

ELENA L. GRIGORENKO
EDITOR

Multicultural Psychoeducational Assessment



SPRINGER PUBLISHING COMPANY

Multicultural Psychoeducational Assessment

Edited by

ELENA L. GRIGORENKO, PhD

 **SPRINGER PUBLISHING COMPANY**
New York

Dr. Elena L. Grigorenko received her PhD in general psychology from Moscow State University, Russia, in 1990, and her PhD in developmental psychology and genetics from Yale University in 1996. Currently, Dr. Grigorenko is Associate Professor of Child Studies, Psychology, and Epidemiology and Public Health at Yale and Adjunct Professor of Psychology at Columbia University and Moscow State University (Russia). Dr. Grigorenko has published more than 200 peer-reviewed articles, book chapters, and books. She has received awards for her work from five different divisions of the American Psychological Association: the Gardner Lindzey Dissertation Award in General Psychology, Sigmund Koch Early Career Award in Theoretical and Philosophical Psychology, Berlyne Early Career Award for Creative Achievement in Psychology of the Arts, Boyd McCandless Early Career Award in Developmental Psychology, and Richard E. Snow Early Career Award in Educational Psychology. In 2004, she won the APA Distinguished Award for an Early Career Contribution to Developmental Psychology. Dr. Grigorenko's research has been funded by NIH, NSF, DOE, Cure Autism Now, the Foundation for Child Development, the American Psychological Foundation, and other federal and private sponsoring organizations.

Copyright © 2009 Springer Publishing Company, LLC

All rights reserved.

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, or otherwise, without the prior permission of the publisher or authorization through payment of the appropriate fees to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400, fax 978-646-8600, info@copyright.com or on the Web at www.copyright.com.

Springer Publishing Company, LLC
11 West 42nd Street
New York, NY 10036
www.springerpub.com

Acquisitions Editor: Philip Laughlin
Project Manager: Julia Rosen
Cover design: Steve Pisano
Composition: Apex CoVantage, LLC

Ebook ISBN: 978-0-8261-0102-0

09 10 11 12 / 5 4 3 2 1

The author and the publisher of this Work have made every effort to use sources believed to be reliable to provide information that is accurate and compatible with the standards generally accepted at the time of publication. The author and publisher shall not be liable for any special, consequential, or exemplary damages resulting, in whole or in part, from the readers' use of, or reliance on, the information contained in this book. The publisher has no responsibility for the persistence or accuracy of URLs for external or third-party Internet Web sites referred to in this publication and does not guarantee that any content on such Web sites is, or will remain, accurate or appropriate.

Library of Congress Cataloging-in-Publication Data

Grigorenko, Elena L.

Multicultural psychoeducational assessment / edited by
Elena L. Grigorenko.

p. cm.

Includes bibliographical references and index.

ISBN 978-0-8261-0101-3 (alk. paper)

1. Psychological tests. 2. Educational tests and measurements.
3. Cultural pluralism. I. Title.

BF176.G75 2009

150.28'7—dc22 2008050932

Printed in the United States of America by Hamilton Printing.

Contents

Contributors xi

Preface by Elena L. Grigorenko xv

- 1** How Universal Are Test Development and Use? 1
Thomas Oakland
- 2** Equitable Assessment Practices in Diverse Contexts 41
Elias Mpofu and Samuel O. Ortiz
- 3** Respecting Local, Cultural Contexts for Assessment Practice in an Era of Globalization 77
Val Klenowski
- 4** Conceptualizing Developmental Assessment Within Africa's Cultural Settings 95
A. Bame Nsamenang
- 5** Assessing the Environment of Children's Learning: The Developmental Niche in Africa 133
Sara Harkness, Charles M. Super, Oumar Barry, Marian Zeitlin, and Jennifer Long
- 6** Assessing Competencies in Reading and Mathematics in Zambian Children 157
Steven E. Stemler, Florence Chamvu, Hilary Chart, Linda Jarvin, Jackie Jere, Lesley Hart, Bestern Kaani, Kalima Kalima, Jonna Kwiatkowski, Aidan Mambwe, Sophie Kasonde-N'gandu, Tina Newman, Robert Serpell, Sara Sparrow, Robert J. Sternberg, and Elena L. Grigorenko

- 7** Assessing Mother Tongue in the Era of Globalization: Promise and Challenge 187
Elena L. Grigorenko, Kelly Nedwick, Dinah Kwadade, Erik Boro, Lesley Hart, Tina Newman, and Linda Jarvin
- 8** The Logic of Confidence and the Social Economy of Assessment Reform in Singapore: A New Institutional Perspective 213
David Hogan, Phillip A. Towndrow, and Kim Koh
- 9** Instructional and Assessment Practices in Singapore 253
David Hogan, Phillip A. Towndrow, and Kim Koh
- 10** Considerations for Developing and Adapting Language and Literacy Assessments in Arabic-Speaking Countries 287
Saleh Shaalan
- 11** The Behavioral Characteristics of Kindergarten Gifted Children in Saudi Arabia: Construction and Validation of a Scale 315
Usama M. A. Ibrahim and Abdullah M. Aljughaiman
- 12** Developing Culture-Specific Assessments 335
Alexander G. Shmelyov and Anna S. Naumenko
- 13** The Use of Foreign Psychodiagnostic Inventories in Differing Methodological Contexts 351
Tatiana V. Kornilova and Sergey A. Kornilov
- 14** Adapting Existing Abilities and Competencies Assessment Devices to Different Cultures 375
Márcia Regina F. de Brito
- 15** The Challenge of Measuring Abilities and Competencies in Hispanics/Latinos 417
Antonio E. Puente and Antonio N. Puente

- 16** Considering Language, Culture, and Cognitive Abilities: The International Translation and Adaptation of the Aurora Assessment Battery 443
Mei T. Tan, Abdullah M. Aljughaiman, Julian G. Elliott, Sergey A. Kornilov, Mercedes Ferrando-Prieto, David S. Bolden, Karen Adams-Shearer, Hilary E. Chart, Tina Newman, Linda Jarvin, Robert J. Sternberg, and Elena L. Grigorenko
- 17** Conclusions: Assessment in an Era of Globalization 469
Peter Tymms and Robert Coe
- Index 487**

Contributors

Karen Adams-Shearer, BA

Hartlepool Borough Council
UK

Abdullah M. Aljughaiman, PhD

King Faisal University
Saudi Arabia

Oumar Barry, PhD

University Cheikh Anta Diop of Dakar
Senegal

David S. Bolden, EdD

Durham University
UK

Erik Boro, BA

Ministry of Education
Ghana

Florence Chamvu, MA

University of Zambia
Zambia

Hilary Chart, MA

Stanford University
USA

Robert Coe, PhD

Durham University
UK

Márcia Regina F. de Brito, PhD

State University of Campinas
Brazil

Julian G. Elliott, PhD

Durham University
UK

Mercedes Ferrando-Prieto, PhD

University of Murcia
Spain

Jacquiline Folotiya-Jere, MA

University of Zambia
Zambia

Elena L. Grigorenko, PhD

Yale University
USA

Sara Harkness, PhD, MPH

University of Connecticut
USA

Lesley Hart, PhD

Yale University
USA

David Hogan, PhD

Nanyang Technological
University
Singapore

Usama M. A. Ibrahim, PhD

King Abdulaziz and his
Companions Foundation
for the Gifted
Saudi Arabia

Linda Jarvin, PhD

Tufts University
USA

Bestern Kaani, PhD

University of Zambia
Zambia

Kalima Kalima, PhD

University of Zambia
Zambia

Sophie Kasonde-N'gandu, PhD

University of Zambia
Zambia

Val Klenowski, PhD

Queensland University of Technology
Australia

Kim Koh, PhD

Nanyang Technological University
Singapore

Sergey A. Kornilov, BSc/BA

Moscow State University
Russia

Tatiana V. Kornilova, PhD

Moscow State University
Russia

Dinah Kwadade, MEd

Ministry of Education
Ghana

Jonna Kwiatkowski, PhD

Emmanuel College
USA

Jennifer Long, PhD

University of Illinois at Chicago
USA

Aidan Mambwe, PhD

University of Zambia
Zambia

Elias Mpfu, PhD

University of Sydney
Australia

Anna S. Naumenko, PhD

Moscow State University
Russia

Kelly Nedwick, MA

Yale University
USA

Tina Newman, PhD

Yale University
USA

A. Bame Nsamenang, PhD

Yaounde University
Cameroon

Thomas Oakland, PhD

University of Florida
USA

Samuel O. Ortiz, PhD

St. John's University
USA

Antonio E. Puente, PhD

University of North Carolina
Wilmington
USA

Antonio N. Puente, PhD

University of North Carolina
Wilmington
USA

Robert Serpell, PhD

University of Zambia
Zambia

Saleh Shaalan, CCC-SLP

University College
London
UK

Alexander G. Shmelyov, PhD
Moscow State University
Russia

Sara Sparrow, PhD
Yale University
USA

Steven E. Stemler, PhD
Wesleyan University
USA

Robert J. Sternberg, PhD
Tufts University
USA

Charles M. Super, PhD
University of Connecticut
USA

Mei T. Tan, MA
Yale University
USA

Phillip A. Towndrow, PhD
Nanyang Technological University
Singapore

Peter Tymms, PhD
Durham University
UK

Marian Zeitlin, PhD
GENSEN EcoYoff Living & Learning
Center
Senegal

Preface

Elena L. Grigorenko

There is always a mix of joy and sadness in completing an edited volume. It is quite delightful to realize that a large-scale project is done; yet, it is quite poignant to complete something that has taken quite a large portion of one's time, devotion, and energy and let it make its own way in the world, even though that is where, by its very nature, it belongs.

This volume is a diverse and exciting collection of contributions sampling from different theoretical perspectives on assessment, exemplifying assessments of different psychological processes and functions, and illustrating international views on assessments and the roles of assessments in different countries around the world. The issues that are raised here are complex and various, but linked together by their shared contemplations on matters integral to the meaning, validity, fairness, and interpretability of today's assessments.

The word *globalization*, although around since the 1960s, has received much attention and use only recently, within the last decade or so. These days we hear and see news of this world everywhere and virtually every day. The reason is that globalization is unfolding now. So, what is it that is happening?

There are many definitions of globalization. Palmer (2002), for example, defines it as "the diminution or elimination of state-enforced restrictions on exchanges across borders and the increasingly integrated and complex global system of production and exchange that has emerged as a result" (p. 1). Notice that what is exchanged is not specified. In the context of this book, the objects of exchange are assessments.

People have always tried to size each other up, whether in an intellectual debate (e.g., in Socrates' defiant defense during his trial by the prominent Athenians) or an armored fight (e.g., Achilles and Hector in Troy). And often the consequences of such assessments were very costly,

such as in these cases. “Sizing up” became more structured and ordered when psychological and educational assessments came around in the late 19th to early 20th century. And in today’s globalized (or globalizing) world, assessment still remains and will remain the main method of sizing people up. Assessment in the globalizing world assumes an agreement on the reasons, methods, and procedures for this attempt to size people up globally, that is, on a “world scale.” Thus, the questions are: why, with what, and how to assess.

Answers to these questions are complex. This volume, collectively, endeavors to address all of these questions, examining assessments across the spectrum of abilities (from disabilities to giftedness), across multiple continents and cultures, across multiple languages, and across multiple domains of functioning. While the book does not offer any final or absolute conclusions, it offers a diverse collection of well-articulated approaches illustrated with data. I am profoundly thankful to the authors not only for their willingness to write for the book, but also for their patience with revisions, suggestions, and their search to find ways to express themselves in English (note that a substantial portion of the contributors to the volume are foreign-language speakers). And last but not least, I express my sincere appreciation for the support, creativity, and camaraderie I received from Springer Publishing’s Senior Editor, Phil Laughlin.

But now, the book is done. And so we will wait to see how it makes its way. We hope, globally.

REFERENCE

Palmer, T. G. (2002). Globalization is grrrrreat! *Cato’s Letters*, 2, 1–6.

1

How Universal Are Test Development and Use?

THOMAS OAKLAND

This chapter discusses the status of test development and use internationally. The chapter opens with a discussion of qualities that impact test availability and then discusses the results of international surveys that provide information on the status of test development and use with children and adults. National and international technical standards and guidelines for test development and use are discussed. Guidelines for adapting tests are highlighted. Ethical issues associated with test development and use as well as national, regional, and international codes of ethics are discussed. Traditional and emerging models used to define and describe disorders are identified. The World Health Organization's International Classification of Functioning, Disability and Health is described in some detail. The impact of external and internal conditions that will influence the futures of test development and use are discussed. The chapter concludes with a description of efforts in three regions to promote test development and use that exemplify many of the themes discussed herein.

This chapter focuses on the international use of standardized tests commonly used in the behavioral sciences, industry, and education, namely those that have well-established standards for administration and scoring and generally provide scores that are interpreted normatively (e.g., comparing one person's scores to those of a norm group).

This chapter does not focus on teacher-made tests, those used exclusively in health care, the use of clinical observations, or other informal and unstandardized testing methods.

Test use is universal. Tests are used in virtually every country, with newborns through the elderly, and most commonly with students (e.g., persons engaged in formal education from preschool through graduate school). Tests are used within the behavioral sciences to describe current behaviors and other qualities, estimate future behaviors, assist guidance and counseling services, establish intervention methods, evaluate progress, screen for special needs, diagnose disabling disorders, help place persons in jobs or programs, and assist in determining whether persons should be credentialed, admitted/employed, retained, or promoted. Tests also are used widely in research and for various administrative and planning purposes. Tests may be administered to groups or individually to assess aptitudes, achievement, adaptive behavior, intelligence, language, motor, perception, personality, and other personal qualities (Oakland, 2004).

QUALITIES THAT IMPACT TEST AVAILABILITY AND USE

Although tests are widely available and enjoy widespread use in many countries, their availability and use differ considerably among the more than 220 countries. The need for tests spurs test development. Test development may occur only when a need for tests is recognized. Educational institutions and those who work in them often constitute the largest consumers of tests within a country. Thus, the development and financing of a country's educational system strongly impact test need and development. Test use and thus test development generally are stronger in countries with well-established and universally attended public education systems that include elementary, secondary, and tertiary education. Countries that lack this infrastructure or are unable to fund it adequately are less likely to need tests.

A country must have a testing industry with sufficient financial and personnel resources to support the development and use of standardized tests. Test development can be expensive, with some costing more than \$500,000. Thus, companies that develop and market tests can expect a sufficient return on their investment to warrant this expense. Additionally, personnel with expertise in organizational management as well as psychometrics and test development must be available to assume leadership for developing and marketing tests.

Test development and use also assume the presence of a sufficiently large and stable market. Most tests are developed in response to requests by professionals to assist them in their work. Test consumers commonly include educators, counselors, management specialists, medical specialists, occupational therapists, physical therapists, psychologists, social workers, speech pathologists, and other professionals. An infrastructure that supports test development and use also requires the presence of a fairly large number of educational programs at the tertiary level that prepare professionals with skills associated with administering, scoring, and interpreting tests. Test development is viable only when there is a sufficiently large workforce that commonly purchases and uses tests. Thus, test development and use most commonly occur in countries that enjoy stable financial support for education, have well-developed economies that depend on test use, tertiary education programs that prepare persons for test development and use, and a large workforce that depends on test data.

A country also must display positive attitudes toward test use. Its citizens must view the use of tests and other assessment methods to be reliable, valid, efficient, humane, and address important social issues. Test development and use are based on science, and test use is a form of technology. Thus, tests are more common in countries that value science and technology.

Tests first were developed more than 3,000 years ago in China (Wang, 1993). In the 1880s tests were developed in the West to assist research efforts that examined individual differences (i.e., whether traits that distinguish people can be validly assessed). Test use assumes the presence of individual differences and the importance of identifying them. Countries differ in the value they place on individuals versus groups (e.g., families, work, and social groups) and differ in their emphasis on the importance of individuals versus groups.

Countries that place a greater emphasis on individualism (e.g., the United States and most Western European countries) expect their citizens to look after one's self and immediate family. Persons in these countries tend to be more competitive and believe their interests are more important than most others. In contrast, countries that place greater emphasis on collectivism (e.g., People's Republic of China, most Latin American countries) expect their citizens to form strong cohesive groups that protect them in exchange for their service and loyalty. Persons in these countries are more inclined to put aside their individual pursuits in favor of those important to the groups in which they are members (Hofstede, 1994).

A country's emphasis on this individualism-collectivism dimension impacts test development and use. Test development tends to occur more frequently in countries that emphasize individualism and favor meritocracy (i.e., the belief that persons should be rewarded based on their accomplishments) than collectivism and egalitarianism (i.e., the belief that all people are equal and should have equal access to resources and opportunities). Some psychologists believe this focus on individual differences may be the discipline's most enduring and unique contribution to the behavioral sciences (Benjamin, 2007).

Nationally developed tests are most common in Australia, Canada, Western Europe, and the United States. These countries generally display a stronger commitment to individualist and merit-based beliefs than to collectivist- and egalitarian-based beliefs. Collectively, they constitute about 10% of the world's population. In contrast, test development and use are lower in countries that have or had strong ties to communism or socialism or strongly value a collectivism (e.g., People's Republic of China, countries that formed the Soviet Union, as well as Mexico and those in Central and South America). Test development and use are lowest among the 55 African countries and 22 Arab countries (Hu & Oakland, 1991; Oakland & Hu, 1991, 1992, 1993).

The following 12 countries have the largest populations (e.g., in rank order: China, India, United States, Indonesia, Brazil, Pakistan, Bangladesh, Russia, Japan, Mexico, Philippines, Vietnam). Their combined population is approximately 4.2 billion. Among them, only the United States has a large number of locally developed standardized tests. An estimated 80% or more of the world's population reside in countries in which locally developed standardized tests either are somewhat uncommon or rare.

INTERNATIONAL SURVEYS ON TEST USE

Test Use With Children

Forty-four respondents with specialized knowledge of psychological and/or educational testing in their countries and who were members of the International School Psychology Association, International Council of Psychologists, or the International Test Commission completed a questionnaire through which detailed information about the status of test development and use with children and youth in

their respective countries was obtained. Respondents identified 455 tests used frequently, especially measures of intelligence, personality, and achievement (Hu & Oakland, 1991; Oakland & Hu, 1991, 1992, 1993).

Among commonly used tests, 46% were developed within other countries and imported for use. Tests imported for use came mainly from one of five countries: United States (22%), United Kingdom (7%), Germany (7%), France (5%), and Sweden (5%). The 99 tests that were published in the United States were cited 301 times. Among those tests used outside their country of origin, 97% were developed in highly industrialized nations.

Foreign-developed tests were used more frequently than locally developed tests in 68% of the countries surveyed. Locally developed tests were used more frequently than foreign-developed tests in only 27% of the countries. Seven countries report no locally developed tests. Many smaller and developing countries that were not surveyed also are likely to lack locally developed tests.

Types of Tests Used

Measures of intelligence (39%), personality (24%), and achievement (10%) were cited most commonly. Measures of perceptual-motor abilities, vocational interests and aptitudes, school readiness, and social development were less common (i.e., 3% to 6% in each of the categories).

Psychometric Standards

Standardized tests can be expected to display three basic qualities: adequate, representative, and recent norms; sufficient reliability to ensure the data are stable; and sufficient validity based on both theory and empirical evidence that pertain to the ways in which tests are used. Many tests used in the more developed countries display these qualities. However, many—perhaps most—tests used in developing countries lack one or more of these basic qualities. The psychometric qualities of tests tend to be most deficient when tests are obtained from a host country (i.e., the test's country of origin) and translated for use in the target country (the location in which the translated test is used). These tests typically lack target country norms and their reliabilities and validities are unknown. As noted later in this chapter, the use of the International Test

Commission's guidelines for test adaptations helps overcome many of the problems seen in translated tests. The psychometric qualities of tests identified in Oakland and Hu's survey are summarized below.

Availability of Norms

Sound testing practices typically rely on locally developed norms. National norms were available on 80% of the achievement tests, 65% of intelligence tests, and 58% of personality tests. Thus, many tests lack local norms.

Reliability Studies

Tests must provide consistent data for them to be useful. Studies estimating internal consistency or test-retest reliability were conducted on approximately 50% to 60% of achievement, intelligence, and personality tests. Thus, the reliability of many tests is unknown.

Validity Studies

Validity (e.g., the degree to which a test accurately measures what it was designed to measure) generally constitutes a test's most valued quality. Validity studies were most common on achievement tests. Among achievement tests, concurrent validity studies were available on 71%, construct validity studies on 48%, and predictive validity studies on 43%. Among measures of intelligence, concurrent validity studies were available on 63%, predictive validity studies on 56%, and construct validity studies on 54%. Thus, the validity of many tests is unknown.

Needs for Tests

Two-thirds of the countries reported critical needs for both group and individual tests of achievement, intelligence, vocational interests and aptitudes, social development and personality as well as more moderate needs for measures of perception, motor development as well as those used for entrance into primary, secondary, and tertiary schools.

Eighty-five percent of the responding countries reported the need for tests that assess qualities important for those who are mentally retarded, blind, deaf, learning disabled, slower learners, emotionally and socially disturbed, physically impaired, and the gifted. The need for tests for those with learning disabilities was most critical.

Professionals Who Administer Tests

Sound testing practices require suitably educated professionals to correctly select, administer, score, and to wisely interpret tests and other measures. At least 16 professional groups commonly administer tests (Oakland & Hu, 1991). In many countries, school or educational psychologists assumed leadership for these activities. Other frequently cited specialists included regular or special education teachers, clinical psychologists, counselors, and professionals engaged in the health service professions.

The amount of postsecondary education found among the 16 groups ranges from a mean of 2.5 years for nurses to 6.5 years for physicians. The correlation between years of postsecondary education and the perceived adequacy of test users is significant ($r = .50, p < .001$). Thus, the adequacy of test users is associated with their level of education. Professionals who use both group and individually administered tests typically have more education than those who use only group tests.

Test Use With Adults

Test use with adults also is common and somewhat universal. Tests are used by tertiary institutions for entrance, retention, and graduation; by state and national boards to certify and license vocations and professions; by professionals to evaluate medical, social, and psychological problems; and in the business community to assist in selecting, training, retaining, and promoting employees as well as certifying attainment of critical abilities and skills of persons at entry and mid-management levels and above (e.g., DiMilia, Smith, & Brown, 1994; Gowing & Slivinski, 1994; Schuler, Frier, & Kauffmann, 1993; Shackleton & Newell, 1994). For example, persons seeking positions that require word and numerical processing skills increasingly are required to demonstrate competence in the use of software programs by passing tests developed for this purpose. One software company, Microsoft, administered credentialing examinations to more than 1 million persons in 15 languages and in more than 30 countries in 1997 (Fitzgerald & Ward, 1998).

The Myers-Briggs Type Indicator, commonly used in personnel selection, training, and team building, may be the most widely used measure in the world, with 2 million administrations reportedly occurring each year (Myers, McCauley, Quenk, & Hammer, 1998). It has been translated into at least 17 languages or dialects for which there is

commercial distribution and another 13 languages for which there is so-called underground (e.g., noncommercial) distribution. The Minnesota Multiphasic Personality Inventory also is widely used. Cross-national research supporting the viability of the big five personality model cross-nationally has contributed to the popularity of the Revised NEO Personality Inventory and NEO Five-Factor Inventory (Costa & McCrae, 1992). Cheung's review (2004) of the uses of Western and indigenously (i.e., locally) developed personality tests in Asia underscores the growing interest in personality assessment for research and clinical use within this region.

An international survey of test-related issues in 29 countries (19 in Europe and others from Central and South America, Asia-Pacific, the Middle East, and South Africa), largely in reference to adults, also reports considerable diversity between countries in their approaches to testing, their uses of tests, and attitudes toward test user qualifications (Bartram & Coyne, 1998; Muniz, Prieto, Almeida, & Bartram, 1999). Assessment methods are more consistent within specialty areas of applied practice (e.g., within clinical, occupational, educational, or forensic practices) across countries than between specialty areas within countries.

STANDARDS AND GUIDELINES THAT MAY IMPACT TEST DEVELOPMENT AND USE

Test development and use generally are strengthened by employing standards or guidelines, including those that address legal uses of tests, standards and guidelines for test development and use, ethical issues associated with test development and use, and professional qualities needed by those who use tests. Legal issues typically differ considerably between countries. Scholarship that examines legal issues impacting test development and use internationally could not be located. Thus, the following comments address other nonlegal issues.

Technical Standards and Guidelines for Test Development and Use

Professional associations can be expected to establish and promulgate standards or guidelines that impact their professional practices. Standards define obligatory practices whereas guidelines suggest advisory practices. Standards are employed when a professional association, typically

at the national level, has sufficient authority and leverage to oversee the display of practices associated with standards by its members and, when needed, can take steps to either curtail or work to improve unsuitable practices. International professional associations rarely have this authority and leverage and thus propose guidelines, not standards, for practice.

Examples of National Standards

Standards for Educational and Psychological Testing (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education, 1999) and its previous editions have served as the authoritative source for test development and use in the United States since 1954. The current edition discusses issues pertaining to test construction, evaluation, and documentation (e.g., validity, reliability, scales, norms, test administration and scoring), fairness in testing (e.g., rights and responsibilities of test-takers, testing persons from diverse linguistic backgrounds and those with disabilities), and testing applications (e.g., test use in psychology, education, employment, credentialing, and program evaluation). These standards and those from the Canadian Psychological Association (1987) and other sources (e.g., Joint Committee on Testing Practices, 1993; Kendall, Jenkinson, De Lemos, & Clancy, 1997; Koene, 1997; Lindsay, 1996) often find acceptance from psychologists in other countries in which national standards on test development and use have not been established.

International Guidelines

The International Test Commission (ITC, www.intestcom.org) plays a central role in addressing cross-national issues that impact test development and use. It sponsors biannual conventions, a journal (*International Journal of Testing*), a newsletter (*Testing International*), and has assumed leadership with respect to developing international guidelines for test use, adapting tests, and testing and the Internet. These guidelines are summarized below.

Guidelines for Test Use

The International Test Commission's commitment to promoting practices that can have a beneficial impact on test use is seen in its original charge. Early records reveal an uneasiness as to the presence

of unqualified persons using tests, their making important decisions despite their limited preparation and experience, and their use of tests that lack suitable norms and sufficient validity (Oakland, Poortinga, Schlegel, & Hambleton, 2001). The International Guidelines for Test Use, developed under the leadership of David Bartram (past president of the International Test Commission), discuss the fair and ethical use of tests with the intent to provide an internationally agreed framework from which standards for training and test user competence and qualifications could be derived (Bartram, 1998). These guidelines were approved by the International Test Commission in 1999 and have been endorsed by the European Federation of Professional Psychologists Associations Standing Committee on Tests and Testing. A number of countries have translated and adopted these guidelines.

Guidelines on Test Adaptations

As noted above, test use in most countries is characterized by importing tests, typically from five Western countries. These tests typically are translated from the original source language to the local language. Efforts to norm these translated tests, establish their psychometric properties (e.g., reliability and validity), and determine the relevance of the test's content to the local culture are sporadic. These practices clearly are below professional standards.

The International Test Commission, under the leadership of Ronald Hambleton (past president of the International Test Commission), together with support and participation from various organizations (i.e., American Psychological Association, Canadian Nursing Association, Collegio de Psicologos, European Association for Psychological Assessment, European Test Publishers Group, International Association for the Evaluation of Educational Achievement, International Union of Psychological Sciences, National Institute for Education Measurement in the Netherlands, and the United States Department of Education) developed guidelines for adapting educational and psychological tests in an effort to overcome these deficiencies (Hambleton, 1994; Hambleton, Merenda, & Spielberger, 2005; Muniz & Hambleton, 1997; van de Vijver & Hambleton, 1996). These guidelines are discussed in some detail below, given their relevance for avoiding common problems associated with merely translating tests.

The test adaptation guidelines are intended to provide assistance to persons attempting to transform a test from one originally intended to

be used with one population (the source) to one suitable for use with a different population (the target). Two examples include the transformation of a test originally developed in the United States to one revised for use in Hungary or transforming a test from one developed in Hungary designed to be used with native-born Hungarians to one revised for use in Hungary with non-native-born Hungarians from Romania. These guidelines discuss cultural and language differences, identify five technical issues and methods, and describe three conditions that possibly impact test interpretations. Each of these is reviewed below.

Measurement Error and Their Sources. Test validity constitutes a test's most important quality (AERA et al., 1999). Validity refers to the accuracy with which a test measures a construct and how the results may be used appropriately. Validity is judged in light of theory and empirical evidence that support the manner in which test data are interpreted and used. Strictly speaking, a test does not have validity.

Validity may be attenuated by various conditions. Two that are most prominent include construct underrepresentation (i.e., when a test fails to measure important aspects of the construct) and construct irrelevance (i.e., when qualities extraneous to the construct attenuate its measurement). Thus, those engaged in test development and use as well as test adaptations strive to reduce measurement error by ensuring a test measures a trait or construct adequately and consistently with the source test and that the target test does not measure extraneous qualities (e.g., that a measure of intelligence also does not measure the ability to read or to use a particular language or dialect) in order to help ensure the scores are accurate. Three broad conditions that may contribute to measurement error when adapting tests or using them are discussed below: cultural and language differences, technical issues and methods, and conditions that impact test performance.

Cultural and Language Differences. The potential for error increases as differences increase between persons who comprise the source (i.e., first) and target (i.e., the one used in the adapted test) tests in reference to languages, culture (e.g., values, dress, food, money, forms of measurement), social class, urban-rural residence, gender, and age. Four methods to address these potential problems include establishing the equivalence of the constructs for the various groups, promoting proper test administrative practices, utilizing suitable norms, and minimizing tests that rely heavily on speed.

First, establish a test's construct equivalence. Knowledge of a test's construct provides the single most important evidence of a test's

validity. Construct validity is likely to be promoted when a test displays other suitable psychometric qualities (e.g., reliability, adequacy of norms) and to not be adversely impacted by unsuitable psychometric qualities. Investigations of construct equivalence consider whether the target culture has and adheres to this construct, whether the construct has a similar meaning within the two cultures, and whether the construct is displayed in a similar fashion.

Methods to investigate construct equivalence initially involve and often rely heavily on judgmental strategies (e.g., interviewing and observing persons, literature reviews, as well as consulting with cultural anthropologists and others who specialize in a region and know the culture well). These judgmental strategies are subjective and thus will benefit from input from multiple sources.

Second, promote proper test administration. For example, test directions should be understood clearly, verbal communication should be minimal, and information should be communicated consistently for source and target groups. Test administrators should be properly selected. They should be drawn from the target community(ies), familiar with the culture and with the languages (dialects), have experience administering tests, and recognize the importance of maintaining standardized methods. Their training and preparation may be needed to obtain these qualities together with ongoing supervision to ensure their presence.

Third, utilize suitable test formats. Cultures differ in their use of different test formats (e.g., multiple choice, short answer, essay). While one typically should emphasize the use of those formats with which people are familiar, the limited use of various formats may be warranted. When using tests to make cross-cultural comparisons, consider using a multiple-choice format as it can be scored more objectively. In contrast, scoring rubrics for essay exams may introduce considerable error.

Fourth, emphasize power and reduce reliance on speed. Cultures differ in reference to the importance placed on completing work, including tests, well or quickly. *Power tests* are intended to assess a person's level of mastery of a topic or trait when tested without time constraints. In contrast, *speed tests* are intended to assess the number of problems a person can complete during a predetermined and often brief time period. Power tests rather than speed tests may provide a more accurate assessment of many qualities.

Five Technical Issues and Methods. Various technical issues and methods used when adapting a test are summarized below.

Focus on the test revision process. When developing a test that is likely to be used in two or more cultures, outline a test development strategy that reflects this goal. Address issues that pertain to the choice of item formats, stimulus materials, vocabulary (e.g., unless the test assesses vocabulary, keep vocabulary simple), sentence structure (again, keep it simple), and cultural differences (e.g., seasons, time, money, weights, foods, dress, gender roles, knowledge content, writing and reading from either the right or left, as well as temperament and personality differences).

Select and prepare translators. A target test is only as good as its linguistic equivalent of the source test. Translators should include two or more persons familiar with both cultures and languages (i.e., more than a literal translation is needed), the subject matter, and with skills in test construction.

Use data analysis methods to establish equivalence and to detect bias. Ultimately data need to be collected and analyzed. The data should examine qualities at the item level (e.g., difficulty, distractibility, and discrimination), construct level (e.g., confirmatory factor analysis), together with means and standard deviations. Note that groups may differ legitimately in their means and standard deviations. Moreover, such differences may be less important when cross-cultural and cross-national comparisons are not being made.

Decenter the test. The term *decentering* refers to a process of revising a test's source language so that equivalent materials can be used in both the source and target language versions. Two typical decentering methods are used when translating tests: a forward translation or a backward translation process. A third translation process, consensus translation, is discussed later in this chapter.

Using a forward translation, a single translator or group of translators first adapts the test from the source language to the target language. Then other translators compare the equivalence of the two versions. This comparison may lead to changes in the target language version. The advantages of the forward translation are that judgments are made directly between the two versions and the process is less costly and quicker. The disadvantages of this process are that considerable inferences are required by the translators, they may be more proficient in one of the two languages, and ratings by those who are bilingual may not reflect the language abilities of those who are monolingual.

Using a backward translation, a single translator or group of translators first adapts the test from the source language to the target language.

Then another group of translators adapts the test from the target language back to the source language. Then other translators compare the equivalence of the two versions. The principal advantage of this process is that it provides a more thorough review of possible language problems. Its disadvantages are that comparisons are specific to only the source language test, the target test will contain shortcomings of the original test (e.g., problems with grammar, content), and this process is more costly in time and personnel.

Both decentering methods are far from perfect. Both methods fail to provide empirical data on the performance of actual people for whom the test is designed as well as the test's psychometric characteristics. Furthermore, the tests were not administered under testlike conditions, and limitations inherent in the scope and concept of the source test are retained by the second target test.

These guidelines propose three empirical designs to help overcome some of these limitations. Bilingual examiners can take the source and target versions. This method helps control for ability levels, is low in costs, and can be completed quickly. However, one cannot assume those who are bilingual are similar to the target group in reference to important qualities (e.g., language).

A second method employs source language monolinguals who take the original and back-translated versions. This method allows for a comparison of item characteristics. However, source language monolinguals are likely to differ from non-source language monolinguals; furthermore, taking the first test may influence performance on the second test.

A third method employs source language monolinguals who take the source language test as well as target language monolinguals who take the target language test. This provides data on both tests. However, the two groups may differ by ability.

Conditions That Possibly Impact Test Performance. Three conditions may impact test performance and thus test interpretations: assumed similarity of cultural experiences, suitable levels of motivation, and sociopolitical qualities.

One may incorrectly assume cultural experiences between the source and target groups are similar. Differences may exist in access to information, school curricula, values, attitudes, and if one reads a book from the front to the back or from the back to the front.

Test results are assumed to be valid if those who took a test displayed suitable levels of motivation. For example, cognitive measures assume

those being tested are highly motivated and strive to do as well as they possibly can. However, some examinees may not strive to achieve high scores. Three test-taking qualities can adversely influence test performance: avoidance, uncooperative mood, and inattentiveness (Oakland & Glutting, 1998). The display of these qualities generally contributes to error that result in scores below one's potential. One should be alert to the presence of these and other qualities that attenuate performance. Determining whether test-takers display avoidance, an uncooperative mood, or inattentiveness is more difficult on group than individually administered tests.

Sociopolitical qualities also may influence test performance. Tests typically require an individual to perform some activity. Some persons function well when asked to perform individually. However, others seemingly do better when involved in a group. For example, those who are extroverted generally prefer group activities that involve talking while those who are introverted generally prefer individual activities that rely more heavily on writing. Orientations to life also may be influenced by many qualities, including one's religion, tribal expectations, regional differences, gender, and individual-collective values.

Ethical Issues Associated With Test Adaptation. Thirty ethical standards from the APA's Ethical Principles of Psychologists and Code of Conduct (APA, 1992) apply to ethics associated with test adaptations and use (Oakland, 2005). Among them, test plagiarism (i.e., the taking of someone's work product for personal benefit without compensating the author) as well as nonadherence to copyright provisions (i.e., when tests either are photocopied or are adapted without the consent of the test's author and publisher) are somewhat common and serious. Possible implications of six ethical principles and 30 standards from the APA Code in reference to 11 stakeholders who use adapted tests are discussed elsewhere (Oakland, 2005). Persons involved in test adaptations should be aware of the potential violation of laws and ethical principles through their work and strive to establish and maintain high standards for themselves and the profession.

International Guidelines on Computer-Based and Internet-Delivered Testing

Technology that has an international reach often requires the involvement of organizations that transcend one nation. Efforts by multinational governmental agencies, multinational companies, as well as regional and

international associations often are needed to envision, revise, create, promote, regulate, and in other ways assist in forming and reforming services in light of changes due to technology. Computer use surely is a pervasive technology that has changed the ways in which we work, shop, communicate, and play.

Through computers, testing technology has acquired an international reach and is increasing. Many of the world's largest testing companies are making their tests available through the Internet. This availability invites various problems. For example, although more than 150 countries are signatures to the 1996 World Intellectual Property Organization Copyright and Performance and Phonograms Treaties, they differ considerably in their adoption and enforcement of the treaty's provisions. This is seen clearly in nonadherence to copyright protection. Encryption and enveloping technologies together with copyright management information on test products are being used to help address this growing problem (Thiemann, 1998).

The legitimate use and potential abuse of computers generally is well known. The potential for abuse warranted standards or guidelines for test administration, security of tests and test results, and control of the testing process. Therefore, the International Test Commission established international guidelines on computer-based and Internet-delivered testing in its *Computer-Based Testing and the Internet: Issues and Advances* (Bartram & Hambleton, 2005; Coyne & Bartram, 2004; <http://www.intestcom.org>).

The goal of these guidelines is to raise awareness among all stakeholders in the testing process of internationally recognized guidelines that highlight good practice issues in computer-based testing (CBT) and testing delivered over the Internet. The development of these guidelines drew on common themes that run through other existing guidelines, codes of practice, standards, scholarship, and other sources to create a coherent structure within which they can be used and understood. In addition, these guidelines are specific to CBT/Internet-based testing. Test developers, test publishers, and test users share responsibilities for ensuring the following four guidelines are enforced.

Attend to technological issues in CBT/Internet testing (i.e., give consideration to hardware and software requirements, ensure the robustness of the CBT/Internet test, recognize human factors issues in the presentation of material via computer or the Internet, consider reasonable adjustments to the technical features of the test for candidates with disabilities, and provide help and information, both off- and on-screen, as well as practice items within CBT/Internet tests).

Attend to quality issues in CBT/Internet testing (i.e., ensure knowledge and competence of CBT/Internet testing, consider the psychometric qualities of the CBT/Internet test, ensure that there is evidence of equivalence when the CBT/Internet test has been developed from a paper-and-pencil version, score and analyze CBT/Internet testing results accurately, interpret results appropriately, give appropriate feedback, and work to ensure equality of access for all groups).

Provide appropriate levels of control over CBT/Internet testing (i.e., ensure control over the test conditions and their supervision, consider controlling prior practice and item exposure, and ensure control over test-taker's authenticity and cheating).

Make appropriate provision for ensuring security and safeguarding privacy in CBT/Internet testing (i.e., take account of the security of test materials, ensure security of test-taker's data transferred over the Internet, and maintain the confidentiality of test-taker results).

ETHICAL ISSUES ASSOCIATED WITH TEST DEVELOPMENT AND USE

An unwritten yet generally recognized social contract binds professions and the societies in which they practice. Within this social contract, a society agrees to provide funds to help prepare professionals and to conduct research as well as allows the profession to select, prepare, and credential neophytes and to establish standards for services. This social contract also requires a profession to serve all of society well and, minimally, to do no harm.

A profession's ethics' code informs society of its commitment to serve society. Such codes are designed to protect the public by prescribing and proscribing behaviors professionals are expected to exhibit. Ethics' codes typically contain principles and standards that reflect both general virtues (e.g., beneficence, fidelity) and specific behaviors (e.g., informed consent, confidentiality).

The first known ethics' code, the Code of Hammurabi (circa 1795–1750 B.C.), reflected a desire to impose rules governing personal and vocational behaviors (Sinclair, 2005). The Hippocratic Oath (circa 400 B.C.) is the first known example of a professionally generated ethics' code. The 1953 American Psychological Association's ethics' code is the first example of one for psychologists.

Psychology is represented internationally by various associations, two of which are most prominent: the International Union of Psychological

Science (IUPsyS) and the International Association of Applied Psychologists (IAAP)—the oldest international association of psychologists. The mission of IUPsyS is to build global interaction among research communities and promote advances in psychological science and technology at the international level. Its members are national associations. The mission of IAAP is to promote the science and practice of applied psychology and to facilitate interaction and communication about applied psychology around the world. Its members are individual psychologists.

The importance of these and other international associations has increased during the last two decades, given the growing science and practices of psychology that transcend country and cultural boundaries. Thus, one may expect the emergence of international and regional codes of ethics that address cross-national practice issues, including test development and use. Such codes would promote and highlight common professional practice, thus conveying psychology's universal commitment and impact.

International Ethics Codes

Neither the IUPsyS nor the IAAP has developed ethics codes. However, they approved the Declaration of Universal Ethical Principles for Psychologists (Gauthier, 2008). Its purpose is to promote unity within the profession of psychology internationally by including ethical practices regardless of country or culture. Although this declaration is not intended to replace existing codes, it is expected to impact their revisions.

The International School Psychology Association has adopted an ethics code, applicable only to its specialty, that addresses issues associated with professional responsibilities, confidentiality, professional growth and limitations, relationships, assessment, and research (Oakland, Goldman, & Bischoff, 1997).

Regional Ethics Codes

The European Federation of Professional Psychologists Associations' (EFPA) Meta-Code of Ethics (www.efpa.be/ethics.php), adopted in 1995, is applicable to practices of members of 31 national psychology organizations. This code is one of the first regional codes to underscore the importance of principles for psychological organizations and psychologists. This brief document addresses four main principles that are

applicable to practice: respect for a person's rights and dignity, competence, responsibility, and integrity. The five Nordic countries (Denmark, Finland, Iceland, Norway, and Sweden) also adopted a unified code of ethics in 1988 that was revised in 1998 to be more consistent with EFPA's meta-code (Pettifor, 2007).

International Studies of Ethics Codes That Address Test-Related Issues

Contemporary forms of tests initially were used in research, not in applied practice, and thus had limited social visibility and impact. However, both the globalization of psychology as well as an increase in test use to address important social, psychological, educational, and vocational needs have deepened their impact and broadened their visibility, resulting in questions as to whether they serve the public good. Thus, ethical and other professional standards that inform the public regarding expectations from those who develop and use tests are needed.

Separate ethical principles and standards that govern only test development and use could not be located. Ethical principles and standards governing test development and use typically are embedded within a country's broader ethics code (APA, 1992; Lindsay, 1996). Moreover, psychological tests are used broadly in some countries and narrowly in others (e.g., mainly to assess mental retardation). Scholarship that discusses ethical issues specific to tests is somewhat meager (Leach, Glossoff, & Overmier, 2001; Leach & Harbin, 1997; Leach & Oakland, 2007).

Testing standards in 31 ethics codes representing 35 countries were compared with those in the APA's 2002 Ethical Principles of Psychologists and Code of Conduct (Leach & Oakland, 2007). Codes from approximately one-third of the countries surveyed do not address test use. This finding should not be surprising in that test development and use are not prevalent in many countries. Thus, their ethics codes do not need to address these issues.

However, ethics codes in some countries with advanced test development and use do not address test issues (e.g., Canada, Switzerland). Codes from some countries emphasize broad and virtuous qualities (e.g., respect, responsibility) intended to have a pervasive impact on psychological practice and do not focus on more specific interests (e.g., clinical practice, testing, advertising). Although these codes do not address test issues directly, their virtuous-centered principles are intended to impact them.

The 20 (65%) codes that address test use included one or more specific standards that also are consistent with the 2002 APA Code of Conduct. The frequency of ethics codes, noted in the parentheses, that address these standards follows: explaining test results (15), using assessment (11), assessment by unqualified persons (11), interpreting assessment results (10), maintaining test security (10), informed consent (8), test scoring and interpretation services (7), the basis of assessment (5), test construction (5), and using obsolete tests and outdated test results (4).

Given the widespread use of imported tests (Oakland & Hambleton, 1995), one may expect standards that directly address test use cross-nationally. However, only Latvia's code explicitly addresses this issue, stating that Latvian psychologists should conduct studies ensuring suitable psychometric properties when tests have been developed elsewhere (Leach & Oakland, 2007). Other commonly endorsed standards require persons who use tests to have suitable training and experience and to ensure the test's security (Leach & Harbin, 1997; Leach et al., 2001).

Few codes address issues associated with test construction or the use of obsolete tests; two issues that often are linked. For example, when psychological tests are more likely to be imported than developed locally, the tests tend to be used for many years despite the availability of newer editions. In addition, standards that address test construction generally are not needed in countries that do not develop their own tests.

Psychologists in many countries do not honor copyright protection of test protocols, manuals, and test kits. Two examples follow. Ten or more pirated copies of the Wechsler Intelligence Scale for Children are in use globally. The Russian government has been using the 16 Personality Factor Test for decades without compensating its author or publisher.

These practices violate international copyright laws and are unethical in those countries with ethics codes that proscribe these practices. International legal agreements prevent the photocopying and reproduction of intellectual and other property, including tests. Issues concerning the protection of copyright become ethical only if they are included in a country's ethics code. Few countries address this issue in their codes.

Psychologists play a leading role in developing and using psychological tests that serve the public and professions and thus are committed to maintaining the integrity and security of test materials and other assessment methods, knowing that their unauthorized release to the public jeopardizes test integrity, results in test use by unqualified persons, and thus harms the public. Some persons are selling tests through

unauthorized sources to the general public. These sales jeopardize test integrity, harm the public, and violate accepted practice as well.

The International Test Commission has urged professionals to become aware of this possible practice in their countries and to take steps to stop such unauthorized sales. Unfortunately, ethics codes typically do not address issues associated with the secondary sale or disposal of tests one no longer uses. The Commission has encouraged national psychological associations to inform their members of this problem and to take preventative measures, including the revision of their ethics codes to help prevent this and similar unauthorized releases of tests to others. National psychological associations also are encouraged to develop standards that promote the safe disposal of outdated tests.

INTERNATIONAL SOURCES USED TO DEFINE AND DESCRIBE DISORDERS

Three Commonly Used Sources to Diagnose Mental Disorders

Test data commonly are used to describe behaviors and, when working in clinical settings, to diagnose disorders. Clinicians typically use one of the following three authoritative international sources to classify mental disorders: the *Diagnostic and Statistical Manual of Mental Disorders, 4th edition, text revision (DSM-IV-TR)*; American Psychiatric Association, 2000); its *International Version* (American Psychiatric Association, 1995); and the *International Classification of Diseases and Related Health Problems, 10th edition (ICD-10)*; World Health Organization [WHO], 1992). The disorders identified by the *ICD-10* generally are consistent with those cited in and are cross-referenced to the *DSM's International Version*. These three diagnostic systems have been developed by the medical community and view behaviors from a medical model.

A New Model for Conceptualizing Health, Wellness, and Disability

The World Health Organization's *International Classification of Functioning, Disability and Health (ICF)*; WHO, 2001) provides a different model for viewing disorders. Its bio-psycho-social framework views behaviors from three broad perspectives (see Figure 1.1): (1) body functions

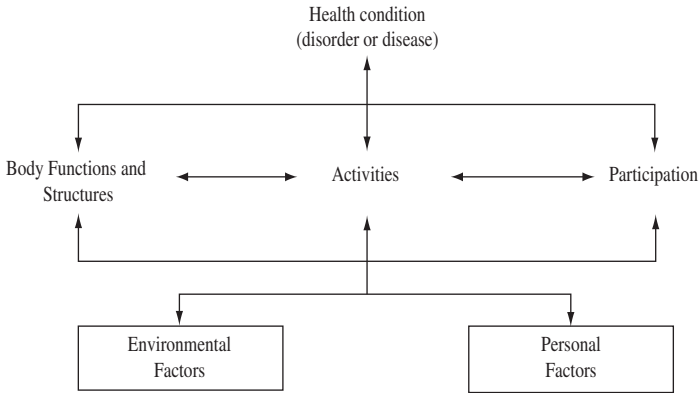


Figure 1.1 Interactions between the components of *ICF*.

From *International Classification of Functioning, Disability and Health: Children and Youth Version* (p. 17), by World Health Organization, 2007, Geneva, Switzerland: WHO. Copyright © 2007 by World Health Organization. Reprinted with permission.

and structures (e.g., physiologic, physical, and psychological functions), (2) activities (e.g., the extent to which persons engage in functional life activities), and (3) participation (e.g., their participation in social settings). A patient's health is understood from knowledge of the interaction between three broad components. Working in this model, professionals integrate medical and psychological information with knowledge of a person's social and adaptive skills (e.g., personality traits, coping abilities, stress, and social support). Professionals may use WHO's *International Classification of Diseases and Related Health Problems, 10th edition*, if a diagnosis is needed.

Thus, an understanding of a person's health requires knowledge of the dynamic nature between body functions, body structures, and activities as well as participation and environmental factors. Each can influence the others. The *ICF* emphasizes the importance of identifying possible conditions that impact activities and performance deficits. An understanding of a person's activities and performance requires knowledge of personal, social, and environmental conditions that may be impacting the person.

For example, a person's adaptive skills may be influenced adversely by his or her body functions (e.g., mental, sensory, and neuromusculoskeletal functions) and structures (e.g., nervous, cardiovascular, and metabolic systems). Additionally, his or her environment may not provide needed opportunities to acquire adaptive skills as well as support and reward their use. Thus, knowledge of a person's adaptive skills in

conjunction with body functions, structures, and environment is important to diagnosis, if one is needed, and is essential to the design, delivery, and monitoring of services intended to have an instrumental and functional impact on a person's life.

The *ICF* places considerable emphasis on identifying functional impairments and thus emphasizes the importance of adaptive behaviors and life skills. Specifically, its activities and participation components address the execution of a task or action by an individual and his or her involvement in life situations. The term *activities* refer to tasks or actions a person is able to perform. Examples for older children and adults include writing, talking, and calculating. The term *participation* refers to activities that become integrated into one's life. Examples include regularly taking others to nearby places, talking by telephone with family and friends, and refraining from embarrassing others.

Activities and participation include the following nine domains: learning and applying knowledge (e.g., functional academics), general tasks and demands (e.g., work), communication (e.g., communication), mobility (e.g., fine and gross motor skills), self-care (e.g., self-care), domestic life (e.g., school and home living), interpersonal interactions and relationships (e.g., social skills), major life areas (e.g., health and safety, leisure skills), and community, social, and civic life (e.g., community use).

A *skill deficit* occurs when a person does not display a needed behavior. A *performance deficit* occurs when a person has displayed a needed skill yet does not use it when needed. For example, a child who does not have the ability to dress oneself displays a skill deficit. In contrast, a child who has displayed the ability to dress oneself and does not do so regularly when needed is described as having a performance deficit. If deficits in behaviors and skills have been identified and an individual is in need of services, then the *ICF* aids in describing the disability in terms of an interaction between the impairment, functioning, and the environment. Strengths or weaknesses may be identified, including the adequacy of one's adaptive skills, in light of environmental needs.

The *ICF* currently cites 10 mental and behavioral disorders: organic, including symptomatic, mental disorders; mental and behavioral disorders due to psychoactive substance use; schizophrenia, schizotypal, and delusional disorders; mood (affective) disorders; neurotic, stress-related, and somatoform disorders; behavioral syndromes associated with physiological disturbances and physical factors; disorders of adult personality and behavior; mental retardation; disorders of psychological development; and behavioral and emotional disorders with onset

usually occurring in childhood and adolescence. A category for unspecified mental disorder also is provided. Other disorders have been added for children and youth in the WHO's (World Health Organization, 2007) version of the *ICF* for children and youth.

The adoption of the *ICF* will lead to various changes in health and health-related professions and thus to the tests needed to assist their work. The traditional medical model, with its emphasis on pathology, will be deemphasized in favor of a bio-psycho-social model that incorporates all components of health described at body, individual, and societal levels. Assessments will focus more on activities (i.e., tasks or actions a person is able to perform), participation (i.e., activities that become integrated into one's life), and contextual factors (i.e., the nature of one's environment together with its supports and impediments). Additionally, assessment will require the use of multidisciplinary methods.

Thus, the *ICF* encourages us to view health and disability differently. Every person, not just a few, can be expected to experience a decrement in health and thereby experience some degree of disability. Both health and disability are universal experiences. The *ICF* shifts our attention from the causes of a disorder to the disorder's functional impact on behaviors. Furthermore, the *ICF* takes into account the social aspects of disability and does not see disability only as dysfunction due to medical or biological qualities.

SOME FUTURES OF TEST DEVELOPMENT AND USE INTERNATIONALLY

The development and use of tests in the behavioral sciences may constitute psychology's most important technical contribution and is consistent with its overriding quest to understand individual growth and development. The futures of test development and use will be impacted by both external conditions—those over which psychology has little control, as well as internal conditions—those over which psychology has more control.

Conditions External to Test Development and Use

External conditions include the stability of a country's educational and political systems; a well-developed economy with financial resources that support test use and development; a testing industry that develops and markets tests for local use; public recognition that prevailing personal

and social problems may be addressed through test use; and the display of positive attitudes and values toward science, technology, and individual differences. Test development and use are stronger when a country has a well-established and well-funded educational system, uses its financial capital to support the development of a testing industry, uses tests to address important personal and social problems, and has a public that holds positive views toward test use and individual differences.

Public attitudes impact test use. Some people see them as a great resource to acquire valid and unbiased information efficiently. Others are skeptical about their use. For example, people in many countries are or have been living under fear that governments could obtain various forms of test-related information to be used against them. They may oppose the use of tests, given their belief that tests are used to serve an oppressive state, not them.

The presence of these qualities will strongly impact the status of test development and use, particularly in those countries that have few test resources. Many countries lacking these desired qualities can be expected to remain underdeveloped with respect to test development and use through most of the 21st century. Thus, although qualities internal to psychology may initially promote test development and use, their continued development and use require the need for tests, a testing infrastructure that responds to this need, and the belief that test use enhances the attainment of personal and national values.

Conditions Internal to Test Development and Use

Psychology and other social sciences have control of other conditions that affect test development and use. These include preparing sufficient numbers of professionals able to develop and use tests properly, developing suitable measures that assess a broad range of important qualities, the presence of strong professional associations that support the value of assessment, standards that address issues important to test development and use, and positive attitudes toward test use held by the profession.

Competencies to develop tests require considerable graduate-level work in psychometrics as well as research design and statistics. Within the United States, the number of graduate programs that prepare persons for this work and the number of graduate students able later to replace forthcoming retirements are meager (Aiken, West, & Millsap, 2008; Rossen & Oakland, 2008). Conditions in developing countries are more dismal. For example, within the Americas, only three countries

offer this graduate specialization: Canada, Brazil, and the United States. Psychology simply is not preparing sufficient numbers of graduate students to assume leadership in test development as well as to assume responsibility for preparing future generations. The discipline and profession of psychology should unite in its efforts to help ensure these vital resources are sustained.

Professionals in psychology and education are fortunate to have an abundance of standardized tests that assess a broad range of qualities important to their work. Other professions often lack this level of abundance (e.g., audiologists, occupational and physical therapists, as well as specialists in recreation and life care planning). Nevertheless, efforts are being made to help increase the number and quality of standardized tests for these and other health care specialists (Mpofu & Oakland, 2010, in press).

The infrastructure that supports test development and use within the United States is strong. For example, three professional associations directly support the value of test use: American Educational Research Association, American Psychological Association, and the National Council on Measurement in Education. Many other professional associations whose members use tests (e.g., school counselors, school psychologists, speech pathologists) also support testing practices. Additionally, a number of companies specialize in developing and marketing tests. The infrastructure that supports test development and use also is strong in Canada and Western Europe and is emerging in Brazil. However, this infrastructure generally is not found in other regions. Some of the larger test development companies have positioned themselves to market tests internationally.

Commercial firms that help support test development, marketing, and distribution are needed in most regions. Early in the 20th century three psychologists in the United States attempted to locate a publishing company willing to assist in marketing and distributing their tests. Unable to locate one, they formed their own company, one that grew into the largest test publisher in the world and was sold recently for almost \$1 billion. Others within or outside of psychology need to come forward to share the risks and benefits of commercial aspects of test development, marketing, and distribution. The success of these efforts requires methods that contain costs for test development (Ilon, 1992) as well as professional associations to advocate for and respect strong laws that uphold copyright protection of intellectual property.

Professions engaged in test development and use are expected to develop and promulgate standards that address these activities. Although

some national associations have developed ethics codes that address these issues, a review of codes in 31 countries in which test use occurs somewhat commonly found many did not address ethical issues associated with testing (Leach & Oakland, 2007). The *Standards for Educational and Psychological Testing* (AERA et al., 1999) provide the most thorough and authoritative standards for test development and use. The International Test Commission has provided important leadership, especially with respect to test adaptation and computer-based and Internet-delivered testing. The two major international associations of psychologists (i.e., IUPsyS and IAAP) have been somewhat silent on issues impacting testing. Thus, efforts are needed to strengthen the commitment of national and international psychological associations to address ethical issues through standards setting.

Test use is stronger when professionals see value in their use, actively support their development, and restrict their sales. Professional attitudes toward test use generally are strong in Australia, Canada, Israel, Western Europe, and the United States. However, attitudes in other regions can be less positive. For example, prevailing attitudes by psychologists in many countries favor theory over research and thus reject the belief that important personal qualities can be measured reliably and validly. The professional community has considerable control over other conditions discussed above.

The strength of national professional associations of psychology constitutes a bellweather of the strength of psychology within a country. Psychology is strong only when its professional associations are strong. More than 80 national psychological associations serve member interests. The associations vary in their support for test development and use. Conditions are more favorable when a national psychological association displays positive attitudes toward assessment, works actively to promote graduate preparation in this and related areas, and establishes high standards for test development and use. Each country needs one or more national associations dedicated to test development and use to serve as a visible and active advocate.

THREE RECENT EXAMPLES OF EXEMPLARY PROGRAMS FOR TEST DEVELOPMENT AND USE

Until recently, Brazil, Romania, the Czech Republic, Hungary, Latvia, and Slovakia lacked resources for developing and using tests. Many

Brazilian psychologists favored social psychology, theoretical orientations, and psychoanalytic orientations—conditions that did not favor the use of quantitative methods, including tests. For decades, Romania, when ruled from the late 1940s to the 1980s by communist and later another oppressive national government, displayed few positive qualities listed above under conditions external to test development and use. Both recently developed exemplary models for test development and use, ones other countries may be able to emulate.

In addition, the Czech Republic, Hungary, Latvia, and Slovakia shared a need for a quality test of intelligence yet lacked the resources to develop one. These countries are small, were caught within the clutch of communism for more than six decades, and lacked the professional resources to develop a test on their own. The efforts in these countries exemplify many of the issues discussed above. In addition, these efforts give support to the principle that a small and dedicated group of able persons who share a vision, devise a plan, and persist in its attainment can change a country and perhaps the world.

Brazil's Recent Efforts to Develop and Use Tests (Wechsler, 2007)

Efforts of Brazilian psychologists to develop and use tests underscore the importance of using their resources well. Luis Pasquali, Professor of Psychology at the University of Brasilia, assumed early leadership for test development in Brazil. He attended the 1993 International Test Commission conference in Oxford University and was introduced to various methodologies that he incorporated in his work that led to his establishing the first psychometrics laboratory in Brazil. Other laboratories were established at the University of Sao Paulo, Pontifica Catholic University of Campinas, University of San Francisco in Itativa, Federal University of Rio Grande do Sul, and the Federal University of Paravia. Other Brazilian universities are organizing similar laboratories for the development and use of psychological tests.

The national psychological association, the Federal Council of Psychologists, provided political support for forming the laboratories. Laboratory funding comes from federal and state sources as well as the private sector. The Federal Council of Psychologists also established high professional standards for test development and use. For example, in 2003, it created a national task force of researchers responsible for evaluating the psychometric qualities of all tests used in Brazil. The task

force found 50% of the tests to be substandard and prohibited psychologists from using them. This decision was a shock to many engaged in test development and use and motivated them to construct and use suitably developed tests.

A professional association, the Brazilian Institute of Psychological Assessment (www.ibapnet.org.br), was founded in 1997 following the establishment of university-based laboratories and in light of growing interests in developing psychological tests and other quantitative methods in Brazil by Brazilians and for Brazilians. This institute is the only known association within the Americas south of the United States dedicated to test development and use. It publishes a journal and has well-attended biannual conferences, with approximately 1,000 persons attending.

These and other efforts by psychologists in Brazil have had a profound impact on test development and use, resulting in psychological tests being regarded more highly by professionals and the public. Tests have been developed to assess creativity, general intelligence, personality, and vocational aptitudes. The need for nationally normed achievement tests is most urgent. Thus, although test availability remains somewhat limited, a significant expansion of the number and types of tests developed in Brazil has been achieved and more are expected.

In Brazil and many other Latin American countries, individuals initiate and fund test development. Commercial firms generally only publish and/or sell tests. They rarely underwrite their development. Furthermore, test authors typically receive little if any royalties from the sale of their tests. Thus, test distributors aggressively pursue test authors, given their high profit margin. As a result, some psychologists have begun publishing their tests and contracting with commercial firms to distribute them.

Romania's Recent Efforts to Develop and Use Tests (Iliescu & Dincă, 2007)

Scientific psychology in Romania was established at the beginning of the 20th century and later was among the humanistic sciences banned by the communist regime, resulting in the virtual decapitation of this discipline. Psychology survived only by being implicitly tolerated by and embedded within other academic departments, including history, philosophy, education, as well as in some personnel departments (e.g., those with industrial and organizational orientations). Although psychology made a fresh start after 1990, it has not yet succeeded in bridging the serious professional gap (David, Moore, & Domuta, 2002).

The lack of nationally developed measures to assess psychological and educational qualities exemplifies this gap. Many of the renowned measures used internationally (e.g., 16 Personality Factors, California Psychological Inventory, NEO Personality Inventory Revised, Myers-Briggs Type Indicator, Eysenck Personality Questionnaire, State-Trait Anxiety Inventory, and State-Trait Anger Expression Inventory, 2nd edition) have been translated into the Romanian language.

On the positive side, the availability of these instruments facilitated the teaching of these instruments in academic settings and their use in research. On the negative side, the tests do not meet common professional standards. For example, test adaptations were not performed consistent with the International Test Commission guidelines on test adaptations, translations are sometimes poorly done, cultural adaptations occur marginally, tests are validated rarely, and normative samples either are too small or display seriously skewed distributions. Thus, the use of these measures when making high-stake decisions is hazardous. Additionally, copyright infringement occurs commonly by psychologists, an issue that constitutes a serious problem.

During the last four to five years a small number of Romanian test companies have formed in response to the needs of the country's 4,000 to 5,000 psychologists and others. The work of one company is described below. Its efforts can serve as a model for other countries that lack testing resources.

TestCentral™ (www.testcentral.ro) was founded in 2003 as a private initiative, forming relationships between a major local research company (D & D Research, Ltd.), many academic and professional psychologists and researchers, as well as international test companies. TestCentral has rapidly established itself as the main Romanian publisher of psychological tests. The company displays acceptable standards for adapting psychological tests in Romania that are widely regarded nationally and internationally. TestCentral publishes a variety of tests from such publishers as Psychological Assessment Resources, Sigma Assessment Systems, Mindgarden, Hogrefe, Organizzazioni Speciali, Management Research Institute, and Robertson-Cooper. Their tests assess personality, educational, clinical, and quasiclinical values and social axioms, and work/industrial/organizational qualities.

TestCentral used the University of California, Berkeley's Institute of Personality Assessment and Research efforts as a model for its work. Within this model, the company attends to financial goals and other issues keenly important to international publishers who look upon

Romania and other small national emerging markets with both interest and skepticism. The company also is committed to engaging in scholarly activities that lead to high-quality products. These activities recognize the importance of making suitable cultural adaptations, performing adequate norming, and conducting postnorming validation studies. Cultural adaptations and norming of tests at TestCentral always have been consistent with professional standards for test development and adaptation, including those from the International Test Commission (www.intestcom.org).

Test Translations/Adaptations

Translations use consecutive and back-translation methods with dyads of translators, sometimes also coupled with panels of subject matter experts. Consistent with the International Test Commission guidelines for test adaptation, the objective of this phase is to ensure the retention of the original meaning of the test items when translating them into Romanian and to work to ensure the translated items are suitable for use in the Romanian culture and reflect the same trait as that found in the culture in which the test first was developed.

Establishing Test Norms

TestCentral follows two major principles when norming tests: the sample should be nationally representative and stratified as well as sufficiently large. All normative research during the last three years has involved randomized samples of Romanians, stratified by gender, age, education, urban/rural residence, and ethnic group member in light of the most recent national census data. Additionally, the standardization samples are large. For example, the California Psychological Inventory-260 was normed on 3,200 (50% males) and a measure of children's temperament (Oakland, Glutting, & Horton, 1996) on 2,400 students (100 boys and 100 girls from each of the age groups from 7 through 18 years).

Test Validation

Romanian psychologists generally distrust tests that only have been translated. The origins of their reservations stem, in part, from copyright infringement and illegal usage, the poor quality of the translations, and the circulation of many forms of the same test. Due to these and other

reservations, TestCentral was committed to the broader issue of working to ensure the tests were properly validated, not merely suitably translated and normed.

Given the limited resources of TestCentral and the relatively small market for tests, extensive validation that may have been used on tests developed in the host countries is not possible. Nevertheless, every test is subject to a thorough criterion-validation process using observer-evaluations of the targeted behavior and additional empirical research in order to determine whether the results of a test and its scales are consistent with predicted behavioral outcomes.

Extensive studies have provided strong support for many test constructs. Examples include the validity of the California Psychological Inventory (Pitariu & Iliescu, 2004; Pitariu, Pitariu, & Ali Al Mutairi, 1998), the reliability of the Nonverbal Personality Questionnaire (Iliescu, Nedelcea, & Minulescu, 2006), establishing the minimum acceptable age of administration for the measure of children's temperament (Iliescu, Dincă, & Dempsey, 2006), and the use of tests for professional selection in the military (Pitariu & Iernutan, 1984). Mountains of test data await analysis and publication.

TestCentral conceptualizes validation studies in terms of long-term perspectives. One year ago it established its own panel of subjects on whom longitudinal research is being conducted. Data from this panel of 250 individuals are used to examine the co-occurrence of personality traits assessed concurrently by two or more personality tests as well as measures of various behavioral outcomes and indicators (e.g., health and health-related behaviors such as coronary heart disease and Type A behavior pattern; smoking; medication ingestion; frequency of medical visits; frequency and duration of hospitalization and therapy, including psychotherapy and counseling); work and work-related behaviors, including tenure and work performance; family and social relationships; quality of life; and academic achievement.

Further Test Development Efforts

At times, research underscores the need to make major changes, including the development of new scales that best reflect Romanian culture and needs, not simply to adapt and norm tests. This work occurs with the permission and assistance of the original test authors. Some examples include the Romanian Anxiety Scale on the California Psychological Inventory to better fit Romanian culture (Albu & Pitariu, 1999), the

revision of the Romanian Female/Male Scale of the California Psychological Inventory (Albu & Pitariu, 1991; Pitariu, 1981), amalgamating Holland's codes within the Jackson Vocational Interest Survey (Jackson, 2000), and the adaptation of the Fleishman Job Analysis Survey to become the backbone of a national occupational network, mirroring the O*Net used in the United States.

Issues Associated With Fairness

Fairness is an ongoing concern in Romania, especially given the emergence of the profession of psychology. Issues associated with fairness warrant special care when adapting tests and later when selecting and using them. For example, the California Psychological Inventory-462 has been reworked and republished in 1996 as a 434-item version because of legal pressures in light of the Americans With Disabilities Act (Gough & Bradley, 1996). However, the test infringes only marginally in light of Romanian law. Nevertheless, the 462-item version of the California Psychological Inventory was discarded even though this is the most researched version of the famous measure in Romania. The attempt to maximize fairness when using psychological and educational tests represents a main goal when adapting, developing, selecting, and using tests.

Due to the large number of the Hungarian ethnic minority within Romania (approximately 7% of the population), TestCentral has begun translating tests reports and other test materials in Hungarian. Supplementary special norms on special samples for this ethnic population are under development.

Regional Test Development: Efforts by Psychologists in the Czech Republic, Hungary, Latvia, and Slovakia to Develop a Test of Intelligence (Furman, 2007)

The need for additional high-quality psychological tests generally is most apparent in countries that are limited by their small size, those that adopted a socialist political system, established clinical service recently, and have few specialists in psychometrics. Although these countries may need tests, they often lack resources needed for their development. Thus, they are likely to rely on obtaining tests developed in other countries and either translating or adapting them for use in their countries.

In 1998, leadership within the International School Psychology Association decided to attempt to acquire a test of intelligence for use in

the Czech Republic, Hungary, Latvia, and Slovakia—countries with histories characterized by the above-stated conditions that limit test development. The association leadership formed an abiding relationship with Dr. Richard Woodcock, the senior author of the Woodcock-Johnson Tests of Cognitive Abilities—3rd edition, as well as its predecessor, the Woodcock-Johnson Tests of Cognitive Abilities—revised, and Riverside Publishing Company, the test's publisher. Selected tests from this battery were offered for use. The Woodcock-Johnson battery was selected because it assesses a broad spectrum of scientifically identified intellectual abilities that compose the widely accepted Cattell-Horn-Carroll (CHC) theory of intelligence.

The Translation Process

Teams consisting of 15 to 30 psychologists from each of the four countries were formed first to become knowledgeable of CHC theory, then to assist in adapting the battery, and later to acquire standardization data. Dr. Anton Furman coordinated the work of teams in the four countries and facilitated support provided by Dr. Woodcock and his staff throughout the duration of the project. Dr. Woodcock and his staff had considerable experience translating and adapting tests.

The four country teams used consensus translation methods. They differ somewhat from the more commonly used translation/back-translation method. Consensus translation methods rely on the development of two or more adaptations of the target test that are completed independently, typically by at least one linguist and one bilingual psychologist.

Once completed, the adaptations then were discussed and compared by the bilingual psychologists to ensure the adaptations were correct linguistically, suitable for the target culture, and measured the targeted CHC constructs accurately. Dr. Woodcock was consulted, when needed, to ensure this last goal was met. Although this adaptation process may be somewhat more complex and difficult than that commonly used in the translation/back-translation method, it helps reduce the dominant influence of one or two individual translators.

The adapted tests then were reviewed and discussed by psychologists who agreed to collect test data within each of the four countries. Issues again centered on the adequacy of the tests' linguistic translations, suitability for the target culture, and accuracy in measuring the targeted CHC construct.

Collection of Pilot Data

The examiners then were trained to administer the tests. Data on approximately 200 children and youth who varied by age, gender, education, and social class were collected in each of the four countries. These data were provided to Dr. Woodcock and his colleagues who conducted statistical analyses using item response theory and Rasch analysis. As expected, verbal ability tests exhibited the most variability compared to the item difficulty found on the original WJ-R Cognitive scale. Items that did not meet criteria were discarded.

Collection of Standardization Data

The resulting adapted scales were published in booklet form and then used to collect standardization data on 1,000 children and youth in Hungary, Latvia, and Slovakia and 500 in the Czech Republic. Again, Woodcock and his colleagues analyzed these data.

Among the seven factors, only the verbal ability factor showed differences in item difficulties between the original U.S. and target country scales. The four verbal abilities tests (i.e., Picture Vocabulary, Synonyms, Antonyms, and Verbal Analogies) were shortened. Thus, instead of providing separate scores for each of these four scales, this regional version provided one score to represent one's verbal intellectual ability.

The Resulting Test

The final version of the regional battery assesses the following seven broad cognitive abilities: crystallized, fluid, visual-spatial, speed, short-term memory, long-term retrieval, and auditory abilities. The test's factor structure is identical to that of the latest U.S. edition of the Woodcock-Johnson Tests of Cognitive Abilities. However, unlike the source test in which two or more subtests are used to assess each factor, the regional versions used one test for each of the seven broad ability factors. A computer scoring program aided the use of this battery.

This project was successful, in part, due to the abiding leadership of Dr. Anton Furman, the generosity of Dr. Woodcock, and support from the Riverside Publishing Company, Measurement Learning Consultants, and the Woodcock-Muñoz Foundation. The project's success also is due to the many dedicated psychologists in the target countries who invested

their talents, time, and efforts to assist in the test's development; became trained; gathered and coded test data; and assisted in other ways. Their efforts have led to the availability of an intelligence test that meets world-class standards in Latvia, a country in which tests of intelligence were officially banned, and in Slovakia, Hungary, and the Czech Republic—countries that lacked the resources to initiate and complete the development of such a test on their own. The success of this project suggests that other similar test adaptations are possible in other countries that need such tests and lack needed resources.

CONCLUSIONS

Methods to develop and use tests have advanced considerably since their introduction in the 1880s to assist research efforts. Although test use occurs universally, its use is uneven. Test use is more common in countries that see the value of testing, are aware of its positive impact on achieving personal and professional goals and values, and in which a viable infrastructure that supports testing has been developed. Professional standards for test development and use provide direction for those engaged in these efforts.

National test publishers can and should be encouraged to assume leadership in test development within their countries. However, most countries have not established an infrastructure needed to support and sustain test development. For them, the use of adapted tests may present the most suitable immediate solution.

Although some countries may believe locally developed measures may better fit a country's culture, evidence to support this belief is sparse in countries in which test development is being born. Proof of a better fit must be based on comparisons with existing and widely used measures. Judgments as to whether new locally developed measures are superior to adaptations of widely used measures obtained from a host country should be based on research findings.

Tests that are internationally renowned constitute a resource that, when properly adapted, normed, and validated, can serve national interests (Pettifor, 2007). The development of indigenous measures at this point in many developing countries would require the advance of international perspectives and technology—qualities that may be in somewhat short supply. In addition, measures developed there are unlikely to find acceptance among the international community.

REFERENCES

- Aiken, L., West, S., & Millsap, R. (2008). Doctoral training in statistics, measurement, and methodology in psychology. *American Psychologist*, 63(1), 32–50.
- Albu, M., & Pitariu, H. (1991). Algoritm de construire a unei scale pentru un test psihologic: Contributii la reprojectarea scalei F/M a Inventarului Psihologic California (CPI) [An algorithm for constructing a psychological scale: Some contributions to a revision of the F/M scale of the California Psychological Inventory]. *Psychologia-Paedagogia*, Universitatea Babes-Bolyai, Cluj-Napoca, Romania, 36(2), 30–35.
- Albu, M., & Pitariu, H. D. (1999). Evaluarea anxietatii cu ajutorul Inventarului Psihologic California (CPI) [Assessment of anxiety with the California Psychological Inventory]. *Studii de Psihologie*, 4, 19–32.
- American Association on Mental Retardation. (2002). *Definitions, classifications, and systems of supports* (9th ed.). Washington, DC: Author.
- 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. (1995). *Diagnostic and statistical manual of mental disorders* (4th ed.). (DSM-IV: International Version with ICD-10 Codes). Washington, DC: Author.
- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text revision). Washington, DC: Author.
- American Psychological Association. (1992). Ethical principles of psychologists and code of conduct. *American Psychologist*, 47, 1597–1611.
- Bartram, D. (1998). The need for international guidelines on standards for test use: A review of European and international initiatives. *European Psychologist*, 3, 155–162.
- Bartram, D., & Coyne, I. (1998). Variations in national patterns for testing and test use: The ITC/EFPPA international survey. *European Journal of Psychological Assessment*, 4, 249–260.
- Bartram, D., & Hambleton, R. (Eds.). (2005). *Computer-based testing and the Internet: Issues and advances*. New York: Wiley.
- Benjamin, L. (2007). *A brief history of modern psychology*. Malden, MA: Blackwell.
- Canadian Psychological Association. (1987). *Guidelines for educational and psychological testing*. Ottawa, ON: Author.
- Cheung, F. (2004). Use of Western- and indigenously-developed personality tests in Asia. *Applied Psychology: International Review*, 53, 201–211.
- Costa, P. T., Jr., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual*. Odessa, FL: Psychological Assessment Resources.
- Coyne, I., & Bartram, D. (2004). *International Test Commission computer-based and Internet delivered testing guidelines*. Draft 5, March 2002. Paper presented to the Council of the International Test Commission, Beijing, and the People's Republic of China.
- David, D., Moore, M., & Domuta, A. (2002). Romanian psychology on the international psychological scene: A preliminary critical and empirical appraisal. *European Psychologist*, 7(2), 153–160.

- DiMilia, L., Smith, P. A., & Brown, D. F. (1994). Management selection in Australia: A comparison with British and French findings. *International Journal of Selection and Assessment*, 2, 80–90.
- Fitzgerald, C., & Ward, P. (1998). *Computer-based testing: A global perspective*. Paper presented at the International Congress of Applied Psychology, San Francisco, CA.
- Furman, A. (2007). Adaptation of Woodcock-Johnson Tests of Cognitive Abilities in the European context. *World-Go-Round*, 34(5), 10–11.
- Gauthier, J. (2008). Universal declaration of ethical principles for psychologists. In J. E. Hall & E. M. Altmaier (Eds.), *Global promise: Quality assurance and accountability in professional psychology* (pp. 98–105). New York: Oxford University Press.
- Gough, H. G., & Bradley, P. (1996). *California Psychological Inventory manual* (3rd ed.). Palo Alto, CA: Consulting Psychologists Press.
- Gowing, M. K., & Slivinski, L. W. (1994). A review of North American selection procedures: Canada and the USA. *International Journal of Selection and Assessment*, 2, 102–114.
- Hambleton, R. K. (1994). Guidelines for adapting educational and psychological tests: A progress report. *European Journal of Psychological Assessment*, 10, 229–244.
- Hambleton, R., Merenda, P., & Spielberger, C. (Eds.). (2005). *Adapting educational and psychological tests for cross-cultural assessment*. Mahwah, NJ: Erlbaum.
- Hofstede, G. (1994). *Cultures and organizations: Intercultural cooperation and its importance for survival*. Boston: Kluwer Academic Publishers.
- Hu, S., & Oakland, T. (1991). Global and regional perspectives on testing children and youth: An international survey. *International Journal of Psychology*, 26(3), 329–344.
- Iliescu, D., & Dincă, M. (2007, Autumn/Winter). Current advances in test adaptations in Romania. *International Association of Applied Psychology Newsletter*, 11–12.
- Iliescu, D., Dincă, M., & Dempsey, A. (2006, July). *Challenges in the Romanian indigenization and norming of the Student Style Questionnaire*. Paper presented at the Biannual Conference of the International Test Commission, Bruxelles, Belgium.
- Iliescu, D., Nedelcea, C., & Minulescu, M. (2006). Noi alternative în evaluarea personalității: NPQ—Chestionarul Nonverbal de Personalitate [New alternatives in personality assessment: The Nonverbal Personality Questionnaire], *Adaptare si etalonare la populatia României. Psihologia Resurselor Umane*, 11(4), 49–61.
- Ilon, L. (1992). *A framework for costing tests in Third World settings*. Washington, DC: The World Bank.
- Jackson, D. (2000). *Manual for the Jackson Vocational Interest Survey*. London, ON: Sigma Assessment Systems.
- Joint Committee on Testing Practices. (1993). *Responsible test use*. Washington, DC: American Psychological Association.
- Kendall, I., Jenkinson, J., De Lemos, M., & Clancy, D. (1997). *Supplement to guidelines for the use of psychological tests*. Sydney: Australian Psychological Society.
- Koene, C. J. (1997). Tests and professional ethics and values in European psychologists. *European Journal of Psychological Assessment*, 13, 219–228.
- Leach, M. M., Glossoff, H., & Overmier, J. B. (2001). *International ethics codes: A follow-up study of previously unmatched standards and principles*. In J. B. Overmier & J. A. Overmier (Eds.), *Psychology: IUPsyS global resource* [CD-ROM]. Hove, East Sussex, UK: Psychology Press.

- Leach, M. M., & Harbin, J. J. (1997). Psychological ethics codes: A comparison of twenty-four countries. *International Journal of Psychology*, 32, 181–192.
- Leach, M. M., & Oakland, T. (2007). Ethics standards impacting test development and use: A review of 31 ethics codes impacting practices in 35 countries. *International Journal of Testing*, 7, 71–88.
- Lindsay, G. (1996). Psychology as an ethical discipline and profession. *European Psychologist*, 1, 79–88.
- Mpofu, E., & Oakland, T. (Eds.). (2010). *Assessment in rehabilitation and health*. Upper Saddle River, NJ: Merrill.
- Mpofu, E., & Oakland, T. (Eds.). (in press). *Rehabilitation and health assessment*. New York: Springer Publishing.
- Muniz, J., & Hambleton, R. K. (1997, August). Directions for the translation and adaptation of tests. *Papeles del Psicologo*, 63–70.
- Muniz, J., Prieto, G., Almeida, L., & Bartram, D. (1999). Test use in Spain, Portugal and Latin American countries. *European Journal of Psychological Assessment*, 15, 151–157.
- Myers, I. B., McCauley, M. H., Quenk, N. L., & Hammer, A. L. (1998). *MBTI manual: A guide to the development and use of the Myers-Briggs Type Indicator*. Palo Alto, CA: Consulting Psychologists Press.
- Oakland, T. (2004). Use of educational and psychological tests internationally. *Applied Psychology: International Review*, 53, 157–172.
- Oakland, T. (2005). Selected ethical issues relevant to test adaptations. In R. Hambleton, C. Spielberger, & P. Meranda (Eds.), *Adapting educational and psychological tests for cross-cultural assessment* (pp. 65–92). Mahwah, NJ: Erlbaum.
- Oakland, T., & Glutting, J. (1998). Assessment of test behaviors with the WISC-III. In A. Prifitera & D. Saklofske (Eds.), *WISC-III: A scientist-practitioner perspective* (pp. 289–309). New York: Academic Press.
- Oakland, T., Glutting, J., & Horton, C. (1996). *Manual for the Student Style Questionnaire*. San Antonio, TX: Harcourt Assessment.
- Oakland, T., Goldman, S., & Bischoff, H. (1997). Code of ethics of the International School Psychology Association. *School Psychology International*, 18, 291–298.
- Oakland, T., & Hambleton, R. (1995). *International perspectives on academic assessment*. Boston: Kluwer Academic Publishers.
- Oakland, T., & Hu, S. (1991). Professionals who administer tests with children and youth: An international survey. *Journal of Psychoeducational Assessment*, 9(2), 108–120.
- Oakland, T., & Hu, S. (1992). The top ten tests used with children and youth worldwide. *Bulletin of the International Test Commission*, 19, 99–120.
- Oakland, T., & Hu, S. (1993). International perspectives on tests used with children and youth. *Journal of School Psychology*, 31, 501–517.
- Oakland, T., Poortinga, Y. H., Schlegel, J., & Hambleton, R. K. (2001). International Test Commission: Its history, current status, and future directions. *International Journal of Testing*, 1(1), 3–32.
- Pettifor, J. L. (2007). Toward a global professionalization of psychology. In M. J. Stevens & U. P. Gielen (Eds.), *Toward a global psychology: Theory, research, intervention, and pedagogy* (pp. 299–331). Mahwah, NJ: Erlbaum.
- Pitariu, H. (1981). Validation of the CPI femininity scale in Romania. *Journal of Cross-Cultural Psychology*, 12, 111–117.

- Pitariu, H. D., & Iernutan, L. (1984). Utilizarea inventarului de personalitate Freiburg (FPI) in investigarea capacitatii de adaptare la viata militara [Using the Freiburg Personality Inventory for the investigation of adaptation to the military life]. *Revista sanitara militara*, 1(1), 47–55.
- Pitariu, H., & Iliescu, D. (2004). Inventarul Psihologic California—CPI260-Ro [The CPI-260 Inventory]. *Psihologia Resurselor Umane*, 5(2), 40–49.
- Pitariu, H. D., Pitariu, H. A., & Ali Al Mutairi, M. (1998, August). *Psychological assessment of managers in Romania: Validation of a test battery*. Paper presented at the meeting of the American Psychological Association, San Francisco, CA.
- Rossen, E., & Oakland, T. (2008). Graduate preparation in research methods: The current status of APA-accredited professional programs in psychology. *Training and Education in Professional Psychology*, 27, 42–47.
- Schuler, H., Frier, D., & Kauffmann, M. (1993). *Personalauswahl, in europaischen Vergleich* [Personal comparisons in Europe.] Göttingen, Germany: Verlag fur Angewandte Psychologie.
- Shackleton, V., & Newell, S. (1994). European management selection methods: A comparison of five countries. *International Journal of Selection and Assessment*, 2, 91–102.
- Sinclair, C. (2005). A brief history of ethical principles in professional codes of ethics. In J. B. Overmier & J. A. Overmier (Eds.), *Psychology: IUPsyS global resource* [CD-ROM]. Hove, East Sussex, UK: Psychology Press.
- Thiemann, A. (1998). *Digital publishing and test publishers' changing copyright opportunities*. Paper presented at the International Congress of Applied Psychology, San Francisco, CA.
- Van de Vijver, F. J. R., & Hambleton, R. K. (1996). Translating tests: Some practical guidelines. *European Psychologist*, 1, 89–99.
- Wang, Z. M. (1993). Psychology in China: A review. *Annual Review of Psychology*, 44, 87–116.
- Wechsler, S. (2007, Autumn/Winter). Test standards, development, and use in Brazil. *International Association of Applied Psychology Newsletter*, 12–13.
- World Health Organization. (1992). *International statistical classification of diseases and related health problems* (10th revision). (ICD-10). Geneva, Switzerland: Author.
- World Health Organization. (2001). *International classification of functioning, disability and health (ICF)*. Geneva, Switzerland: Author.
- World Health Organization. (2007). *International classification of functioning, disability and health (ICF): Children and youth version*. Geneva, Switzerland: Author.