# The NHEAI-NAAC Framework for Technical Documentation of Alternate Assessments

Jacqui Kearns & Scott Marion

CCSSO Large-Scale Assessment Conference

June 17-20, 2007

Nashville, TN

NHEAI

National Alternate Assessment Center

New Hampshire Enhanced Assessment Initiative: Technical Documentation for Alternate Assessments

# Overview of Session

- Introduction to the project and framework for technical documentation
  - Jacqui Kearns, NAAC
  - Scott Marion, Center for Assessment (NCIEA)
- State implementation reactions
  - Dan Weiner, Massachusetts
  - Susan Kennedy, Connecticut
- Discussion and insight into the state TAC perspective
  - Peter Behuniak, University of CT

2

# A Sketch of the Projects

- The goal of NHEAI and the first goal for NAAC--design approaches for documenting the technical quality of AA-AAS
  - Expert panel
  - Pilot with four states
  - Revisit with Expert Panel
  - Test with an additional eight states
  - Interact with individual state TACs
  - Produce multiple tools and documents to help states and further understanding of "the field"

3

# There is no "best"

- We are not evaluating tests or systems; rather we are trying to appropriately document the technical quality of whatever system is being used
- We approach this scientifically and collect data to support or challenge our assertions
- The field is so new that there will be no "gold standard" systems
- We need everyone to approach things with an open mind to help us all learn more

# The Problem of Technical Documentation

- Most psychometricians would likely rate validity as the most important technical criterion.
- Yet, most technical manuals include only a superficial treatment of validity.
- In fact, in a recent call for the standardization of assessment technical reports, Becker and Camilli (2004) include validity as part of the required information, but it clearly appears secondary to reliability and other statistical concerns.
  - By their own admission, Becker and Camilli were focusing only on the "nuts and bolts" and expected more would be added to the state's technical manual.

# **Validity Should be Central**

- We argue that the purpose of the technical manual is to provide data to support or refute the validity of the inferences from the alternate assessments at both the student and program level.

- But, it is not so easy…

6

# KWSK

- Fortunately, we are undertaking this work after the publication of *Knowing What Students Know: The science and design of educational assessment* (NRC, 2001), which synthesized a tremendous body of learning and measurement research and set an ambitious direction for the development of more valid assessments. *Knowing What Students Know* (KWSK) builds off of Mislevy's (1996) notion of assessment as a "process of reasoning from evidence" (p. 39).

# The Assessment Triangle

Observation                    Interpretation
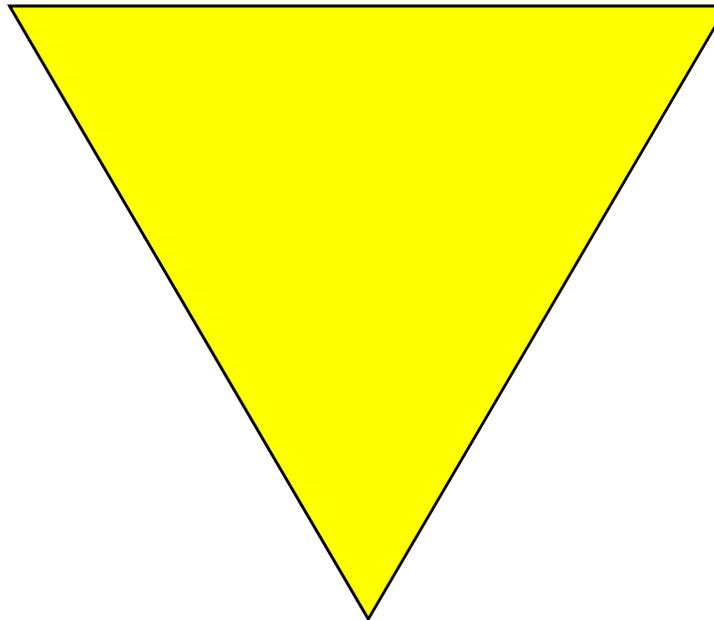
Cognition

# A Heuristic

- We are using the assessment triangle as a heuristic to organize the validity evaluation.
- The triangle immediately reveals an important piece of missing information:
  - When we started this project, we were lacking models of cognition that could be applied generally to students with the most significant cognitive disabilities.
  - However, work by Kleinert, Browder, & Towles-Reeves (2005) and others have started to provide some important insights in this realm.

# The Challenge of Alternate Assessments

- Documenting the technical qualities of alternate assessments is very difficult for many reasons:
  - heterogeneity of the group of students being assessed;
  - relatively small numbers of students/tests; and
  - often "flexible" assessment experiences.
- Additionally, as Linn, et al. (1991) noted, the measurement field has been slow to move away from traditional correlational indicators.

10

# Flexibility and Standardization

- Nominal categories are NOT often useful for characterizing the technical aspects of the assessment (see Gong & Marion, 2006).

- The evaluation of technical adequacy interacts with the types of alternate assessments (i.e., degree of flexibility-standardization) being employed.

- In evaluations of technical quality, state leaders now have some additional tools to consider how they might want to increase or decrease standardization.

- This does NOT mean that standardization is good and flexibility is bad—it all depends on purposes!

# Rethinking Technical Documentation

- As mentioned earlier, this project is based on intense intra- and inter-disciplinary interaction.

- Feedback from both the expert panel and several state TACs have led us to reconsider representing the initially-proposed technical documentation as a single document.

# A Set of Technical Documents

- The "Nuts and Bolts"
- The Validity Evaluation
- The Stakeholder Summary
- The Transition Document

# "Nuts and Bolts"

**Author**: Test contractor or other test developer with state DOE input and review

**Audience**: State DOE, district assessment and special education directors, state TAC members, and others with some technical backgrounds. This also will serve as document for legal defensibility.

**Notes**: This is what we typically expect in technical documents with more attention to validity. The overview of the system and several of the first few chapters will be replicated in both this volume and the validity evaluation volume.

# The Validity Evaluation

**Author**: Independent contractor with considerable input from state DOE

**Audience**: State policy makers, state DOE, district assessment and special education directors, state TAC members, special education teachers, and other key stakeholders. This also will contribute to the legal defensibility of the system.

**Notes**: This will be a dynamic volume where new evidence is collected and evaluated over time.

15

# The Stakeholder Summary

**Author**:  State DOE with the support of consultants

**Audience**:  State policy makers, school and parent stakeholders, business and education communities, special education and general education teachers, and other key stakeholders.

**Notes**:  This volume will essentially serve as a user friendly synthesis drawn from the validity evaluation with additional information drawn from the "nuts and bolts" volume.
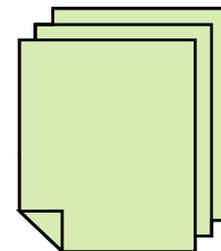
# The Transition Document

**Author**:       Test contractor and state DOE

**Audience**:   Future contractors and state DOE.

**Notes**:        This volume is intended to collect extensive procedural technical documentation so that future test contractors and future state DOE personnel will be able to have a reference for past design and implementation decisions.

# Timing and Publishing

- There is no expectation that the full set of documents be produced each year, but it is crucial that there be a plan for systematic data collection and reporting.

- Many of these documents can be published on the web, which would allow regular updating as new evidence is collected and analyzed.
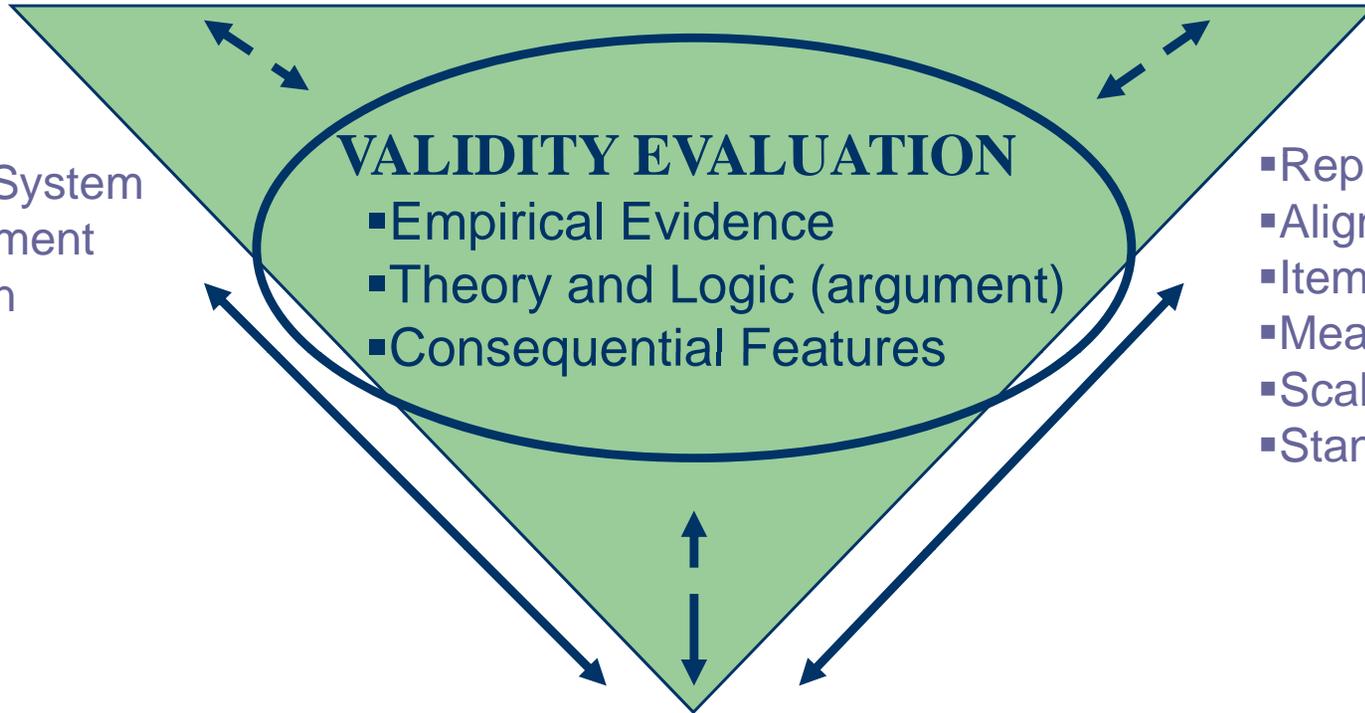
# The Triangle Revisited

- So how does this proposed design for technical document mesh with where we started with the assessment triangle?

- Funny you should ask…

# The Assessment Triangle and Validity Evaluation

**(Marion, Quenemoen, & Kearns, 2006)**

**OBSERVATION** ⟷ **INTERPRETATION**

- Assessment System
- Test Development
- Administration
- Scoring

**VALIDITY EVALUATION**
- Empirical Evidence
- Theory and Logic (argument)
- Consequential Features

- Reporting
- Alignment
- Item Analysis/DIF/Bias
- Measurement Error
- Scaling and Equating
- Standard Setting

**COGNITION**

- Student Population
- Academic Content
- Theory of Learning

# Some Major Challenges

- **Technical**
  - **Defining the Student Population**
  - **Defining the learning progression in academic content**
  - **Alignment**
  - **Comparability**
  - **Characterizing and quantifying error**
  - **Standard Setting**
  - **Validity argument (including consequences)**
- Practical
  - The instantiation of this model will be different in every state due to different policy and cultural contexts
  - That's what we will here about from Dan and Susan

# Addressing the Challenges

- Defining the Population:
  - A conceptual paper "Models of Cognition" mirrors discussion *KWSK*.
    - (In press in Review of Educational Research)
  - Developed a tool to describe/track the population.
  - Four states have provided data sets using the Learner Characteristics Inventory
    - (Towles-Reeves, Kearns, Kleinert, Kleinert, 2007).
  - Preliminary analysis suggests that the population is very consistent across states.

# Addressing the Challenges

- Alignment (NAAC Partner UNCC)
  - Browder, D. M., Flowers, C., Ahlgrim-Delzell, L., Karvonen, M., Spooner, F., & Algozzine, R. (2004). The alignment of alternate assessment content to academic and functional curricula. *Journal of Special Education ,37*, 211-2
  - Browder, D. M., Spooner, F., Wakeman, S. Y., Trela, K., & Baker, J. (2006). Aligning instruction with academic content standards: Finding the link. *Research and Practice for Persons with Severe Disabilities*, *31*, 309-321.
  - Browder, D. M., Wakeman, S. Y., & Flowers, C. (2006). *Level of symbolic communication classification for students with significant cognitive disabilities*. Manuscript submitted for publication.
  - Browder, D. M., Wakeman, S. Y., Flowers, C., Rickelman, R., Pugalee. D., Karvonen, M. (2007). Creating access to the general curriculum with links to grade level content for students with significant cognitive disabilities: An explication of the concept. *Journal of Special Education, 41*, 2-16.

# Addressing the Challenges

- Comparability and Characterizing Error
  - Work shop materials available @ www.naacpartners.org
  - Paper in progress NCIEA.
- Standard Setting
  - Seminar Materials available www.naacpartners.org
  - New paper available through NCIEA

24

# Addressing the Challenges

- Validity argument Consequences
  - What impact on teaching and learning?
  - Teacher and principal surveys available through NAAC – a start but insufficient
  - Multiple measures
  - Comprehensive Evaluation

- Putting it all together
  - "Annotated technical manual table of contents-Volumes 1 & 2" designed to function like a user's guide. Available at www.naacpartners.org

25

# MARCES conference

- Technical quality and technical documentation of alternate assessment systems
- University of Maryland—October 11-12
- www.MARCES.org