Designing Content Targets for Alternate Assessments in Science:
Reducing depth, breadth, and/or complexity

Brian Gong
Center for Assessment

Web seminar on "Best Practices in Science Teaching and Testing for Students with Significant Cognitive Disabilities"
Sponsored by the National Center on Educational Outcomes
March 12, 2007

Context for Policy Decisions

- State assessment in science annually once per grade span (elementary, middle, high school)
- State content standards and state achievement levels
- Alternate assessment of science
  - Content targets "linked" to grade span content
  - Alternate achievement standards
  - Alternate assessments
- Required to include all students; 1% cap on students "proficient" on alternate achievement standards; implement by 2007-08; Peer Review of standards and assessments in science; not in AYP accountability now, but proposed by USED for reauthorization of NCLB

Designing Content Standards

- Content standards are a policy decision
  - Informed by purpose, values, and context
- Content standards for regular assessment form foundation – get these strong *before* developing the content targets for the AA if possible
- Content targets for alternate assessment must be related to grade-level regular content standards – but may have "reduced depth, breadth, and/or complexity"

- We'll first consider establishing regular science content standards, and then consider ways to "reduce" for the alternate assessment, and justify those decisions

Key Steps in Science Assessment Design

1. Be clear on purpose for assessment
2. Design reports and think about uses
3. Identify key science learning targets
4. Identify key science assessment targets
5. Decide what is most important and what should/can be reduced for alternate
6. Check your reduction: item content/skill, cognition/behavior, domain representation
7. Check your system

Be Clear on Your Purposes for Assessment

- Give valid information about student's knowledge and skills in science in relation to state science standards

Not NCLB Essential Purposes

- Encourage inclusion in instructional environment
- Evaluate effectiveness of a program or person
- Encourage teaching of particular content and skills
- Get information about students' performance in order to hold school accountable
- Inform how to help a learner on a particular skill, concept, etc.
- Other purposes

Design Assessment Information and Uses

- Standards-based achievement levels (e.g., Proficient)
  - Regular achievement levels – *"A proficient student can…"*
  - Alternate achievement levels – *"A proficient student can…"*

Not NCLB Designs

- Individual student-referenced – *"right for IEP," "show what student can do no matter if off grade-level" – but… what about growth models?*
- Individual item-referenced – *"got item right 80% of time," "scored a 3, so proficient"*
- Norm-referenced – *"in second quartile of reference group," "in top 90% of this subgroup"*
- Diagnostic – *"what can/can't student do; why; so what should be done instructionally"*

Identify key learning targets

- State science standards – get these right first!
  - What (content standards)
  - How well (verbs, proficiency level descriptions)
  - When (grade span expectations)
- Selection and Prioritization – a challenge in science!
  - Can it be learned well? Is it centrally important? – *e.g., fewer than 15 big concepts per grade/year?*
  - Balance of emphasis – *Life, Earth/Space, Physical, Inquiry, etc.*
  - Depth of knowledge – *facts, skills, concepts, principles, models, evidence, applications, etc.?*
  - Linkages – *within disciplines, across disciplines, math, technology, etc.?*
  - Sequencing w/in grade span – *e.g., if assessed in grade 5, is all content/performance level at grade 5?*

Identify key assessment inferences and claims

- Construct-performance interpretation
  - Inference to proficiency in domain – *"**In science**, the student will likely demonstrate proficient performance…"*
  - Inference to proficiency across situations, time – *"across these types of situations and time periods…"*
  - Inference to proficiency across student performances – *"because the student has performed like this…"*

- What is sufficient evidence for the state and others to justify these claims (reliable and valid)

Decide what is most important and what can be reduced

- Reducing content scope
- Reducing complexity
- Reducing inferential scope
- Prerequisites
  - Reducing grade-level

Reducing Scope

- Select most important content
  - "Big ideas" – most significant science from scientists' point of view
  - Learning power standards – needed to learn other important content and skills
  - Unifying themes – what cuts across, unifies topics and science disciplines
  - Central messages vs. validations, applications, examples
  - Instructional power – most engaging, persuasive, illustrative of science
- Select assessable in this context
- Justify! – primarily content and assessment validity views, *not* "what students are currently learning"

Reducing Complexity

- Identify "Essences"
  - in content
  - in cognition/performance
- Develop multiple entry points of performance
  - More complex to less complex

- Presentation/response
- Prerequisites

Reducing the Scope of Inferences

- Reduce the domain – *e.g., "Provide evidence in three of eight content standards," "These 20 tasks that assess these ten content standards are the assessment target"*
- Reduce the complexity – *e.g., "provide various supports and assistance up to fully independent work," "reduce the depth of knowledge from analyze to identify," "student chooses from six provided options rather than generating options"*
- Reduce the situations, time – *e.g., "Perform in these (few) given situations," "situations are known ahead of time," "situations will not change over time," "Performance is 'banked' before the end of the year," "Must have multiple raters agree"*
- Reduce the number and type of student performances – *e.g., "Best performance counts rather than typical performance,"*
- Should provide justifications for reductions in scope of inferences – more than just a process description!

Prerequisites and "Linkage"

- Within-grade (span) prerequisite
  - Cognitive or content analysis, curriculum

USED guidance and Peer Review practice not clear

- Below-grade prerequisite
  - Grades P-minus, P, K, 1, 2, 3, 4, 5, 6…
    - What is the "grade-level" relationship?
      - The lowest levels are prerequisites for everything above
- Task analysis prerequisite
  - Is the task decomposition retaining the essence, or is it substituting an out-of-grade or out-of-content area procedural skill or factoid? – *example: "Number sense and operation → multiplication → two-digit multiplication → line-up columns procedure → one-to-one correspondence → discriminate symbolic numerals (identify '6,' '7,' etc.)*
- Instructional context as content

Cautions About Inferences from Prerequisites

- Inference: If student doesn't know X, then s/he probably doesn't know Y.
  - If student doesn't know X, then s/he must learn X in order to learn Y.
- Inference: If student knows X, then s/he probably knows Y.

- Cautions – need strong prerequisite chains for claims like these (be especially cautious about instructional claims)
  - "Grade level/grade span" – will the state specify? – many different ways to structure and order topics by grade

Check your reduction

- Item content/skill
- Domain representation per grade span
- Domain cohesion across grade spans
- Evidence required

- Do they make sense?

Check your system

- Correct, Complete, Coherent, Constructive
  - Fair, Useful, Sustainable
- Monitoring validity
  - Validity of assessment inferences about student
  - Validity of assessment design
  - Validity of accountability decisions, uses, and impact
- What are the differences between your alternate and your regular assessment systems? Justified?

For more information:

Center for Assessment

www.nciea.org

Brian Gong

bgong@nciea.org