The Chimera of Validity
Bringing the Validity Conversations Home: When Educational Measures Go Public

Eva L. Baker
CRESST National Center for Research on Evaluation, Standards, & Student Testing
UCLA
Educational Assessment, Accountability, and Equity: Conversations on Validity Around the World
Teachers College, Columbia University
29 March 2012, New York
The Chimera of Validity
Chimera

- An entity comprised of distinct animals
- Unreal
- Of the imagination

Validity

- Validity meaning and Test Standards’ Interpretations
- “…When Education Measures Go Public”
- Projections of validity into the future
Quality = Validity

Zen and the Art of Motorcycle Maintenance
R.M. Pirsig, 1974

“Real Quality must be the source of the subjects and objects, the cone from which the tree must start.”
Common Language Validity

1. Well grounded, sound, supportable
2. Producing the desired results; efficacious
3. Legally sound and effective; incontestable; binding
4. a. Containing premises from which the conclusion may be logically derived
   b. Correctly inferred or deduced from a premise as a conclusion
Validate!

- To declare or make legally valid;
- To mark (like a transfer or passport);
- To substantiate and verify (decisive, conclusive, capacity to resist attack, capable of ending a debate)

Conflicting Interpretations

Public & Policy Makers

Testing Professionals

Users
Face Validity
Reasons for Early Scholarly Studies

• Thorndike
• Exploration of human capacities
• Uses for selection
• Normal distributions as foundation of design and interpretation
• Largely assumed stability (fundamentally not subject to change by schools or programs)

“A test is valid for anything it correlates with”
Guilford, 1946
Faces of Key Testing Innovators

- Ralph Tyler
- Art Lumsdaine
- B.F. Skinner
- E.F. Lindquist
- Ben Bloom
- Chester Harris
- Lee Cronbach
- Sam Messick
- Bob Glaser
- Ed Gordon
Standards for Educational and Psychological Testing
“Standards”

• Mid 20th century
• 1999 APA-AERA-NCME sponsored
• Synthesis of accepted state of knowledge and best practice
• Not “aspirational”
• Serve all areas in education and psychology using testing—neurology, clinical, etc.
“Standards” Cont’d

- Revised every decade or so
- Documentary
- Guidance—not enforceable by organizations
- Power through legal precedent, prophylactic
Emerging Notions of Validity

- Unitary concept
  - Not types of validity, like criterion, predictive
- Purpose-driven
- Validity argument–theory & evidence Vs. claims
- Assessments designed to measure learning rather than traits
Key Elements of the 1999 Standards

• Enormous effort to involve the field

• Validity as unitary concept, argued with a range of theory, designs and data

• Purpose and claims

• Guidance for assessment designers involving cognitive demands and other test attributes
1999 Standards

- More standards
- Longer chapters to clarify concepts
- Fairness chapter and standards
- Weak technology chapter
- Thin policy chapter
- **No rating scheme** for Standards, choices depended on professional judgment
- Not universally acclaimed
“Going Public” What Happened?

• Given my hopes, not much
• Swamped by NCLB requirements
• Ambitious schedules for developers
• Restricted testing time
• BOTA
• Cost, cost, cost
Making a Validity Argument

- Consequences
- Meaningfulness
- Fairness
- Cognitive complexity
- Content quality
- Transfer and generalizability
- Instructional sensitivity
- Cost and efficiency

Linn, Baker, & Dunbar, 1991
Key Practical Barriers

- Clarity of standards
- Technical quality of complex tasks
- Evidence of transfer and generalized learning
- Time to conduct studies before use
- Scalable evidence of classroom or other learning experiences
- Confounded by accountability
Challenge of Multiple Purposes

- Assessments for certification, progress and improvement
- Assessments of certification and admissions
- Assessments of children used to determine adult performance (teacher evaluation)
- Research use as dependent measures
Looking to New CCSSs

- Carefully designed—many standards
- Consortia specifications of useful assessments
- Graphics to clarify meaning and relationships to avoid narrowness and trivial alignment
  - Maps, blueprints, ontologies
- Strategies for data management and validity studies
 Ontology Design

DOMAINS

- Specifications
- Relevant texts
- Expert judgment
- Best Practices

Principle

- definition
- requires

Includes components

- requires

Procedure

Concept

- related-to

Fundamental principles

© Regents of the University of California
My Picks

• Unitary validity argument
• Process measures of learning
• Evidence that assessment scores will rise because of high quality instruction/learning
• Translational graphics
• Interrupting predictions
• Choice by students, as in badges and qualifications
Going Public in the Future

• Short term questions
  • *Who is taking the test?*
  • *Hacking*

• Favorite solution: Zacharias—Give up security
  • *Publish everything in advance, menu to guide choice*
Technology Now and Soon

- Technology affects behaviors, brain function, preferences
- Personalized, point-to-point communication
- Crowdsourcing Vs. expertise
- Desire to be noticed—metric: millions over weeks
- Constrained choices
- Weakened formal institutions
Everyone Can Know Everything About You

- Data capture from multiple sources (sensors, cameras, phone)
- Integrated profiles of individuals
- Achievement, tacit preferences or psychological states
- Reports
- Privacy, personal choice, new institutions
- Generational problem
To Educate for the Unpredictable

• Focus on challenging skill learning in content:
  
  *Adaptive problem solving, situation awareness, risk, decisions, metacognitive skills, social and collaborative skills*

• Transfer and generalization tasks—learning to learn

• Emphasize choice
Validity?

- Pertaining to skills
- Managed by different institutions, corporations (who are just like people)
- Quality and mobility issues
- Competition
- Adaptability rather than prediction
Why the Chimera?

• Imagine options
• Imagine validity as a hybrid
• Chimeras are scary
http://www.cse.ucla.edu

Eva L. Baker

email eva@ucla.edu
310-206-1530
Credits for Figures

- FACE:


Credits for Pictures

• Cronbach: http://oaks.nvg.org/lee-cronbach.html
• Gordon:
  http://www.google.com/imgres?q=edmund+gordon&hl=en&client=safari&sa=X&rls=en&bih=986&tbm=isch&prmd=imvnso&tbclid=QPa0aELm65205M:&imgrefurl=http://www.tc.columbia.edu/news.htm%3FarticleID%3D7866&docid=Q1HvISuu5URgSM&imgurl=http://www.tc.columbia.edu/i/media/7866_EdmundGordon.jpg&w=250&h=364&ei=qHVqT8GjCeeWiAL-0iBBQ&zoom=1&iact=hc&vpx=185&vpy=121&dur=6186&hovh=271&hovw=186&tx=79&ty=154&sig=115080220995703348838&page=1&tbnh=154&tbnw=116&start=0&ndsp=39&ved=1t:429,r:0,s:0
• Bloom:
• Tyler:
Credits for Figures

- http://gpaper139.112.2O7.net/b/ss/gpaper139,gntbcstglobal/1/H.3--NS/0" height="1" width="1" border="0" alt=""
- http://www.education.com/reference/article/thorndike-edward-lee-
- Test standards: teststandards.org
- Early test standards Amazon
- http://www.bates-communications.com
- http://www.frontdoor.com
- http://www.bates-communications.com
- http://library.thinkquest.org/04oct/00492/Quiz_pencil.jpg