

# 13th Panhellenic Conference of Psychological Research

## Test Development and Use Internationally

Thomas Oakland  
University of Florida

# Honored to be with you today

- Greece
  - established important principles for a democratic society
  - Provided the philosophical foundation for our discipline through the writings of Aristotle, Socrates, Plato, and others.
  - The word *assessment* has a Greek origin
  - Has a number of excellent scholarly journals
  - Am with colleagues who see value in tests and have interests in
    - Research and other forms of scholarship
    - Individual differences

# Psychological testing: the flagship of applied psychology

- Used prominently by all specialty areas
- To describe current behaviors, estimate future behaviors, provide guidance, help establish interventions, evaluate progress, screen for special needs, diagnose, place, credential, retain/promote, and in research
- Other behavioral sciences envy our resources

# Four Issues to be addressed

- Status of test development and use internationally
- How advanced levels of development were reached in some countries and not in others.
- Country qualities needed to develop and use test
- Changes occurring to which tests need to respond

# Test use is universal

- Some forms of tests are used
  - in every country
  - from birth to death
  - most commonly with students
  - prominently with adults

# Signature historical events

- Origin of testing in China 3000 years ago for its civil service examination
- Other countries did not develop similar methods until the last century.

# Testing Cubicles in China



**Figure 2-1**  
**Testing Booths in China**

# Later signature efforts to develop tests

- Wundt in Leipzig
- Galton in London
- Binet in Paris
- World War I (e. g. Army Alpha and Beta)
- World War II and its aftermath in Western Europe and the US (leading to the development of clinical psychology)



# Test development and use with children

- ~ 600 tests are used frequently
- Measures of intelligence (39%) and personality (24%) are most frequent. Achievement tests are used in only 10% of the countries.

## Test development and use with children

- Imported tests are used more frequently than locally developed tests in 70% of the countries. Deficiencies commonly exist in their
  - Local norms
  - Reliability
  - Validity
  - Ethical practices (e.g., copyright infringement)

# Test development and use with children

- the most widely used measures of intelligence
  - WISC for individual assessment
  - Ravens for group assessment
- Personality tests
  - theory-based tests commonly are used.

## Need for tests for children and youth

- 85% countries report the need for additional tests to assess
  - mental retardation,
  - learning disabilities,
  - emotional and social adjustment, and
  - gifted and talented students

# Test development and use with adults

- A more complex picture
  - clinical practitioners generally rely on imported tests
    - Lack local norms
    - Lack evidence of reliability and validity
    - Need and want more test
    - Considerable consistency within the psychological specialties (clinical neuro-, and rehab) in the tests used

# Test development and use with adults

- Most testing occurs by organizations
  - Thus bypassing psychologists
    - Credentialing: Microsoft and other multinational software firms
    - Employment and advancement: Price-Waterhouse, SHL, and other multinational consulting firms
  - Thomas Friedman's book *The World is Flat*:
    - Anyone can provide assessment services so long as they have a laptop and Internet access

## Common adult tests

- Myers/Briggs Type Indicator
- Minnesota Multiphasic Personality Inventory
- NEO Personality Inventories
- WAIS and Ravens

In summary,  
test development and use are

- Strongest in the US, Canada, and some Western European countries as well as Australia
- Emerging in Brazil and Eastern European countries, including Russia
- Low yet beginning to emerge in Asia and Latin America
- Lower among the 22 Arab countries
- Lowest in the 54 African countries



## Five conditions needed for test development and use

- Test availability requires
  - our society recognizes a need for tests
  - educational institutions that prepare professionals to develop and use tests
  - a test industry infrastructure
  - a sufficiently large number of professionals who use tests wisely and ethically
  - professional associations that establish standards for test development and use and that advocate for test use

## #1: A society recognizes a need for tests

- Test development occurs in response to social and professional needs.
- The public must display positive attitudes toward science, technology, and test use.
- Test use is stronger in organizations and countries that value individual differences and meritocracy over collectivist and egalitarian beliefs

## A need for test consumers include...

- The market for a test must be sufficiently large to justify a test's development.
- Educators, counselors, management specialists, medical specialists, occupational therapists, physical therapists, psychologists, social workers, speech pathologists, and other professionals.
- A test must have a commercial value and thus must be purchased, not photocopied.

## A need for tests and thus test consumers

- Thus, test development and use occur commonly in countries that have
  - publically recognized needs for tests,
  - public attitudes that support test use to solve important social and personal problems, and
  - many professionals who purchase and use tests.

## #2: Educational institutions that prepare professionals who develop and use tests

- University-based undergraduate and graduate programs are needed to prepare specialists in test development and use.
- Professors also commonly develop tests and conduct research that examine a test's psychometric qualities.
- There is a shortage of personnel with expertise in psychometrics

## #3: A test industry infrastructure is needed

- Companies that develop and market tests are needed.
- Companies may be supported publicly (e.g., by the government) or privately.
- They must have sufficient financial and personnel resources to support the development and marketing of tests.

## A test industry infrastructure is needed: an example from the early 20<sup>th</sup> Century

- Publishers would not publish psychological products, believing there was no market.
- In 1921, James Cattell and two students, Robert Woodworth and Edward Thorndike, formed The Psychological Corporation.
- In 2007 it was sold to Pearson for approximately US \$1 billion, thus reflecting its current value

#4: Professional associations are needed to establish standards for test development and use and to advocate for test use

- A profession's strength is seen in the strength and leadership of its professional associations.
  - Locally
  - Nationally
  - Internationally



# Advocacy by professional associations

- Professional associations must
  - value individual differences,
  - be quantitatively oriented,
  - advocate for establishing and enforcing high technical, clinical, ethical, and legal standards for test development and use:
  - advocate for the value of using tests to address important social and personal issues

# Advocacy by professional associations: technical standards

- *Standards for Educational and Psychological Testing* (1999; American Educational Research Association, American Psychological Association, and National Council on Measurement in Education)

# Advocacy by professional associations: clinical standards

- Generally are found in
  - peer reviewed journals
  - best practice documents from professional associations, and
  - authoritative textbooks

## Advocacy by professional associations: clinical standards

- Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision.
- Its International Edition.
- International Classification of Diseases and Related Health Problems, Tenth Edition

# Advocacy by professional associations: ethical standards

- American Psychological Association's 2002 *Ethical Principles of Psychologists and Code of Conduct*

# Advocacy by professional associations: legal standards

- Often create the standards to which we must be most alert.
- Legal issues typically differ considerably between countries and will not be discussed today.
- International Organization of Standards is increasing its influence in reference to testing issues in I/O psychology

# The role of two prominent international associations

- International Union of Psychological Science
- International Association of Applied Psychologists
  - Neither has developed standards for test development or use or ethics codes.
  - However, they have approved the Declaration of Universal Ethical Principles for Psychologists

## Declaration's 4 Principles

- **Respect for the Dignity of Persons and Peoples**
- **Competent Caring for the Well-Being of Persons and Peoples**
- **Integrity**
- **Professional and Scientific Responsibilities to Society**



# The International Test Commission (ITC)

- Its mission is to address cross-national issues that impact test development and use. Members include national psychological societies, universities, test companies, and individuals. Three of its guidelines are discussed below.

## ITC's Guidelines For Test Use

- 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.

# ITC's Guidelines on Test Adaptations

- Guidelines are provided to adapt tests, not merely translate them
  - How 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).
  - They address cultural and language differences, identify technical issues and methods, and describe conditions that possibly impact test interpretations.

# ITC's Guidelines on Test Adaptations

- Two examples:
  - (1) adapting a test originally developed in the United States to one revised for use in Greece or
  - (2) transforming a test from one developed in Greece designed to be used with native-born Greeks to one revised for use in Greece with non-native-born Greeks from Albania.

## ITC's International Guidelines on Computer-Based and Internet-Delivered Testing

- Technology that has an international reach often requires the involvement of international organizations. Computer use is a pervasive technology that has changed the ways in which we work, shop, communicate, play, and test.
- Through computers, testing technology has acquired an international reach and is increasing. The potential for abuse warranted guidelines for test administration, security of tests and test results, and control of the testing process

## ITC's International Guidelines on Computer-Based and Internet-Delivered Testing

- Their purpose is to highlight good practice issues in computer-based testing and testing delivered over the Internet.

## Some of ITC's other initiatives

- test-taker's guide to technology-based testing
- guidelines on assessing persons with language differences
- scoring and reporting test data
- methods that survey health-related issues
- test security
- standards for programs that prepare professionals engaged in test development.

## More on test ethics: Do national ethics codes address test-related issues?

- Ethics codes from 35 countries were compared with standards found in APA's Ethical Principles of Psychologists and Code of Conduct.
- Codes from 15 countries do not address test use, including UK and Switzerland.



## Do national ethics codes address test-related issues?

- The 20 codes that address test use included one or more of the following specific standards found in the 2002 APA Code of Conduct.

## Issues addressed in the 20 codes.

- explaining test results (15),
- using assessment (11),
- avoiding assessment by unqualified persons (11),
- interpreting assessment results (10),
- maintaining test security (10),
- providing informed consent (8),
- using test scoring and interpretation services (7),
- constructing tests (5), and
- using obsolete tests and outdated test results (4).

## More on ethical and legal issues associated with test translations & adaptations

- Two most prominent ethical and legal violations
  - Test plagiarism (i.e., the taking of someone's work product for personal benefit without compensating the author)
  - Non-adherence to copyright provisions (i.e., when tests either are photocopied or are adapted without the consent of the test's author and publisher)

## An example of test ethics among 471 psychologists

- Self-reporting and thus skewed: respondents are more inclined to work and act responsibly
- 77% women
- 44% clinical
- 36% school/educational
- 20% I/O
- Median years practiced: ~ 12

## The ethics of testing practices of 471 psychologists

- 96% use 4 or fewer tests
- 40% never saw the original test materials
- 58% do not use national norms
  - When used, the norms were acquired > 10 years ago
- 96%
  - use 4 or fewer tests
  - use photocopied materials
  - mainly obtained from universities and colleagues
  - do not own copies test manuals they use

## The ethics of testing practices of 471 psychologists

- 80% report having < 5 hours of training on the tests they use
- 86% rely exclusively on their university training
- 99% do not obtain consent in writing

## The ethics of testing practices of 471 psychologists

- However, despite these ethical shortcomings,
  - 95% consider themselves to be qualified to use tests competently
  - Almost 100% report their practices are ethical

## 7 social and professional changes to which test development and use should respond

- Decrease emphasis on traditional forms of diagnosis.
- Increase emphasis on describing behavior within the environmental contexts in which it occurs.
- Emphasize interventions that either develop or rehabilitate behaviors
- Acquire functional information: link data to interventions.



7 social and professional changes to which test development and use must respond

- Improve efficiency of test use
- Higher stakes decisions require higher stakes assessment methods.
- Testing resources are needed in countries that currently lack the resources needed to initiate their development.

## #1 Decrease emphasis on diagnosis

- Knowing a person's diagnosis provides little information on what interventions are needed or the manner in which the interventions should be delivered.
- Thus, the public and professionals want and need to know more than a diagnosis.

# #1 Decrease emphasis on diagnosis

- Our first responsibility is to describe behavior.
- Three broad personal qualities impact behavior that need to be described
  - biology
  - environment
  - personal choice
- Much of the information one needs may be acquired through interviews, observations, and other non-test methods.
- The process of assessing and interpreting test data is enhanced when it is viewed as collaborative activity involving professional and those with whom we work. Everyone has a stake in the assessment process.

# #1 Decrease emphasis on diagnosis

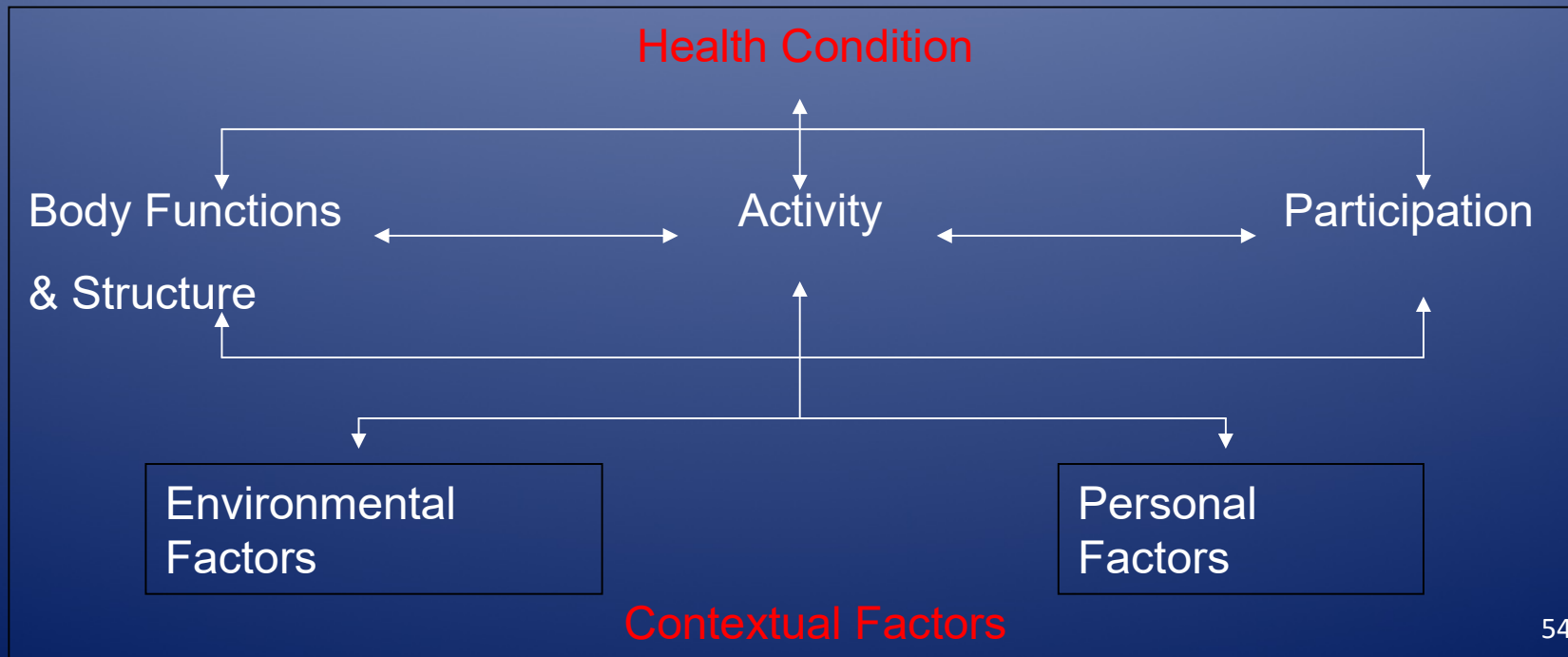
- When conducting clinical work, keep in mind the five pillars of assessment.
    - Assess multiple traits
    - Use multiple assessment methods
    - Acquire data from multiple reliable sources
    - Describe behaviors in multiple settings
    - And over multiple time periods
- We will return to the issue of diagnosis shortly.

## #2 Emphasize describing behavior within the contexts in which they occur.

- The World Health Organization's International Classification of Functioning, Disability and Health (ICF; World Health Organization) provides a bio-psycho-social framework for viewing behaviors from two broad and different perspectives: health conditions and contextual conditions

# ICF's Two Broad Parts

- Health Conditions
- Contextual Conditions



# ICF

- A person's physical, environmental, and personal factors impact
  - Body functions and structures
  - Activities: whether the person has the needed skill
  - Participation: whether he/she uses the needed skill
- These three factors need to be considered within the person's environment and in light of personal choices when describing behaviors

## ICF

- Its purpose is to describe the complex interaction various qualities exert on health
- Its purpose is not to diagnose. Practitioners can continue to rely on the WHO's International Classification of Diseases and Related Health Problems, Tenth Edition, if needed.



## #3 Emphasize interventions that either develop or rehabilitate behaviors

- The effectiveness of interventions increase when we focus on behaviors that
  - are modifiable
  - are small and well defined,
  - occur naturally in one's environment
  - are seen by our client and others as important and
  - the display of the newly learned behaviors is reinforced by one's self and others

Emphasize interventions that either develop or rehabilitate behaviors

- General concepts of intelligence, achievement, personality, adaptive behavior, fine and gross motor development do not meet these standards
- In short, attempts to improve broad concepts and traits important to psychology have not been successful.

## #4 Acquire Functional Information

- Acquire functional information—that which can lead to a meaningful impact on one's life.
- Link data to interventions.
- View behavior as an interaction between the person and his or her environments.

## 8 qualities for emerging standards for assessment

- Acceptable: social worth
- Authentic: natural methods and context sensitive
- Collaborative: involves teamwork
- Evidence-based: disability design/evidence-base
- Multifactored: synthesis of ecological data
- Sensitive: finely tailored measurement gradations
- Universal: equitable design/special accommodations
- Utility: data are useful for interventions

## #5 Improve test use efficiency

- Testing can take many hours and thus may be costly.
- We are being asked to reduce the amount of time needed to collect, score, and report data.

# Improve test use efficiency by

- Constructing two-tier tests: one with many items and a companion test that can be used first to screen the desired behavior.
- Constructing briefer measures
- Having respondents (e.g., family members, teachers, employers) complete test forms on their own

# Improve test use efficiency by

- Acquiring data using the Internet
- Employing technicians to acquire and perhaps score data
- Using computers to record and score data and to draft the initial report

## Improve test use efficiency through item response theory (IRT)

- IRT methods allow us to use item difficulty data to build different forms of a test that are appropriate to each examinee
- Use of IRT methods allows one to identify where each item lies along a continuum from easy to difficult (called “item fit”).
- Knowing this, a test can be designed to include items that represent a range of difficulty and to select only those items that are most discriminating for the person being tested.



## IRT assessment methods

- Thus, using computerized assessment methods, one can more quickly determine one's proficiency level, administer only those items likely to fall within one's proficiency range, thus
  - building tests that are appropriate for each examinee,
  - improving testing efficiency by eliminating the administration of items that are too easy and too difficult.

## Computerized methods for neuropsychological assessment

- Physicians can purchase and install computer-driven assessment devices that provide information on a patient's neuropsychological disorders.
- These methods are
  - Quick
  - Inexpensive
  - Tailored to the physician's interests
  - Do not need psychologists

## #6 Low vs. high stakes testing

- Test data can have different impacts on person's lives.
  - Low stakes impact: whether you pass a spelling test
  - High stakes impact: whether you are diagnosed with a disorder, are admitted to a university, or are hired or fired.

Different standards exist for low and high stakes test use.

# Higher stakes testing requires...

- Higher reliability estimates (e.g.,  $> .90$ )
- Validity data that support test-based decisions
- Tests that minimize bias (e.g., due to cultural and linguistic reasons)

# Higher stakes testing is enhanced by

The use of data collection methods that

- Assess multiple traits
- Use multiple assessment methods
- Acquire data from multiple reliable sources
- In order to describe behaviors in multiple settings
- and over multiple time periods

## An example from I/O psychology: 360 assessment methods

- The purpose of a 360 assessment is to provide information that leads to improved work performance.
- This method is used in most of the Fortune 1000 companies.
- Raters include the employee, supervisor, and others: co-workers and team members, internal and external customers.

## 360 assessment methods

- Test data along with other sources of data (e.g., length of employment, absentees) are communicated to the employee and compared with his/her self-ratings.
- The supervisor and employee discuss the information, leading to the development of a plan to further improve the employee's behaviors.

## 360 assessment methods

- Thus, the use of 360 assessment methods may
  - Assess multiple behaviors
  - Use multiple assessment methods
  - Acquire data from multiple reliable sources
  - And thus describe behaviors in this setting
  - Displayed over some time period.



## #7 Testing resources are needed in countries that have limited resources

- Persons in all countries may benefit from the use of standardized tests.
- Most countries have few locally developed standardized tests.
- Many countries lack needed resources
  - Some will acquire them
  - Others are unlikely to acquire them during the 21<sup>st</sup> Century

## Examples of test development from three countries

- In which psychology as a discipline and profession is emerging

## Brazil: Using its university system to build an infrastructure

- A large country intent on becoming more internationally prominent.
- The discipline of psychology that has been grounded in theory, strong in social psychology and concerned about social justice
- Tests generally are not developed in this climate.
- No companies invested in test development, only sales.

## Brazil: Using its university system to build an infrastructure

- Brazil recently established psychometric laboratories at 8 universities—each working on different types of test
  - Creativity
  - General intelligence
  - Personality
  - Vocational aptitudes
- Their goal is to develop both tests and the next generation of test developers

## Brazil: Using its university system to build an infrastructure

- Professional Association support
    - The national Federal Council of Psychologists provided
      - Political support
      - Established standards for test use.
- 50% of the tests were found to be deficient,  
→ to a wake-up call to psychologists to build better tests

# Two examples of test adaptations

- Czech Republic, Hungary, Latvia, and Slovakia: a general measure of intelligence
- Romania: establishing a company that uses cutting edge methods to develop and market tests.

## Czech Republic, Hungary, Latvia, and Slovakia: a general measure of intelligence

- Countries are small in size, striving to recover from years of oppressive communist rule, and have no specialists in psychometrics.
- Psychologists needed a measure of intelligence to assist in their school-based work.
- International School Psychology Association helped link the countries with needed resources.

## Czech Republic, Hungary, Latvia, and Slovakia: a general measure of intelligence

- Access to the Woodcock-Johnson Test of Cognitive Ability. It was selected because it
  - Is consistent with current CHC theory of intelligence
  - Was made available by the test author, Dr. Richard Woodcock, and the test publisher.
  - Dr. Woodcock selected tests that were most culture fair (e.g., minimized crystallized abilities)



## Czech Republic, Hungary, Latvia, and Slovakia: a general measure of intelligence

- A team of psychologists from four countries adapted the test.
- Psychologists from each of these countries collected data on 500 to 1000 children.
- Dr. Woodcock and his staff conducted the data analyses.

## Czech Republic, Hungary, Latvia, and Slovakia: a general measure of intelligence

- The international version of this intelligence
  - Retains seven broad cognitive abilities
  - Has a factor structure consistent with the original test

The success of this effort could be a model to other countries.

Romania: on the cutting edge of test development, in part, through test adaptation

- Scientific psychology was established at the beginning of the 20<sup>th</sup> Century and after WWII was banned by the communist government.
- Few tests were developed nationally.
- Some reliance on translated internationally renowned measures (e.g., 16 PF, CPI, MBTI, Eysenck Personality Inventory, State-trait measures)

Romania: on the cutting edge of test development, in part, through test adaptation

- During the last few years a small number of test companies have formed to develop and market tests to the country's 4000-5000 psychologists.
- One company, TestCentral, started in 2003, exemplifies this work and is discussed below

Romania: on the cutting edge of test development, in part, through test adaptation

- TC utilizes a model developed at the UC/Berkeley's Institute of Personality Assessment and Research, including
  - attending to financial issues
  - maintaining high professional standards, including
  - adhering to ITC's test adaptation guidelines, adequate norming

Romania: on the cutting edge of test development, in part, through test adaptation

- TC first adapts and then publishes various tests through contractual relationships with many of the international test publishers
  - Psychological Assessment Resources
  - Sigma Assessment Systems
  - Hogrefe
  - Management Research Institute
  - Organizzazioni Speciali

Romania: on the cutting edge of test development, in part, through test adaptation

- Examples include the
  - California Psychological Inventory (N = 3200)
  - Student Styles Questionnaire (N = 2400)
  - Nonverbal Personality Questionnaire

A somewhat large ethnic minority population (e.g., Hungarians) → TC translating reports and other test materials into Hungarian. Supplementary norms also are being developed.

Romania: on the cutting edge of test development, in part, through test adaptation

- Tests are extensively reviewed and revised prior to data collection.
- When being normed, two or more tests may be co-normed, thus assisting in conducting validity studies.
- A panel of 250 persons has been formed with whom long term validity studies are being conducted.



The history of psychology in Greece  
is long and important

- In closing, remember **psychological testing is the flagship of applied psychology**
- Psychology's current services to society in Greece may be enhanced by the further development of its testing resources.

Thank you for allowing me to discuss  
these issues with you today.