Implications of Next Generation Assessment Systems for Criteria and Tools to Support Technical Quality: 
*Four conceptual design examples using “Growth”*

Brian Gong

Center for Assessment

Presentation at the 2010 Reidy Interactive Lecture Series, sponsored by the Center for Assessment and WestEd/AACC

October 21, 2010  Cambridge, MA
Context

- Unprecedented consensus, speed, ambition for basis of next generation of operational assessment systems

- Last year: What will Common Core State Standards look like? Who will adopt? What will federal Race to the Top common assessment RFR requirements look like? Who will win? What will they propose?
This year

- CCSS adopted by 36+ states thus far
- Two RTTT common assessment consortia awarded $350 million; involve large majority of states
RTTT Consortia

- Proposals: performance assessment, innovative items, through-course, CAT/CBT, formative/interim/summative systems, deliver by 2014-15
Two challenges

- Can we think of better assessment systems for 2014-15?
  - More valid
  - More useful

- Can we develop tools and criteria to help us design, construct, and implement these systems
  - Frameworks and models of how to design and create
  - Criteria and procedures to evaluate to improve and communicate
My emphasis today

- Call for conceptual clarity first
  - Validity & choices made explicit
  - Design efficiency built in

- Reflect on roles and nature of technical quality tools and criteria
  - Conceptualization (scientific models)
  - Less about evaluation for now
Procedure: Use an example (growth)

- Present an example of what I mean about conceptual clarity and choices

- Ask: What tools and criteria (models) do we have that would help us choose and create these types of assessment designs?

- Ask: Would our current practices get in the way?
Why this example (growth)

- 2009 RILS I presented “three wishes” for future “educational, assessment, and accountability systems”

- Now turn our attention to importance of “model building” and improvement
  - What tools and criteria will help us?
3 wishes for the next generation of assessments, accountability, and educational systems:

- 1: Clearer depiction of student knowledge & skills within a perspective of developing-expertise

- 2: More useful assessment information within a systems perspective of improving learning/organizational functioning

- 3: Powerful theory of actions of how to achieve improved system capacity and commitment, as well as desired outputs; means to improve our theory of actions
Wish 1: Clearer depiction of student knowledge & skills within a perspective of developing-expertise

• Coherent, developmental sequence of content knowledge, e.g., topics/bodies of knowledge

• Coherent sequences of developmental knowledge representations and skills of expertise

• Inclusion of students’ roles and purposes (e.g., independence from teacher, own purposes/goals CONTEXTS, affective, socio-cultural membership)

• Clear specification of desired degree of expertise (three previous dimensions)

• “Existence proof” curriculum & instruction models

• Learner/teacher helps (e.g., common difficulties)
“Refined/Expanded” in...

- **Breadth of Content**
- Deepening or generalizing understanding of the same content; ability to “transfer”
- **Use of more sophisticated reasoning**
  - More powerful representations and reasoning
  - Flexible use of strategies to make meaning and solve problems
  - Metacognition to improve
- **Fluency**

Adapted from K. Hess
Wish 2: More useful assessment information within a systems perspective of improving learning/organizational functioning

- Assessment characteristics
  - Appropriate assessment design & use

- System characteristics for using assessments well
  - Coherent educational systems
  - Key processes to achieve educational goals
  - Assessment information used to inform key processes (primarily formative program evaluation)
Comprehensive, coherent systems

- Assessment information that informs constructive actions in key processes

<table>
<thead>
<tr>
<th>Key Processes</th>
<th>Level of Action</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>National/State</td>
</tr>
<tr>
<td>Set Goals</td>
<td></td>
</tr>
<tr>
<td>Manage Inputs/resources</td>
<td></td>
</tr>
<tr>
<td>Support Learning/Teaching</td>
<td></td>
</tr>
</tbody>
</table>
Wish 3: Powerful theory of actions of how to achieve system capacity, commitment, and desired outputs; means to improve our theory

- Students who not only have learned the “right things” but can and will continue to learn
- Teachers and administrators who can and will continue to learn how to help students learn better and in whatever new situations may arise
- Healthy social systems
  - “Trustworthy measures” versus “Trusted teachers”
- How to get from here to where we want to be – theory and capacity get better
What is desired distributions of student scores – for schools; students in classroom?

- Start: most/all students proficient, variation same as start
- End: most/all students proficient, little variation; equal scores
- End: most/all students proficient, more variation
Growth – Four examples

- Growth and change in performance over time – design of what content/skills to assess over time (target sequence)
- Growth and individualization over time (modifications to target sequence)
- Growth and multiple measures in a validity-evidence framework
- Growth interpretations: content & scales
Procedure: Work from example

- Present an example of what I mean about conceptual clarity and choices
- Ask: What tools and criteria (models) do we have that would help us choose and create these types of assessment designs?
- Ask: Would our current practices get in the way?
1. Growth and assessing target sequence
Common Curriculum & Pacing

Scope & Sequence of Learning Targets

A  B12  C1234  D12  E  F123  G  H123  I  J

Sept  Oct  Nov  Dec  Jan  Feb  Mar  Apr  May  June
Pacing of instruction

C4,
D2,F2,
G, H13, I

End-of-year Assessment
Design of Interim Assessments - 2

Sequence of Learning Targets

A B12 C1234 D12 E F123 G H123

End-of-year Assessment

C4, D2,F2, G, H13

Predictive, Practice Interim Assessments

C4, D2,F2, G, H13

C4, D2,F2, G, H13

C4, D2,F2, G, H13
Design of Interim Assessments - 3

A  B₁₂  C₁₂₃₄  D₁₂  E  F₁₂₃  G  H₁₂₃

A, B₁₂

C₁₂₃₄, D₁₂, E

F₁₂₃, G, H₁₂₃

C₄, D₂, F₂, G, H₁₃

Recent Instruction
Design of Interim Assessments - 4

Cumulative Instruction
Design & Use
(Assessment & Instruction Structure)

A B12 C1234 D12 E F123 G H123

C4, D2, F2, G, H13

A, B12

C1234, D12, E

Interpretation & action

C4, D2, F2, G, H13

A, B12, C1234, D12, E

Interpretation & action

C4, D2, F2, G, H13

A, B12, C1234, D12, E, F123, G, H123

Interpretation & action

C4, D2, F2, G, H13

A, B12, C1234, D12, E, F123, G, H123

Interpretation & action
Procedure: Step 1

- Present an example of what I mean about conceptual clarity and choices – are these differences clear? Are they important choices?
Procedure: Step 2

- Ask: What tools and criteria (models) do we have that would help us create and choose these types of assessment designs?
  - Validity & validation theory
  - Evidence-centered design procedures and criteria (?)
  - Alignment methodologies and criteria
  - Instructional models and research
  - “Learning progressions” (?)
  - Others ?
Procedure: Step 3

- Ask: Would our current technical quality tools, criteria, practices get in the way?
  - Test specification practices
  - Test development practices
  - Peer Review?
2. Growth and individualization over time

- Individual growth that differs from general, target specifications
  - “Super Mastery” / “Mastery of Parts”
  - “Curricular Advancement” / “Master of Prerequisites”
  - “Expertise/Application Development” / “Different Degree of Expertise”
  - “Student Choice”
Tools and Criteria

■ Ask: What tools and criteria (models) do we have that would help us choose and create these types of assessment designs? (e.g., intended vs. implemented curricula; extended learning standards; examples of AP courses & student exhibitions)

■ Ask: Would our current practices get in the way?
3. Growth and multiple measures

| Design Considerations for Multiple Measures in Terms of Validity Concerns |
|----------------------------------------------------------|--------|--------|--------|
| Concern                                                                 | Design                  | Example Assessment Target | Measure 1 | Measure 2 |
| Reduce error due to construct-under-representation when each assessment measure is known to incompletely assess the whole construct | Complementary content (or skills, performance levels, etc.) | AB       | A       | B       |
| Detect/reduce error due to construct-irrelevant variance when occasion or person is thought to be a factor | Repeated administration of the same assessment | A       | A – occasion 1 | A – occasion 2 |
| Reduce error due to construct-under-representation and construct-irrelevant variance when each assessment is known to incompletely assess a part of the construct | Variations in assessing the same construct | A       | A’      | A”      |

Nature & rationales of performance assessments: format, nature of learning, impact on schooling; multiple measures over time
Tools and Criteria

- **Ask:** What tools and criteria (models) do we have that would help us choose and create these types of assessment designs? (e.g., multi-trait/multi-method analyses)

- **Ask:** Would our current practices get in the way?
4. Growth interpretations: content and scales

- **Vertical scale**

<table>
<thead>
<tr>
<th>Grade</th>
<th>Level 1</th>
<th>Level 2</th>
<th>Level 3</th>
<th>Level 4</th>
<th>Level 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>86 - 1045</td>
<td>1046 - 1197</td>
<td>1198 - 1488</td>
<td>1489 - 1865</td>
<td>1866 - 2514</td>
</tr>
<tr>
<td>4</td>
<td>295 - 1314</td>
<td>1315 - 1455</td>
<td>1456 - 1689</td>
<td>1690 - 1964</td>
<td>1965 - 2638</td>
</tr>
<tr>
<td>5</td>
<td>474 - 1341</td>
<td>1342 - 1509</td>
<td>1510 - 1761</td>
<td>1762 - 2058</td>
<td>2059 - 2713</td>
</tr>
<tr>
<td>6</td>
<td>539 - 1449</td>
<td>1450 - 1621</td>
<td>1622 - 1859</td>
<td>1860 - 2125</td>
<td>2126 - 2758</td>
</tr>
<tr>
<td>7</td>
<td>671 - 1541</td>
<td>1542 - 1714</td>
<td>1715 - 1944</td>
<td>1945 - 2180</td>
<td>2181 - 2767</td>
</tr>
</tbody>
</table>
Some sample student sequences of scores

<table>
<thead>
<tr>
<th>Grade</th>
<th>Level 1</th>
<th>Level 2</th>
<th>Level 3</th>
<th>Level 4</th>
<th>Level 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>86 - 1045</td>
<td>1046 - 1197</td>
<td>1198 - 1488</td>
<td>1489 - 1865</td>
<td>1866 - 2514</td>
</tr>
<tr>
<td>4</td>
<td>295 - 1314</td>
<td>1315 - 1455</td>
<td>1456 - 1689</td>
<td>1690 - 1964</td>
<td>1965 - 2638</td>
</tr>
<tr>
<td>5</td>
<td>474 - 1341</td>
<td>1342 - 1509</td>
<td>1510 - 1761</td>
<td>1762 - 2058</td>
<td>2059 - 2713</td>
</tr>
<tr>
<td>6</td>
<td>539 - 1449</td>
<td>1450 - 1621</td>
<td>1622 - 1859</td>
<td>1860 - 2125</td>
<td>2126 - 2758</td>
</tr>
<tr>
<td>7</td>
<td>671 - 1541</td>
<td>1542 - 1714</td>
<td>1715 - 1944</td>
<td>1945 - 2180</td>
<td>2181 - 2767</td>
</tr>
</tbody>
</table>
Interpretations of growth?

- Same scale scores, same Achievement Levels, different grade levels

| Three Students, Same Scale Scores and Achievement Levels, Different Combinations of Tests |
|-----------------------------------------------|-----------------|---------------|-------------|-------------|---------------------|
| 1300 (Level 1) | 1400 (Level 2) | 1650 (Level 3) | 1950 (Level 4) | Interpretation of Growth |
| Student A | Grade 4 | Grade 4 | Grade 4 | Grade 4 |
| Student B | Grade 5 | Grade 5 | Grade 5 | Grade 5 |
| Student C | Grade 4 | Grade 5 | Grade 6 | Grade 7 |
Tools and Criteria

- Ask: What tools and criteria (models) do we have that would help us choose and create these types of assessment designs?

- Ask: Would our current practices get in the way?
Putting it together

- Growth four aspects (intended target sequence, individualization, multiple measures, interpretations [content, scales])
- Three wishes (educational, assessment, uses – theory of actions)
- Tools, criteria for design and improvement
For more information:

Center for Assessment

www.nciea.org

Brian Gong

bgong@nciea.org