

Establishing Cutscores for Alternate Assessment on Alternate Achievement Standards

Scott Marion
Center for Assessment
CCSSO-Large Scale Assessment Conference
San Francisco, CA
June 25-28, 2006



Outline of the session

- Scott will present a general framework with some options
- Rachel will provide suggestions for what data and which stakeholders to use to generate PLDs
- Jacqui will present specific information about one possible approach
- Melissa will react from a state DOE perspective



Forthcoming paper

- This presentation is based on a forthcoming paper that goes into considerably detail on all of these points.
- We intend for the paper to be posted on the NCEO, NCIEA, and NAAC websites by the end of the summer and discussed at the October seminars.



Overview of this talk

- A brief background
- Review of four key issues that make this challenging for AA-AAS
- What the regulations say
- Focus in on writing PLDs
- Types and numbers of standards
 - Growth, status
 - Single, multiple



A potential hybrid model

Relax

- This is not an impossible task, but there are some challenges unique to AA-AAS that require us to think harder (sometimes much harder) than we have to for general education assessments



A little review

- Achievement (as opposed to content) standards describe how good is “good enough” for student performances to be classified into specific categories
- Standard setting is the process by which we convert these performance descriptions into operational definitions, i.e., points on the score scale of the test.



A Reality Check

- Cutpoints are set by **policy makers** (hopefully, informed by technical analyses).
- Cutpoints are **NOT** *determined* by panelists or technicians.
- So, does this mean standard setting is an exercise in futility?
 - No, the deliberative judgments help inform the policy and should result in clear, empirically-based performance descriptors.



Methodological Choices

- There are many familiar and legitimate approaches for conducting the actual the operational work—we don't have to invent entirely new methods
- Obviously, not all methods fit with all assessment forms, e.g., it would be tough to use a bookmark or Angoff with portfolio data



Some challenges with AA-AAS

- Heterogeneity of students in the tested population
- Relatively low numbers of students
- Flexibility inherent in the assessment approach
- Writing content-based performance level descriptors (PLDs)

– We discuss each of these in more detail



Student diversity

- Urban legend— “more diversity in this 1% than there is the remaining 99% of students”
- If this is at least close to being true, does it make sense to think about a single standard for the full AA-AAS group?
 - Kearns and her colleagues at UK-NAAC have found at least two distinct groups of students in terms of symbol use
 - There are many good policy/advocacy reasons for setting a single standard



Low numbers of students

- This might be a concern in some of our lowest population states, but in most states it is probably not a significant issue, at least in terms of standard setting
- Depending on the unit of analysis (e.g., grade level or grade span) and the distribution of student scores, it might be worth it to consider the initial standard setting results as “tentative” until they can be validated in subsequent years
 - Actually, many of us would argue for this approach for all assessments



Flexibility in the assessment

- A recent paper (available on the NCEO website) describes the flexibility, or the converse, the degree of standardization for various components of the assessment system
- Flexibility in the learning and assessment goals creates particular challenges when trying to establish a common cutscore with “content-based” meaning



Regulations & Guidance

- USED Peer Review Guidance represents the most up-to-date instantiation of the USED regulations and non-regulatory guidance. Therefore, we use the peer review guidance as the main reference to frame our work.



Critical Element 2.1

- *For students under section 602(3) of the Individuals with Disabilities Education Act with the most significant cognitive disabilities who take an alternate assessment, a State may, through a documented and validated standards-setting process, define alternate academic achievement standards, provided those standards (1) are aligned with the State's academic content standards; (2) promote access to the general curriculum; and (3) reflect professional judgment of the highest achievement standards possible.*



Critical Element 2.1

- First part of this element simply requires states to use a “documented and validated standard-setting process.”
- Second part makes clear that the academic achievement standards must set ambitious goals and be tied to the academic content standards.



Critical Element 2.5

- *If the State has adopted alternate achievement standards, how has the State ensured alignment between its academic content standards and the alternate academic achievement standards?*



Critical Element 2.5

- “Front-end alignment” is a term originally used by Marge Petit and Karin Hess from the Center for Assessment to build the alignment evidence into the assessment design process in the early stages so that a state does not have to wait until a post-hoc external alignment is completed to learn of any misalignment.
- In this context, “front-end alignment” is used to describe the way that the PLDs will be drafted with careful attention to the knowledge and skills described in the states academic content standards.



Critical Element 2.6

- *How did the State document involvement of diverse stakeholders in the development of its academic achievement standards and (if applicable) its alternate achievement standards?*



Critical Element 2.6

- This work will be accomplished by convening a multidisciplinary committee (Element 2.6) comprised of content, curriculum, special education, learning, and assessment experts to consider the content requirements along with the best information about expectations for acquiring proficiency in the respective domains for these groups of students.



Alternate Achievement Standards for Students with the Most Significant Cognitive Disabilities: Non-Regulatory Guidance, USDE (2005, August)

- *May States develop multiple alternate achievement standards to address the range of abilities of students with the most significant cognitive disabilities?*
- *Yes. A State may set more than one alternate achievement standard. If, however, a State chooses to define multiple alternate achievement standards, it must employ commonly accepted professional practices to define the standards; it must document the relationship among the alternate achievement standards as part of its coherent assessment plan; and for AYP purposes it must apply the 1.0 percent cap to all proficient scores based on alternate achievement standards that meet the requirements in 34 C.F.R. §200.1(d)...*



The Performance Descriptors

- The standard setting process must begin with a general descriptor and end with a more specific descriptor—linked to the knowledge and skills associated with the specific cutscores—once student work or test items can be linked to the descriptors
- This is the most difficult step in the process



PLDs

- How to write descriptors tied to specific knowledge and skills for assessment systems with the type of flexibility described earlier is a huge challenge
- The regulations allow for grade level or grade span achievement standards
 - Many are concerned that writing grade level PLDs that can describe unique aspects of performance across adjacent grades is almost impossible, therefore many (including us) advocate writing grade span PLDs.



The Tension

- PLDs that are too general are meaningless
- PLDs that are too specific—even for general assessments—are usually not defensible
- For AA-AAS, including any specificity at all is a challenge



Who Should Draft PLDs?

- I had some ideas about this, but Rachel told me I was wrong 😊
- Rachel will discuss what stakeholders can be most helpful for this task
- These PLDs will be considered drafts until after a standard setting process—my bias is that these descriptors need to be validated with student work not items.



How Many Standards?

- States have the authority to establish more than one set of AA-AAS achievement standards
- But, doing so will require reconciling many psychometric, policy, and advocacy issues
 - Jacqui and Rachel will offer some opinions regarding this important decision



What about Growth?

- NHEAI/NAAC expert panel supported using growth measures as the metric for proficiency.
- Ideally...individualized growth targets...too many uncertainties at this time
- More reasonable approach would involve establishing several different (2-3 different sets) growth targets contingent upon students' starting points



Major Challenge with Growth

- How much is enough?
- Tremendous variability in students' rates of progress
- Types of Goals
 - Policy— “what should be”...tend to be ambitious, but can be unrealistic
 - Empirical— “what is”...achievable, but tends not to advance reform



An “Existence Proof” Goal

- Bob Linn suggested using an “existence proof” when setting NCLB AYP targets
 - Achievable, but ambitious...he suggested that AYP targets be set at the 75th percentile of school growth
 - Using the same idea, we could set goals for the amount of progress typically made by well-instructed students, high-performing students



What do I do tomorrow?

- In the near term, we do not have the data to establish appropriate growth goals
- We can establish a single set of status-based cutscores
- We will collect longitudinal data on students at various points along the score continuum so that we can eventually learn what “75th percentile” growth looks like



How long do we have to wait?

- That depends on the number of students you are testing each year, but it also depends on how long you think it will take until you have “good” instruction
- I suggest that most states should be able to establish appropriate goals in five years or less
- Can set interim growth standards as soon as you have two years of data and then be willing to adjust the cutscores as you gather more data—a good validity process



A Hybrid Model

- While using a growth targets for all AA-AAS students has some appeal, there is some security in still have status-based cutscores
- I recommend a hybrid system where the state establishes
 - a relatively high set of status cutscores, and
 - 2-3 sets (depending on starting point) ambitious growth targets
- A student would be declared proficient if they met EITHER the status or growth targets



For More Information

- To download this presentation and the full paper (in early September):
www.nciea.org
- Contact me: smarion@nciea.org

