, Brooklyn
Brooklyn, NY

Richard Rogers, PhD

University of North Texas
Denton, TX

Edward D. Rossini, PhD

Roosevelt University
Chicago, IL

Kathryn M. Rourke, MPE

Center for Risk Behavior and Mental Health Research
Research Triangle Institute
Research Triangle Park, NC

Sandra W. Russ, PhD

Case Western Reserve University
Cleveland, OH

Amanda E. Schweder, MS

Yale University
New Haven, CT

Daniel L. Segal, PhD

University of Colorado
Colorado Springs, CO

David Shaffer, MD

Columbia University
New York State Psychiatric Institute
New York, NY

Alissa Sherry, PhD

University of Texas, Austin
Austin, TX

Harry J. Sivec, PhD

Northcoast Behavioral Healthcare
Case Western Reserve University
Cleveland, OH

Charles D. Spielberger, PhD

Center for Research in Behavioral Medicine and Health
Psychology
University of South Florida
Tampa, FL

George Stricker, PhD

The Derner Institute of Advanced Psychological Studies
Adelphi University
Garden City, NY

Charles A. Waehler, PhD

The University of Akron
Akron, OH

Irving B. Weiner, PhD

Department of Psychiatry and Behavioral Medicine
University of South Florida
Tampa, FL

Malcolm West, PhD

Department of Psychiatry
University of Calgary
Calgary, Alberta, Canada

P. Michiel Westenberg, PhD

Leiden University
The Netherlands

SECTION ONE

**OBJECTIVE ASSESSMENT OF
PERSONALITY AND PSYCHOPATHOLOGY**

CHAPTER 1

Objective Assessment of Personality and Psychopathology: An Overview

DANIEL L. SEGAL AND FREDERICK L. COOLIDGE

HISTORICAL OVERVIEW	3	RANGE OF POPULATIONS	9
RANGE OF TESTS	4	CROSS-CULTURAL FACTORS	10
THEORETICAL CONSIDERATIONS	5	ACCOMMODATION FOR POPULATIONS WITH DISABILITIES	10
Theoretical Bases	5	LEGAL AND ETHICAL ISSUES	11
Empirical Bases	5	COMPUTERIZATION	11
Diagnostic Bases	5	FUTURE PERSPECTIVES	12
PSYCHOMETRIC CONCERNS	6	REFERENCES	12
Standardization	6		
Reliability	7		
Validity	7		

Psychological assessment pervades nearly every aspect of clinical and research work in the broad area of mental health. In general, psychological assessment techniques are designed to evaluate a person's cognitive, emotional, behavioral, and social functioning. One specific group of tests, called personality tests, strives to uncover the structure and features of one's personality, or one's characteristic way of thinking, feeling, and behaving. Another group of tests is designed to measure signs and symptoms of psychopathology or psychiatric disorders. Tests of personality and psychopathology can be further subdivided into two specific types: objective and projective. Objective tests include standardized, clear, specific items and questions that are presented to the respondent, as well as a limited choice of responses (e.g., choosing "yes" or "no" to a test item). In contrast, projective tests present novel or ambiguous stimuli and include an open-ended response format, such as a story from the respondent (an overview of projective tests is presented in Chapter 23). In this chapter, we discuss the major issues concerning the objective assessment of personality and psychopathology, including analysis of theoretical issues in test development, range of tests, cross-cultural factors, ethical and legal concerns, and the status of computerization of objective tests. We begin with a synopsis of the history of objective tests.

HISTORICAL OVERVIEW

Objective tests of personality and psychopathology received their first important recognition during World War I. With the immediate and sudden burden of large numbers of recruits, the U.S. armed services were in dire need of a means of assessing the capabilities of recruits quickly and efficiently and required a classification system for making determinations of who was mentally fit for service and who was not. The American Psychological Association volunteered its services and developed (with psychologist Lewis Terman, who developed the Stanford-Binet Intelligence test) the Army Alpha (verbal) and Army Beta (nonverbal) tests of intelligence for literate and nonliterate recruits, respectively. At the same time, American psychologist Robert S. Woodworth at Columbia University was developing a paper-and-pencil test of psychiatric fitness for the armed services, called the Personal Data Sheet (Woodworth, 1920).

The Personal Data Sheet became one of the first personality tests ever to be administered on a large basis. Woodworth, an experimental psychologist by training, had designed his test to detect Army recruits who might be vulnerable to emotional breakdowns during combat. He first created 200 questions based on neurotic symptoms described in the literature and on symptoms common to soldiers who had emotional and

4 Objective Assessment of Personality and Psychopathology: An Overview

behavioral problems in the service. The questions covered abnormal fears, excessive anxiety, depression, impulse problems, sleepwalking, nightmares, memory problems, hypochondriasis, compulsions, shyness, and depersonalization. In the final version of the Personal Data Sheet, items were included only if the symptoms occurred twice or more frequently in a group of neurotics (according to prior psychiatric diagnosis) compared to purportedly normal people. The original 200 test items were subsequently reduced to 116 “yes-no” items. The test yielded a single score, which Woodworth considered a measure of psychoneurosis. One innovative feature of the test was that it was based on norms, including education, ethnicity, and clinical versus normal samples. The average new recruit scored about 10 (10 positive psychoneurotic symptoms out of 116). Those who were deemed unfit for service generally had scores of about 30 or 40. By the time Woodworth had finished the final improvements on the Personal Data Sheet, it was too late in the war to use the test to screen recruits. The test later became known as the Woodworth Psychoneurotic Questionnaire, and it became the forerunner for later personality inventories.

One other interesting and innovative personality test proposed during this same period was the Cross-Out Test (Pressey & Pressey, 1919). Also known as the X-O Test, respondents were given lists of pleasant and unpleasant words. They were instructed to cross out or make an X over words they considered wrong, unpleasant, inappropriate, or worrisome. The Presseys believed that the resulting pattern could help categorize emotional states. They also emphasized the test could be administered in a group format.

The first commercially sold test that yielded more than one score was the Bernreuter Personality Inventory (Bernreuter, 1933), which consisted of 125 items answered in a “yes,” “no,” or “?” format and was also based on normative samples. The Bernreuter yielded six subscales: Neurotic Tendency, Self-Sufficiency, Introversion-Extraversion, Dominance-Submission, Sociability, and Confidence. The test became and remained popular in the first half of the twentieth century and was actually still commercially available (with 1938 norms!) into the 1990s (see Aiken, 1989). Interestingly, the Personal Data Sheet and the Cross-Out Test have some features that heralded some of the current objective tests of psychopathology such as the Minnesota Multiphasic Personality Inventory (MMPI) and the Symptom Checklist-90. As noted earlier, the scoring of these early tests was based on published literature, patient interviews, and intuition. As such, the items could be said to be logically keyed (i.e., the test makers used their subjective judgment based on the origin of the items and an item’s face validity to decide what answers were

pathological or not). However, none of these early tests were applied widely in the clinical setting.

Notably, the numerous challenges associated with defining and measuring personality came into much sharper focus in the 1930s with the publication of two famous books on the subject. Gordon Allport’s *Personality* (1937) and Henry Murray’s *Explorations in Personality* (1938) analyzed the topic from different perspectives, but collectively, they focused the field on the measurement of individuality and personality and paved the way for more sophisticated measurement of the constructs. Around that same time, two clinicians associated with the University of Minnesota Hospital began work on the most widely employed test in the history of objective testing, the MMPI. Starke R. Hathaway, a psychologist, and J. Charnley McKinley, a psychiatrist, wanted to provide a more efficient way, other than a one-on-one clinical interview, of obtaining a psychological diagnosis. Like Woodworth and the Presseys, they wanted to create a pencil-and-paper objective test of psychopathology that could be group administered. However, one unique feature of the MMPI was that it was not to be logically keyed but empirically keyed. The problem with logical keying for Hathaway and McKinley was that the items could be too easily faked or manipulated by the test takers. Instead, Hathaway and McKinley chose to use empirical keying where items were grouped on the empirical basis of their ability to differentiate between known psychiatric and normal groups. The full history and nature of the MMPI will be dealt with in Chapter 3, but the creation of the MMPI set the standard for innovative and empirical objective test development that has persisted to the present day.

RANGE OF TESTS

The type and nature of objective tests is astoundingly diverse. It is safe to say that an objective test has been developed to evaluate all of the major psychiatric disorders, most of the relatively uncommon disorders, and almost all of the major constructs that are relevant in clinical psychology. Major distinctions among tests are whether the test is designed for children or adults as the respondent group and whether the test evaluates mental illness (psychopathology) or normal-range personality traits. Yet another distinction is whether the test focuses on a single construct or disorder of interest (e.g., potential for child abuse, depression, or anxiety) or on multiple constructs or disorders (e.g., 10 clinical scales are included in the MMPI-2). The final types of objective tests included in this volume are structured and semistructured interviews. Although they are not classically defined as objec-

tive tests, they are objective tests from the standpoint that the questions are clear, standardized, and presented in a specified order, and responses are coded in a specified way. Notably, tests in each of the categories described here are well represented in this volume.

THEORETICAL CONSIDERATIONS

Although the referral or research question is perhaps the most important reason for the selection of an objective psychological test, it is also important to note that objective tests vary considerably as to their theoretical bases for construction, and this basis may also aid in the selection process. There are three broad methods by which tests are constructed: *theoretical*, *empirical*, and *diagnostic*. It should be noted at the outset that these methods overlap, and it could be argued that no objective test completely lacks a theoretical basis and no objective test can be judged sufficiently reliable and valid without strong empirical methods. Yet, as will be shown by the following examples, objective tests may be driven by one method more than another.

Theoretical Bases

All objective tests are constructed on some theoretical basis. A test maker must have some prior conception of what a test is designed to measure, and test items are initially picked based on some theoretical relationship the test item has to the construct being measured. However, some tests are more tightly linked to a particular theory or theoretician, and other tests have been created with a more general purpose in mind. An exceptional example of a theoretically driven test is the Millon Clinical Multiaxial Inventory-III (MCMI-III; see Chapter 9), which is based on Theodore Millon's innovative and comprehensive theory of personality disorders. Another good example is the Child Abuse Potential Inventory (see Chapter 19), which is derived from psychological factors noted in the literature that are theoretically related to child physical abuse. A final example is that of a recent operationalization of Karen Horney's tridimensional interpersonal theory of personality that postulates three basic personality styles: Moving Towards People, Moving Against People, and Moving Away From People. Coolidge, Moor, Yamazaki, Stewart, and Segal (2001) recently created a new test, called the Horney-Coolidge Type Indicator, that is based on Horney's theory and has demonstrated the usefulness of her three dimensions in the prediction and understanding of modern personality disorder features.

Empirical Bases

Empirical models, although having some theoretical basis, are usually driven by their statistical methods or procedures and they frequently use factor analyses. Factor analyses involve the testing of large groups of participants. In the initial stage of a factor analysis, a correlation matrix is examined between every item on a test with every other item on the test. The second stage is the identification of clusters of related items. The goal of a factor analysis is usually to reduce the number of items on a test to only its nonredundant items or to identify the underlying factor structure of a test. Empirical models also frequently employ discriminant studies where particular traits are demonstrated statistically to pertain more to one identified group than another (e.g., 8-year-olds as opposed to 12-year-olds, or males as opposed to females).

For example, in the 1940s, psychologist Raymond B. Cattell sought to understand the basic building blocks of personality by studying and cataloging all of the words in language that describe personality features. Based on several decades of research and factor-analytic techniques, the Sixteen Personality Factor (16PF) Questionnaire (see Chapter 4) was created. In a similar vein, Tupes and Christal (1961), in a review of thousands of English words describing personality traits, theorized through factor analyses that personality traits could be summarized by as few as five factors. Later, Costa and McCrae (1985) created a famous test of the five-factor model and claimed that it could be extended to abnormal personality traits as well. Thus, 5-factor models and 16-factor models are initially driven by only a general theoretical framework (i.e., do 5 factors underlie personality trait descriptions?). The subsequent creation of an objective test of 5 factors or 16 factors is largely *empirically* and *lexically* driven; that is, the authors were concerned only with what the factor analyses (empirically driven) revealed regarding the relationships among the words (lexically driven).

Diagnostic Bases

The main purpose of a diagnostically based objective test is to produce a psychiatric diagnosis. The Beck Depression Inventory (see Chapter 5) was created to measure the severity of depressive symptoms that the test taker is experiencing (e.g., mild, moderate, or severe levels of symptoms). It was not created to be an "official" diagnostic measure of depression, although it may be useful in that endeavor. In order to become an "official" diagnostic objective test, a measure should be aligned with a current diagnostic system like the *Diagnostic and Statistical Manual of Mental Disorders* (4th ed., text revision; DSM-IV-TR) published by the American

Psychiatric Association (2000) or the *International Statistical Classification of Diseases and Related Health Problems* (10th ed.; ICD-10) published by the World Health Organization (1992). The MMPI could certainly be considered a diagnostic test, although it is also famous for its innovative empirical underpinnings. It has been the most widely used objective diagnostic test for the past 55 years; however, it is not diagnostically aligned with either the *DSM* or ICD. The original Millon Clinical Multiaxial Inventory was created to diagnose personality disorders, and its most recent version is aligned closely with the *DSM-IV*. Certainly, all of the structured and semistructured diagnostic interviews are designed specifically to aid in psychiatric diagnosis. These instruments are based explicitly on a particular diagnostic system (usually the *DSM-IV*) and, therefore, are as useful and valid as the specific criteria are proven to be. When the diagnostic system is updated, parallel changes for the structured interviews are usually not far behind.

There are also many objective tests, such as the Myers-Briggs Type Indicator (MBTI), whose proponents vehemently reject the notion that the test is diagnostic (at least in a psychopathological sense). The MBTI does allow the classification of people into types, but the types are all considered to be variants of normal personality styles. Interestingly, even when a test like the MBTI is shown to have diagnostic implications (e.g., Coolidge, Segal, Hook, Yamazaki, & Ellett, 2001), the findings remain at odds with the original theoretical conception of the test.

In summary, all objective tests may be said to have some theoretical basis. It can be seen that this basis may vary from some general theoretical notion (e.g., five general factors underlie all personality traits) to ones that test a specific theory of personality or psychopathology (e.g., Millon's theory of personality disorder prototypes). Also, all good objective tests must have established their reliability and validity through empirical methods and studies. As such, all objective tests have an empirical basis. These empirical studies often employ factor analytic methods and discriminant validity studies. Finally, the theoretical basis of a test often determines whether it will be used diagnostically (in a psychopathological sense) or whether it will be used primarily to establish variations in normal personality, such as is the case with the 16PF (see Chapter 4).

PSYCHOMETRIC CONCERNS

What distinguishes between a psychological quiz in a popular magazine and a valid psychological measure? Classically, psychometricians have proposed that any measure must be

standardized and possess evidence of reliability and validity. All of these characteristics are actually complex and an elaboration of their features will aid clinicians and researchers in the selection of an appropriate test. At the outset of the test selection, there are some practical considerations. For example, the test user should be clear about the purpose and ultimate outcome of the testing. If the referral question is to assign a *DSM-IV* diagnosis, then tests that were not aligned to make this assessment would not be appropriate. Test users must be aware of how their objective test results will be used and interpreted. One useful place to start in the selection of a test is a resource that reviews psychological tests. Historically, one of the most popular descriptions of psychological tests has been the *Mental Measurements Yearbook* (<http://www.unl.edu/buros>).

Standardization

The word *standardization* implies that the construct being assessed is being measured in a relative way rather than an absolute way; that is, the test reflects the performance of a single respondent compared to a large group. Thus, all major objective psychological tests are standardized, which also means that there is a fixed procedure for administration and scoring, and the test has been given to many kinds of people so that statistical averages and ranges for age, grade, gender, ethnicity, and so forth, are established. A test manual, therefore, should present the characteristics of the standardization sample, including when and where the participants were tested, their characteristics (age, etc.), and how many were tested. Again, one practical consideration in the selection of an objective test is whether the test is appropriate for the people whom the test user wishes to test. For example, if a new objective test is standardized on college students from ages 18 to 22, such a test would be inappropriate to use with an older adult population. Thus, the test user should be thoroughly familiar with a test's manual and the standardization sample.

Interestingly, many objective tests are standardized on a stratified sample. Generally, this means that the standardization sample reflects the same ethnic characteristics of the U.S. population (e.g., 74% Caucasian, 11% Hispanic, 10% Black, etc.) and/or that it reflects the same residential characteristics as the U.S. population (e.g., 62% urban, 38% rural). However, the test user must keep in mind the meaning of a stratified sample. If ethnicity is an important variable on a particular objective test, then how appropriate is it to use norms that are based on a sample that is 74% Caucasian? If a test respondent is an American Indian, how appropriate are the test results if the stratified sample contained 2% American

Indians? The answer to this dilemma is that if ethnicity is deemed to be an important consideration, then a stratified sample does not mean the test results will automatically be valid. If ethnicity is an important variable, then the respondent's results should be compared to a standardization sample similar to the respondent. Objective tests do not often have separate norms based on ethnicity, so the test user should carefully note when separate norms might be an issue for a particular person.

As noted earlier, the word *standardized* also implies that the test comes with a manual that presents specific administration and scoring instructions, and these instructions should be followed diligently. A good manual should spell out the scoring procedures clearly and also provide information about how to handle missing data, prompting the patient for answers, and so forth. Scoring procedures should also be clear about the definitions of right and wrong answers and provide examples of each.

Reliability

The reliability of a test refers to its consistency. A test with good reliability means that the test taker will obtain the same test score over repeated testing, as long as no other extraneous factors have affected the test score. The reliability of a test must be established before its validity can be determined (the validity of a test is the extent to which a test accurately measures the construct that it purports to measure). The most common forms of reliability are *test-retest reliability* and *scale reliability*.

Test-retest reliability is a measure of a test's consistency over a period of time. Test-retest reliability assumes that the construct being measured is relatively stable over time, such as IQ or personality features. A good test manual should specify the sample, reliability coefficient, and the test-retest interval. Many objective tests report intervals of about one week to one month. If the trait is likely to change over time (for example, state anxiety), then test makers generally choose a shorter interval (for example, one week). Test-retest reliabilities are reported and interpreted as correlation coefficients. Test-retest reliabilities are considered to be excellent if they are .90 or better and good if they are about .80 or better. If a trait is thought to be relatively stable but the test-retest reliability coefficient for a test of that trait is around .50, then it may mean that the measure is unreliable. Perhaps there are too few questions on the test, or perhaps they are poorly worded (e.g., double negatives are difficult for nearly everyone). It is also possible that some extraneous variable or variables intervened upon the trait during the test-retest interval. One final problem for the interpretation of test-retest

reliabilities is that they may be spuriously high because of practice effects or memory effects. A respondent may do better on the second testing because the trait being assessed improves with practice. Also, some people may respond similarly to a test because they remember many of the answers that they gave on the test earlier. One possible solution to this problem is the use of *alternate forms*. Although this is not common among objective tests, some tests do come with an alternate form. If a test user is interested in a trait's change over time and is worried about practice or memory effects, then alternate forms of the test may be given.

Scale reliability (commonly called internal consistency) is a measure of how well the items on a test relate to each other. The most common statistic for scale reliability is Cronbach's (1951) coefficient alpha, which has become the virtual standard of scale reliability in objective testing. One intuitive way of interpreting Cronbach's alpha is to view it as kind of an average of all of the correlations of each item with every other item on a test. The alpha coefficient is interpreted much like a correlation coefficient (i.e., it ranges from 0.00 to 1.00). Values above approximately .80 are considered good and generally reflective of reliable (internally consistent) scales. The alpha coefficient is dependent, however, upon two other variables. First, all things being equal, shorter tests (less than about eight items) will yield lower alpha coefficients than longer tests. This also means that scales or tests with seven or less items may possess reliability, but it may not be reflected in the alpha coefficient. Scales or tests of 30 or more items will usually yield alpha coefficients around .90. Second, the alpha coefficient is dependent upon a high first factor concentration (i.e., the scale or test is measuring a unidimensional concept or trait). For example, if there is a scale measuring psychoticism and the items were derived to measure equally two major components of psychoticism (aberrant thinking and social withdrawal), then the coefficient alpha will be lower than it will be for a different scale of psychoticism that measures only one underlying concept.

Validity

Validity refers to the extent to which a test accurately assesses the construct it purports to measure. Essentially, validity has to do with the meaningfulness and usefulness of the specific inferences made from test scores. In the previous example about a psychoticism scale, the question of its validity would be whether it actually measures psychotic traits or psychotic behavior. There is an old adage that states a test can be reliable (i.e., stable and reproducible) but not valid, but a test cannot be valid without first being reliable. The question of a test's validity is critically important and complex. The va-

lidity of any psychological test cannot be absolutely established but only relatively established because there is no gold standard of validity in psychological science. There are also many aspects of a test's validity including *face*, *content*, *criterion*, and *construct*.

Face validity is perhaps the simplest of the four types of validity. Face validity can refer to a single item or to all of the items on a test, and it indicates how well the item reveals the purpose or the meaning of the test item or the test itself. For example, the test item "Recently I have thought of killing myself" has obvious face validity as an item measuring suicidal ideation. The downside of items on tests with clear face validity is that they are more subject to being manipulated by respondents, either to deny or hide problems or to malingering or exaggerate problems. Some psychometricians like tests that lack face validity but still possess general validity. Tests or items that still measure what they purport to measure but lack face validity are harder for respondents to manipulate. For example, the item "I believe in the second coming of Christ" appeared on the first version of the MMPI (Hathaway & McKinley, 1943) and loaded on the Depression scale. Because most of the people in the original normative sample of the MMPI were good Christians, only a depressed Christian would think Christ was not coming back. Obviously, this question would not be a good one for people of many other faiths and beliefs. Thus, although a lack of face validity may have some attractive features, items that have some face validity may, in the long run, make for a better test.

The *content validity* of a test refers to the adequacy of sampling of content across the construct or trait being measured. Given the published literature on a particular trait, are all aspects of that concept represented by items on the test? Let us use the example of the conduct disorder in childhood. If a literature search reveals two major aspects of a conduct disorder, namely delinquency and aggression, then the items on the tests should measure these two aspects in relatively equal proportion. Some test makers also rely on experts in that field. The test makers will devise a means of summarizing what the experts claim to be the nature of a particular trait, and then create the test items to reflect what the experts' consensus was about that trait. The items measuring a trait should appear in equal proportion to what the literature search reveals or what the experts claim about that particular trait. Are there cases where the items might become unbalanced? Yes! An imbalance may occur as the test makers are checking the test's scale reliability. Statistics software used to calculate Cronbach's alpha typically allows the evaluation of each item's reliability with the overall scale's reliability. Sometimes the most unreliable items are those that are tapping only

one underlying concept of the construct trait being measured. In the present example, a test maker may find that the least reliable items may come predominately from the aggression concept. If the test maker eliminates most of the unreliable items primarily from the concept of aggression, then the test maker is unbalancing the content of the test of conduct disorder. The test maker is sacrificing content validity on the altar of coefficient alpha. In this case, perhaps the test maker might consider rewording the aggression questions to make them more reliable.

Criterion validity (also called predictive or concurrent validity) refers to the comparison of the scores on a test with some other kind of external measure of performance. The other measure should be theoretically related to the first measure, and their relationship can be assessed by a simple correlation coefficient. Some psychometricians further divide criterion validity into predictive or concurrent validity. With predictive validity, the new test is given to a group of participants who are followed over time to see how well the original assessment predicts some important variable at a later point in time. For example, a new measure of college success is initially given to high school seniors. Then, after their first, second, and subsequent years in college, their success in college is measured by a different objective standard (for example, GPA). To establish the new measure's predictive validity, there should be a substantial correlation (e.g., $r > .50$) between the new measure and subsequent college GPAs. Thus, in predictive validity, a test is given first, and other measures are subsequently taken and correlated with the original test.

In concurrent validity (which is far more common), a proposed test is given to group of participants who complete other theoretically related measures concurrently (meaning at the same point in time). As an example, if Jones creates a new measure of conduct disorder, then Jones might give the new measure at the same time as Jones administers the Smith Conduct Disorder scale, which should be a well-known or already validated measure of conduct disorder. Jones can demonstrate the concurrent validity of the new Jones Conduct Disorder test if there is a substantial correlation with the Smith Conduct Disorder scale. How can Smith demonstrate concurrent validity if Smith was the first to create a conduct disorder scale? Unfortunately, this is not done as easily. Smith must use other forms of validity (other than concurrent) if there are no other known measures of conduct disorder. This problem is particularly thorny for diagnostic measures in psychology. Because there are no definitive biological markers and no blood tests used for the diagnosis of any mental disorder, this lack of a so-called gold standard for diagnostic accuracy makes it difficult to assess the criterion-related validity of any psychological test (Segal & Coolidge, 2001).

Construct validity refers to the extent to which a test captures a specific theoretical construct or trait, and it overlaps with some of the other aspects of validity. This requires a test to be anchored in a conceptual framework or theory that clearly delineates the meaning of the construct, its uniqueness, and its relationship to other variables measuring similar domains. Psychometricians typically assess construct validity by giving other measures of a trait along with the new proposed measure of a trait and then testing prior hypothesized relationships among the measures. In the example of the new Jones Conduct Disorder scale, Jones might also give measures of attention deficit hyperactivity disorder (ADHD), altruism, and executive functions deficits (organization and planning problems). Jones might hypothesize that if the new measure of conduct disorder possesses construct validity, then it should positively correlate with ADHD (because the literature suggests a strong comorbidity between the two disorders), negatively correlate with altruism (which might be a clinical intuition without evidence from the literature), and positively correlate with executive function deficits (also consistent with the literature). Note that the hypothesized relationships include a mixture of what the construct (in this example, conduct disorder) should show a meaningful positive relationship to and show a meaningful negative relationship to. The new measure should also show weak relationships to other constructs that are theoretically unrelated to it (e.g., conduct disorder and eye color). The type of relationships found, should they be consistent with expected results, help to establish the construct validity of the new test.

Interestingly, there is no single method for determining the construct validity of a test. Usually many different methods and approaches are combined to present an overall picture of the construct validity of a test. Besides the correlational approach described earlier, another frequently used method is factor analysis. The new test is given to a large group of participants (for a proper factor analysis, the number of participants should be at least 10 times the number of items on the test) and the results are analyzed to see how many different constructs or dimensions underlie the measure. In the previous example of a conduct disorder test, the factor analysis should reveal two underlying constructs, delinquency and aggression. If the factor analysis reveals only one main factor, then it might mean that the construct of a conduct disorder is a unitary concept, and perhaps, those who are delinquent are often aggressive and vice versa. However, the factor analysis might reveal a three-factor structure: delinquent-nonpersonal (vandalism to structures), delinquent-personal (damage that hurts people), and aggression. A factor analysis helps a test maker clarify the underlying nature of a new test, and it can help the test maker in modifying the new test to make it better

(e.g., more comprehensive, more consistent with the literature, etc).

Another method of establishing a test's construct validity is discriminant validity. For example, a group of repeat male juvenile offenders should score higher on the new conduct disorder scale than a group of choirboys. School bullies should score higher than their victims on the conduct disorder scale. All of these methods and designs should be used to establish the construct validity of a test. A test manual should report all of the evidence for a test's construct validity, and the more evidence, the better, because as stated earlier, there is no single or absolute measure of a test's construct validity.

RANGE OF POPULATIONS

The range of populations served by objective tests can be subsumed under several broad categories: child versus adult focus and psychopathology versus normal-range focus. Most tests fit neatly into one combination (e.g., a child psychopathology test; an adult test of normal personality). Within these broad categories, however, objective tests have been applied to measure constructs of interest in widely diverse and numerous populations (e.g., medically ill persons, psychiatric inpatients and outpatients, persons with almost every kind of specific form of mental illness, war veterans, spousal abusers, job applicants, self-mutilators, and persons of different cultural and ethnic backgrounds). Indeed, there are countless more diverse populations that have been assessed using objective tests. One important issue to be noted here is that when one selects a test for use in a particular population, one must ensure that there is adequate evidence for reliability and validity of the test in that unique population.

What are the typical ways in which objective tests are used? Notably, objective tests have been used in many different venues and for many different purposes. Application of objective tests can be broadly subsumed under three, non-mutually exclusive areas: research, clinical, and training use. In the research domain, for example, objective tests are used (typically as part of a more thorough assessment process) to classify participants into diagnostic groups so that etiology, comorbidity, and interventions (among other topics) can be investigated for a particular diagnosis or group of diagnoses. Objective tests are also widely used as outcome measures for intervention studies. For example, the Beck Depression Inventory-II (BDI-II; see Chapter 5) and the Hamilton Depression Rating Scale (HAM-D; see Chapter 7) have widespread application in studies of depression.

In the clinical setting, objective tests may be used as a way to ensure standardized initial assessments. For example, each

client may be given a standard battery of tests at the initial intake. Objective testing is also commonly conducted to facilitate treatment because test data can help to clarify diagnostic or personality features in a complex case, to assist in case conceptualization, and to provide data used to monitor and evaluate treatment progress over time (e.g., the client may be asked to complete a symptom checklist before each session, and scores can be plotted and tracked over time). Use of objective tests in the forensic setting has increased dramatically in recent years, and such tests are frequently admitted as evidence in court proceedings.

Use of objective tests for training in the mental health field is an important application because the test output can help beginning clinicians more thoroughly understand important dimensions of personality and psychopathology that may substantially influence case conceptualization and intervention. Structured interviews are particularly conducive to training in mental health because interviewers have the opportunity to learn (through repeated administrations) specific questions and follow-up probes used to elicit information and evaluate specific diagnostic criteria of the *DSM-IV*. Modeling one's own questions and flow of the interview from a well-developed structured interview can be an invaluable source of training for the mental health clinician (Segal & Coolidge, 2003).

CROSS-CULTURAL FACTORS

All of the major objective tests have been applied in different cultural and subcultural settings and most have been translated into different languages. Important considerations regarding application of a test in a different culture include the relevance of the diagnostic or conceptual model on which the test is based and the relevance of individual items in the translated version. For example, a test used to aid in psychiatric diagnosis according to the *DSM-IV* diagnostic system is only valid in a culture or subculture if the *DSM-IV* system itself is valid in that culture or subculture. Although the *DSM-IV* strives to be relevant and useful in diverse countries across the globe and is considered the standard in North America and western Europe, it simply is not the prevailing model in less developed countries (Segal & Coolidge, 2001). Regarding specific items of tests, consideration must be made on an item-by-item basis as to whether the item is relevant and appropriate in a given culture. As an example, a test asking respondents if they "are currently the quarterback of the Buffalo Bills" may be relevant in much of the United States but will be confusing to many in foreign cultures and simply cannot be translated well.

Another interesting cross-cultural research issue is the interpretation of cultural differences on a given test. Let's assume, for example, that a personality test is given to citizens of the United States and India, and that the Americans score higher on Scale X. How can one interpret this group difference? One possibility is that the finding is veridical; in other words, that it reflects a true difference between the cultures on whatever it is that Scale X measures. Another equally valid possibility however, is that the translated item used in the two cultures means something *different* to each culture. In this case, the item does not actually measure the same construct in the two cultures, and thus group comparisons become meaningless.

ACCOMMODATION FOR POPULATIONS WITH DISABILITIES

All of the major objective tests provide for at least minimal accommodation for certain disabled groups. For example, big print versions of some tests are available for visually impaired test takers, and adaptive technology may be used to help a respondent read the test items. Audiotaped versions of some tests (e.g., MCMI-III, MMPI-2, Personality Assessment Inventory) are available as well. Moreover, questions on the typical self-report paper-and-pencil objective test may be administered orally to a respondent who experiences difficulty with the standard format and, in most cases, the responses can be considered valid. An exception to this general rule concerns those tests that have a distinct focus on psychopathology (rather than normal traits). For these tests, having the respondent share his or her responses to some test items with the examiner may substantially alter the context in which the test is normally given. Indeed, some respondents may be prone to minimize the presence of certain dysfunctional traits or experiences if they must verbalize their response to the examiner, whereas they may feel less pressure to deceive when rating the item in private. Notably, no major objective tests have a braille version, and translation into American Sign Language (presented via videotape) is rare.

Objective assessment of those with mental retardation or other cognitive disabilities presents another challenge because self-report tests require that the respondent be able to read and comprehend the items. The exact reading level required for a particular test is usually stated in the test manual. Assessment of medically ill populations is complicated by the facts that emotional, behavioral, or cognitive symptoms measured by a test may be caused by the medical illness or the medications used to treat the condition, or the symptoms may be exacerbated by the stress of having a serious medical

problem. Finally, computerized administration of tests (discussed fully in a following section) may be useful for individuals with motor skills deficits who are more comfortable using a computer keypad rather than a pen or pencil.

LEGAL AND ETHICAL ISSUES

Although objective tests are an integral part of psychological research, clinical work, and training, tests also carry significant legal and ethical responsibilities for the test user. The “Ethical Principles of Psychologists and Code of Conduct” (American Psychological Association, 2002) highlights many of the potential ethical issues regarding the use of psychological tests (regardless of the objective or projective nature of the test). Another important reference is the *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 1999). Finally, test manuals also typically spell out requirements for using the test appropriately. Some broad ethical and legal standards are described next.

Not surprisingly, it is the test user who is obligated to select the most appropriate instrument for a given application. Test users are further obligated to understand the purpose of the testing, its probable consequences, and the necessary procedures to ensure effectiveness and reduce test biases. Users of any test must have appropriate training in the purpose, administration, format, scoring, and interpretation of the test. Test users should understand the psychometric properties of the test, the normative data for the test, and the nature and impact of measurement error. Users should understand appropriate uses of the test and only use tests for their designated and validated purposes. In the clinical setting, the limits of confidentiality should be discussed prior to any assessment, and feedback about testing results should be presented to the respondent in a manner that the person can understand and minimizes the potential for harm.

Test users should also remain alert to the ethical issues that arise specifically regarding computer-generated narrative (or interpretive) reports that are primarily used in the clinical setting. Whereas this trend presents clear advantages to clinicians in regard to time management, one ethical concern is that the decision rules used by the computer to generate the report are sometimes not explicitly stated. Another concern is that, given the ease of utilizing these narratives, it can be tempting to substitute computer-generated interpretations for comprehensive, integrative reports. This practice is a clear violation of ethical standards and likely also will result in a disservice to the testing client. Specifically, Standard 9.01

(American Psychological Association, 2002) mandates clinicians to integrate additional available data, such as behavioral observations or other clinical evidence, into their overall evaluation, a dimension totally neglected by computer-generated reports.

Another ethical concern is the potential availability of interpretive programs to individuals lacking the proper professional qualifications to administer and interpret the tests. Interestingly, concerns such as those noted here contributed to the publication of ethical guidelines that specifically address the use of computer-assisted testing (American Psychological Association, 1986). These guidelines were designed to help clinicians utilize computer testing tools more appropriately and maintain the integrity of psychological assessments. By limiting use of computer programs to individuals with the necessary qualifications, these guidelines strive to protect the public from the use of tests by unqualified individuals. It is clearly the responsibility of the individual clinician to evaluate carefully computer programs they may use and to resist the temptation to blindly or passively accept computer-generated narratives as fast and easy substitutes for more thorough and integrated evaluations. Rather, computer-generated narratives are best viewed as a source of hypotheses about the test taker that require further scrutiny and evaluation.

A final important ethical issue is that objective tests should *never* be used as the sole basis for making a psychiatric diagnosis or drawing any important conclusion. Rather, testing data should always be combined with additional sources of information (e.g., clinical interview, behavioral assessments) that will provide a more complete picture of the respondent's strengths, limitations, and experiences.

Given the specific strength and limitations of different tests, the choice of a particular instrument is often complex and influenced by many factors. Among the many variables a test user has to weigh are: the purpose of the assessment, the psychometric properties of the instrument, his or her experience with the assessment, and the strengths and limitations of the person to be assessed. Accordingly, it would be misleading to conclude that one type of instrument or test is superior to the other. Thus, the decision regarding selection of a test is based on whether a specific type of test is more appropriate in a specific situation with a specific client (or research participant) for a specific purpose.

COMPUTERIZATION

Computerized personality assessment has a long history dating back to the early 1960s with the first computer program

written to interpret the original MMPI (Butcher, 1995). However, computerization of diverse personality and psychopathology tests has grown exponentially since the 1980s (when personal computers became commonplace) and is now the rule rather than the exception. Administration of test items via computer is common for the major self-report tests. Computer scoring and profiling of results is also available for most of the major tests, either by a program that can be stored on the test user's personal computer or through the mechanism of mailing or faxing test sheets to the publisher for scoring. Scoring by computer is ideal because it eliminates scoring errors (once the data are entered correctly). Some scoring programs use optical scanners to read the test responses and enter them automatically (thus eliminating all manual data entry problems).

In many cases, sophisticated computer programs are also available that provide the test user with a narrative interpretation of the test scores. Computer-assisted narrative reports are available for most of the major objective personality inventories, and they are widely used in clinical practice, although use of such reports is subject to significant ethical debate (discussed earlier). Most computer programs (typically purchased from the test publisher) are able to score and interpret an unlimited number of cases, whereas less commonly, the test user pays for scoring and interpretation on a case-by-case basis. An interesting trend for the future will be the use of the Internet to allow for computerized self-administration of objective tests in diverse languages; the 16PF (see Chapter 4) already has applied some of this emerging technology. Finally, the point should be made that computerization is not desirable for all forms of testing. Indeed, some of the semistructured interviews are sufficiently complex and require a significant amount of clinical judgment and experience so that computerized administration is neither desirable nor possible.

FUTURE PERSPECTIVES

As discussed in this chapter, objective psychological tests contribute strongly to research, clinical services, and training in mental health, and there is no indication that the widespread application of testing will decrease in the future. As we look toward the future, it will be important for the major objective tests to continue to be refined and updated as society and the field of psychology evolves. Many of the major psychological tests have already undergone several significant revisions since their initial development, typically including more current normative data and revamping of items to match current conceptualizations of psychiatric disorders. It is probably best to think of test development as an ongoing

process, when there is never a final version of any test that will withstand all the new developments in the field that will come over time. With this caveat in mind, however, if the past is any indication of developments to come, the future looks extraordinarily bright and interesting for the objective assessment of personality and psychopathology.

REFERENCES

- Aiken, L.R. (1989). *Assessment of personality*. Needham Heights, MA: Allyn & Bacon.
- Allport, G.W. (1937). *Personality: A psychological interpretation*. New York: Holt, Rinehart, & Winston.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.
- American Psychological Association. (1986). *Guidelines for computer-based test interpretations*. Washington, DC: Author.
- American Psychological Association. (2002). Ethical principles of psychologists and code of conduct. *American Psychologist*, 57, 1060–1073.
- Bernreuter, R.G. (1933). The theory and construction of the personality inventory. *Journal of Social Psychology*, 3, 387–405.
- Butcher, J.N. (1995). How to use computer-based reports. In J.N. Butcher (Ed.), *Clinical personality assessment: Practical approaches* (pp. 78–94). New York: Oxford University Press.
- Coolidge, F.L., Moor, C.J., Yamazaki, T.G., Stewart, S.E., & Segal, D.L. (2001). On the relationship between Karen Horney's tripartite neurotic type theory and personality disorder features. *Personality and Individual Differences*, 30, 1387–1400.
- Coolidge, F.L., Segal, D.L., Hook, J.N., Yamazaki, T.G., & Ellett, J.A.C. (2001). An empirical investigation of Jung's psychological types and personality disorder features. *Journal of Psychological Type*, 58, 33–36.
- Costa, P.T., Jr., & McCrae, R.R. (1985). *The NEO Personality Inventory manual*. Odessa, FL: Psychological Assessment Resources.
- Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297–334.
- Hathaway, S.R., & McKinley, J.C. (1943). *The Minnesota Multiphasic Personality Inventory*. Minneapolis: University of Minnesota Press.
- Murray, H.A. (1938). *Explorations in personality*. New York: Oxford University Press.
- Pressey, S.L., & Pressey, L.W. (1919). Cross-Out Test, with suggestions as to a group scale of the emotions. *Journal of Applied Psychology*, 3, 138–150.