



**ATP**  
**Innovations  
in Testing**  
Orlando, Florida

**2010**

# **A Standards Based Integrated Test Development and Validation Model**

**James Olsen  
David Scrams  
Chad Buckendahl  
Steve Hegerhorst  
Alpine Testing Solutions**

# Presentation Goals

- *Present a balance of theory and practice for using an integrated test development and validation model.*
- *Give psychometric advice and practical application suggestions for specific elements of the test development and validation model.*

# Agenda

- Theoretical overview of validity in assessments
- An exam-development framework centered around validity
- A cautionary tale

# Validity from the Joint Testing Standards

- **Validity refers to evidence and theory for intended test score interpretations and proposed uses of tests.**
- **Test score interpretations for specific uses are evaluated not the test itself.**
- **What is the rationale and summary of evidence and theory supporting the recommended interpretations and uses of test scores?**
- **Is there a sound validity argument integrating the various strands of validity evidence?**
- **Validity relies on all the available evidence relevant to the technical quality of a testing system.**

# Evidence-Centered Design

## (Mislevy & Riconscente, 2006)

- Domain Analysis
- Domain Modeling
- Conceptual Assessment Framework
- Assessment Implementation
- Assessment Delivery

What are we measuring?

How do we measure it?

Where do we measure it?

How much do we need to measure it?

Knowledge Representations?

Assessment Implementation?

Assessment Delivery?

# **Validity Centered Design**

## **(Bunderson, 2003, Olsen, 2006, 2007)**

### **I. User Centered Design**

- a. Overall Appeal , b. Usability, c. Value and Positive Consequences**

### **II. Design for Inherent Construct Validity**

- a. Test Content , b. Substantive Thinking Processes, c. Test Structure**

### **III. Design for Criterion Related Validity**

- a. Generalizability, b. External Validity, c. Consequential Validity**

# Develop Content: Messick's Three Questions

1. *What complex of knowledge, skills or other attributes should be assessed?*
2. *What behaviors or performances should reveal those constructs?*
3. *What tasks or situations should elicit those behaviors? (Messick, 1994)*

# ATS Test Development & Validation Process Diagram





# Design Program (Purpose)

- Intended claims based on candidate results
- Program mission
- Program goals

# Armchair Psychometrics

- “Should we provide diagnostic feedback to candidates?”
- “Can we write items that focus on product flaws?”
- “Should we incorporate (more) performance-based items?”

# Southern Regional Testing Agency

- Dental and Dental Hygiene licensure examinations
- Clinical judgments and skills
- Primary: Minimum competency
- Secondary: Feedback to training programs

# National Assessment of Educational Progress (NAEP)

- Educational assessment of students in select grades/ages in select content areas
- Inform national achievement and progress
- Secondary goals: Many. . . .

# Program Manager

- Responsibilities
  - Manage client's test development projects
  - Create schedules, schedule internal resources
  - Consult
  - Manage relationships
- Projects from the perspective of a Program Manager

# Case Study

- Real world experience
  - Day and a half into a 3-day JTA management changes scope of exam
    - One exam became two exams
    - Exam's target audience and MQC changed, etc.
  - Facilitator was able to roll with the changes
  - Was not able to finish primary exam in the 3 days
  - Extra time for follow-up calls to finish JTA work
    - Extra costs (fees for consulting and extra SME time)
  - Schedule was pushed out

# Retrospect

- If a Program Design had been done
  - The two exams would have been identified up front
  - Specs for both exams defined
  - Higher quality output from the 3-day JTA
  - Less time and money
  - Schedule would have been met
  - Measure twice, cut once (carpentry maxim) (figurative) “Plan and prepare in a careful, thorough manner before taking action.” (Wiktionary)

# Contact Information

- James B. Olsen, Ph.D.  
[jim.olsen@alpinetesting.com](mailto:jim.olsen@alpinetesting.com)
- David Scrams, Ph.D.  
[david.scrams@alpinetesting.com](mailto:david.scrams@alpinetesting.com)
- Chad W. Buckendahl, Ph.D.  
[chad.buckendahl@alpinetesting.com](mailto:chad.buckendahl@alpinetesting.com)
- Steve Hegerhorst  
[steve.hegerhorst@alpinetesting.com](mailto:steve.hegerhorst@alpinetesting.com)



# Thank You

## Alpine Testing Solutions