

---

# Evaluating Consequential Validity of AA-AAS

---

Presented at OSEP Conference  
January 15, 2008  
by Marianne Perie



---

# Purpose of this Presentation

- Define and summarize the literature on consequential validity
- Discuss the rationale for examining consequential validity of AA-AAS
- Propose hypotheses for testing
- Develop research designs for testing hypotheses

---

# Definition of Validity

- “Validity refers to the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests.” (AERA, APA, NCME, 1999, p. 9)

# Developing a Validity Argument

- Validity argument provides “an overall evaluation of the plausibility of the proposed interpretations and uses of test scores.” (Cronbach, 1988; Kane, 2002)
- In order to develop a validity argument “it is necessary to be clear about what the interpretation claims.”... The interpretive argument “provides an explicit statement of a proposed interpretation/use and a framework for developing a validity argument.” (Kane, 2002, p. 31)

# Consequential Validity

- Consequential validity focuses on the implementation and use of the test as well as decisions made and actions based on test interpretation.
- Concept first appeared in third edition of *Educational Measurement* when Messick, 1989 talked about the consequential basis or aspect of validity.
- Controversy over whether consequences should be considered a part of validity or a separate aspect of test use. Regardless, there is much agreement that the consequences/ results/impact of using a test should be evaluated.

# Application to AA-AAS

- NCLB requires that state assessment systems (including alternate assessments) “be valid for the purposes for which the assessment system is used; be consistent with relevant nationally recognized professional and technical standards, and be supported by evidence of adequate technical quality for each purpose” (NCLB, 2002 §200.2(b)(4)(i,ii))
- Developing a “unified view of validity” is an important part of evaluating and documenting the AA-AAS (Marion & Pellegrino, 2006)
- EAG(s) examine consequential validity of AA-AAS

# Application to AA-AAS (continued)

- Some states have still not had their alternate assessment systems approved by peer review
- While *consequential* validity is not a requirement of peer review, developing a coherent validity argument will go a long way toward satisfying peer review requirements
- Those who are in the early stages should focus on construct validity but incorporate consequential validity into the full validity argument
- Those who are further along should be paying close attention to consequential validity

---

# Considering Consequential Validity from the Beginning

- You should not wait until the assessment has been in place for several years before beginning an evaluation of consequential validity
- Early on, document purposes and intended outcomes of your alternate assessment program
- This will later provide the basis for determining what evidence you should gather on consequential validity



# Think of the Big Picture

- Clearly state the theory behind test development and the ideology that supports test use.
- Develop statements of expected consequences
- Collect evidence to support planned uses and consequences *Reckase, 1998*
- Let's take a moment to do this – jot down some ideas about the effects you expect the AA-AAS to have in your state.

# What Parts of Education Might Statewide Assessment Systems Impact?

- Implemented curriculum
- Instructional content and strategies
- Content and format of classroom assessment
- Student, teacher, and administrator motivation and effort
- Improvement of learning for all students
- Nature of professional development support
- Teacher participation in the administration, development, and scoring of the assessment
- Student, teacher, administrator, and public awareness and beliefs about assessment

*From Lane, Parke, & Stone, 1998*

---

# Additional Impacts of AA-AAS

- IEP development
- Identification rates of special education students
- Attitudes towards special education students
- Self-efficacy of special education students
- Classroom inclusion

---

# Also Consider Negative Impacts

- Narrowing of curriculum
- Cheating
- Bad test preparation activities
- Further stigma on or bias toward special education students
- Changing student designations based solely on test assignment
- Increased turnover of special education teachers

# Develop Hypotheses on How Testing Might Impact Education Systems

- From Lane & Stone, 2002
  - School administrators and teachers are motivated to adapt the instruction and curriculum to the standards
  - Professional development support is being provided
  - Instruction and curriculum will be adapted
  - Students are motivated to learn as well as to put forth their best effort on the assessment
  - Improved performance is related to changes in instruction
- For AA-AAS, add
  - Students are appropriately identified for participation in the AA-AAS
  - Student IEPs are written to reflect appropriately the content assessed

---

# To Test Hypotheses

- Use multiple stakeholders to get multiple perspectives
- Look at outcomes at multiple levels
- Triangulate results using multiple measures
- Cautions:
  - When making causal attributions be sure to rule out alternative hypotheses
  - Be open to finding unexpected (and sometimes negative) consequences

---

# Stakeholders

- Teachers
- Students
- Parents
- School administrators
- Community members
- Policymakers

---

# Outcomes at Multiple Levels

- Consider different levels of outcomes when planning studies
- How does the use of AA-AAS affect the
  - State
  - District
  - School
  - Classroom
  - Family
  - Community



# Multiple Measures

- Surveys
- Focus Groups
  - Gather richer more qualitative data to explain or probe more deeply into survey findings
- Interviews
  - e.g., with district curriculum or special ed coordinator
- Direct measures
  - e.g., instructional materials
- Direct (classroom) observation

# Surveys

- Much work has been done at UKY (e.g., Towles-Reeves & Kleinert, 2006)
- The primary focus of these surveys has been on changes in instruction and IEP development
- They also consider changes in inclusion and perception
- Cautions:
  - May appear biased against finding negative impacts
  - May provide more informative results if baseline data are collected first

# Sample Questions from UKY Surveys

- What benefits to students have accrued from participation in AA-AAS?
- What is the impact of the AA-AAS on students' IEP development?
- What student, teacher, and instructional variables influence parents'/teachers' perceptions regarding the AA-AAS?
- What benefits to teachers have accrued from the participation of students in the AA-AAS?
- To what extent are AA-AAS [activities]\* a part of daily classroom routine?
- What is the impact of the AA-AAS on teachers' daily instruction?
- What is the relationship between student performance in AA-AAS and post-school life outcomes?

\* substitute as appropriate: portfolios, structured tasks

---

# Other Questions to Consider

- Questions about the performance levels:
  - What impact does a student's performance level have on instruction?
  - What impact do the PLDs have on determining the sequence for curriculum and instruction?
- Questions about affect
  - Student and teacher
- Questions about test preparation
  - Look for negative consequences
- Specific questions about score usage
  - Look for both positive and negative consequences
- Questions about costs of using AA-AAS

---

# Focus Groups and Interviews

- Use information from surveys to find areas to explore in depth
- Conduct focus groups or interviews with various stakeholders
- Provides qualitative information to support quantitative information gathered through the surveys

# Direct Evidence and Observation

- Evidence of changes in curriculum or classroom instruction could be gathered and analyzed in a document review
- Evidence of changes in classroom assessment also might be available in document form
- Other changes (e.g., attitude, motivation, time allocation) may need to be observed through systematic classroom observation
  - Researcher
  - School administrator filling out specific form

# Triangulate Results

- Draw conclusions when you have evidence from multiple sources
- Example
  - A teacher says on a survey response that she has changed her instruction to include more math curriculum after seeing the AA-AAS emphasis on academic content. This result is confirmed through examining copies of lesson plans from before and after the teacher implemented the change. In addition, the principal observed the classroom and provided information on the amount of time spent on academic math. A parent survey also indicated that students in this classroom were learning more academic math.
- Caution
  - Remember to disprove alternate hypotheses—are there other reasons for the change in emphasis?

# Prioritizing Studies

- What is critical to know?
  - Give more weight to studies that are central to your premise of how the AA-AAS operates
- What do you really not know?
  - Give more weight to studies that provide new information rather than those that confirm what you already know
- What are the costs
  - May choose the least expensive option, at least to start with
- Timing
  - Some studies will be more appropriate at certain points in the assessment cycle
  - Constantly re-examine the priorities over the life of the assessment program



# Conclusion

- Start by looking at the big picture
  - Ask what impacts you expect the alternate assessment to have and on whom
- Develop hypotheses regarding impacts of AA-AAS
- Determine appropriate stakeholders
- Decide on the best methods for collecting data
- Determine the best timing for collecting each type of data
- Collect and examine data and make adjustments to the AA-AAS program as needed
- Data collection should be continuous over the life of the assessment

# References

- AERA, APA, NCME. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Cronbach, L. (1988). Five perspectives on validity argument. In H. Wainer (Ed.) *Test Validity* (p. 3-17). Hillsdale, NJ: Erlbaum
- Kane, M. (2002). Validating high-stakes testing programs. *Educational Measurement: Issues and Practices*, 21(1), 31-41.
- Lane, S., Parke, C., & Stone, C. (1998). A framework for evaluating the consequences of assessment programs. *Educational Measurement: Issues and Practices*, 17(2), 24-28.
- Lane, S. & Stone, C. (2002). Strategies for evaluating the consequences of assessment and accountability programs. *Educational Measurement: Issues and Practices*, 21(1), 23-30.
- Marion, S. & Pellegrino, J., (2006). A validity framework for evaluating the technical quality of alternate assessments. *Educational Measurement: Issues and Practices*, 25(4), 47-57.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.) *Educational Measurement* (3<sup>rd</sup> ed.). New York: American Council on Education.
- Reckase, M. (1998). Consequential validity from the test developer's perspective. *Educational Measurement: Issues and Practices*, 17(2), 13-16.
- Towles-Reeves, E & Kleinert, H. (2006). Impact of one state's alternate assessment upon instruction and IEP development. *Rural Special Education Quarterly*, 25(3), 31-39.