Revising of the *Standards for Educational and Psychological Testing:* *Summary of Comments on the Initial Draft*

Moderator: Wayne Camara, College Board

2011 Annual Convention of the APA
August 6, 2011
Washington, DC

Scope of the Revision

- Calls for revision received from sponsor organizations in 2007-08
- Summarized by the Management Committee
  - Wayne Camara, Chair, APA
  - Suzanne Lane, AERA
  - David Frisbie, NCME
- Appointed the joint committee to draft the revisions

Joint Committee Members

- Barbara Plake, Co-Chair, University of Nebraska, Emeritus
- Lauress Wise, Co-Chair, HumRRO
- Linda Cook, ETS
- Fritz Drasgow, University of Illinois
- Brian Gong, NCIEA
- Laura Hamilton, Rand Corporation
- Jo-Ida Hansen, University on MN
- Joan Herman, UCLA

Joint Committee Members

- Michael Kane, ETS
- Michael Kolen, University of Iowa
- Antonio Puente, UNC-Wilmington
- Paul Sackett, University of MN
- Nancy Tippins, Valtera Corporation
- Walter (Denny) Way, Pearson
- Frank Worrell, Univ of CA- Berkeley

Five Key Areas for Revision Identified by the Management Committee

- Access/Fairness
- Accountability
- Technology
- Workplace
- Format issues

Timeline

- First Joint Committee meeting January, 2009

- Release of draft revision January 2011

- Public comment through April 2011

- Revised draft for organizational review and approval projected 2012

- Projected publication late 2013, pending approval by the sponsoring organizations

Presentation Overview

- Description of Initial Draft
- Summary of Comments
  - Fairness – Frank C. Worrell
  - Validity, Reliability, and Operations – Lauress L. Wise
  - Psychological Testing Applications – Antonio E. Puente
- Discussant Reactions
  - Wayne Camara

Key Changes in Initial Draft

- Chapters organized into three sections:
  - Core Principal, Operations, and Applications
- Separate chapters on fairness combined into a single chapter under Core Principals
- Updated discussion of technology issues under test development, scoring, administration
- Discussion of use of tests for accountability
  - Mainly in Education and Policy Application chapters
- Format improvements
  - Standards clustered by topic parallel to background
  - Technical edit for consistency across chapters

Comments on Initial Draft

- 4,000 comments received
  - 30-50% were primarily editorial comments
  - Included official comments from sponsoring organizations:
    - About 12 APA Boards and Committees
    - One set of comments from NCME's Standards and Test Use Committee
    - No official comments from AERA
- All comments reviewed by the committee at our May 2011 meeting
- Each comment will be considered in completing a revised draft over the next 6 (or so) months

Revising of the *Standards for Educational and Psychological Testing:* Comments on the Fairness Chapter

Frank C. Worrell,
University of California - Berkeley

Key Fairness Themes

1. Universal design to minimize barriers to valid test interpretations for all individuals.

2. Validity studies for each of the intended examinee subgroups

3. Appropriate accommodations to remove barriers to the accessibility

4. Guard against inappropriate interpretations, use, and/or unintended consequences

Key Fairness Comments

- Combined chapter received kudos
- But many suggestions
  - Fairness as fundamental validity issue
  - Conflation of LEP and IWD
  - Incorporation of Universal Design
  - Tension between responding to individual vs. subgroup differences
  - "Accommodations" murkier than portrayed (especially validity evidence)
  - Role of opportunity to learn
  - Diversity of examples

Revising of the *Standards for Educational and Psychological Testing:* Comments on Validity, Reliability and Operations

Lauress L. Wise,
Human Resources Research Organization (HumRRO)

Key Comments on Validity

- Chapter largely unchanged from 1999 version, and comments raised no major issues
- New material on consequences was fairly well received
- Some minor suggestions to be addressed:
  - Need for more diversity in the examples provided
  - Clarifications regarding "who is responsible" for meeting some of the standards that are stated in passive voice

Key Comments on Reliability

- Most comments supported the broader focus on precision
  - reliability framework incorporated in this version
- Some concerns with background material:
  - Perhaps too long and sometimes too prescriptive (sounded to more like standards than background)
- A number of comments advocated for more or less coverage of different theoretical models
  - (e.g., generalizability theory, coefficient alpha)
  - May have reflected personal preferences as opposed to an imbalance in the content coverage

Key Comments on Operations

- Scaling and Equating
  - No major objections
    - Many comments editorial in nature
  - Call for more detail on multiple ways to link
    - Scaling and linking for adaptive tests
  - Call for more examples outside of ability and achievement testing
- Test Development
  - Support for increased discussion of design issues
  - Some confusion over design elements that seem to require operational data
    - Need to emphasize iterative nature of design

Key Comments on Operations

- Reporting
  - Other uses of data (e.g., research use; release to media, legislative restrictions)
- Detail of qualification for Test Administrators
- Security (Test Administrator Training)
  - Disclosure of test content to others
  - Different types of cheating
  - To be coordinated with Test Users Chapter)
- Disclosure of test content to others
- Different types of cheating

Revising of the *Standards for Educational and Psychological Testing:* Comments on the Application Chapters

Antonio E. Puente
University of North Carolina at Wilmington

Key Comments on Psychological Testing Chapters

- Most comments editorial in nature or intended to improve clarity
- Call for broader range of examples from different areas of psychological testing
- Questions concerning requirements for administering psychological tests
- Questions concerning requirements for test and test score security
- Few substantive comments on Workplace Testing Chapter

Key Comments on Educational Testing Chapter

- Concern with burden on test developers
  - Particularly on educational agencies
- Questions about specific issues
  - Specific issues with growth scores and value-added models
  - Composites and other derived variables
- Coordination with the Policy Applications Chapter
- Call for more examples from educational counseling (not just achievement tests)

Key Comments on Policy Chapter

- Add psychological testing examples
- Clarify relation to other chapters
  - Coordinate with other application chapters
  - Concerns about going beyond testing
- Public policy concerning use of tests
  - Should the standards support test-based school accountability?
  - Who are "users of information" and who is responsible for what?
- A few technical questions (e.g., use of change scores)