Psychometric Support for the Transition towards the CCSS-Aligned Assessments

Presentation to the 2013 CCSSO National Conference on Student Assessment

Yi Du
Dave Chayer
Data Recognition Corporation
June 20, 2013

Transition to CCSS

- In an effort to align the state assessments to the CCSS, numerous states have modified their assessment programs in a variety of ways:
  - Curriculum and classroom instruction
  - Test designs
  - Content standards and blueprints
  - Item types and coverage
  - Test administration protocols and modes

During transition years,
- Can a state assessment program maintain scale continuity?
- How can meaningful achievement comparisons over the transition years be performed?
- How can we ensure the state achievement comparisons are valid, fair to all subgroups, and ultimately, useful for improving teaching and learning?
- How can we assess the impact of the changes to assessments and make appropriate score inferences?
Psychometric Challenges In the Transition

• Comparisons under a continuous scale over the transition years are very attractive
• A number of considerations on the validity of the results of the linking must be addressed
• Linking can be achieved statistically, but the meaning of the scores and performance levels in the comparison could be questionable.
• Comparability of assessments over years might be determined by the degrees of similarity of the assessments and levels of comparability

Variability of Assessments during Transition

– Content standards and testing blueprints
– Types of items and coverage
– Test administration protocols and modes of assessments
– Quality of linking items and linking approaches
– Score inferences and interpretations

Rigor of the Comparisons Over the Transition Years

• The level and nature of the comparisons:
  – The Level of Comparisons
    • State
    • District/school
    • Individual students
  – The Nature of Comparisons
    • Rigor of performance levels
    • Group performance on the performance levels within a year or over years
    • Individual students' performance on the performance standards
    • Skills profiles for a specific group or individual
    • Individual growth
• Some comparisons can be made through various linking approaches, but not all.
Rigor of the Comparisons Over the Transition Years

• When two assessments have great similarity in content framework, structures, and items, the linking can be strong
• Some variability in the assessments can be tolerated, but additional data collection designs and analysis may be required
• Some variability may damage the scale continuity. We may have to treat the modified assessment as a new one and validate the performance levels.

Purposes of The Presentation

• Present our thoughts and approaches from the state transition assessments on
  • What may be possible in the ways of comparing results over the transition years?
  • How much variability can be tolerated for such a comparison?
  • What kinds of linking approaches are useful?
  • What alternative approaches may be used to ensure the validity of the transition assessments?
• Most importantly, we attempt to continue this discussion with colleagues for valid, meaningful, and fair comparisons over transition years.

State Transition Assessments

• The state assessments began to be aligned to the more rigorous CCSS in 2012
• Multiple steps and procedures have been used during the transition:
  – In 2012-2013, small transition occurred
  – In 2013-2014, complete shift to CCSS
**Transition Process**

<table>
<thead>
<tr>
<th>Year</th>
<th>English Language Arts</th>
<th>Mathematics</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012-2013</td>
<td>The writing Prompts in ELA Tests</td>
<td>1. Aligned to the State Standards (GLE).</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2. The blueprints changed based on GLE.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3. About 30-40% of common items remained.</td>
</tr>
<tr>
<td>2013-2014</td>
<td>1. Aligned to the CCSS</td>
<td>1. Aligned to the CCSS</td>
</tr>
<tr>
<td></td>
<td>2. Blueprints changed</td>
<td>2. Blueprints changed</td>
</tr>
<tr>
<td></td>
<td>3. Included CCSS items</td>
<td>3. Included CCSS items</td>
</tr>
</tbody>
</table>

**Similarities between 2012 and 2013 Assessments**

- The State curriculum and instruction
- Measure similar constructs, especially in ELA.
- Most components of the assessments measure similar knowledge and skills
- Performance level descriptors
- Item format
- Substantial number of common items were included in both tests
- Testing administration protocols were the same in:
  - Testing time
  - Conditions of test administrations
  - Rules for granting testing accommodations
  - modes

**Psychometric Strategies for 2013 Transition**

The similarities allow for the continuity of the scale

- IRT linking approaches and other psychometric analysis were used
- 2013 assessments were linked back to 2012 OP scale
- Comparability was established in terms of scores of the assessments over the two years
Psychometric Considerations for 2013 Transition

Issues may be addressed with linking analysis:
  – Possible context effect on the assessment
  – Validity of the transitional assessments
  – Scale continuity over the transition years
  – Inference of the results

Additional data collection designs and analysis are required for the linking.

Additional Data Collection for the Math Transition in 2013

• In 2012, the newly-developed items were field tested in 2012.
• These field tested items were linked to the State baseline scale through common-student linking.
• These linked items were included in 2013 assessments.
• Including the linked 2012 FT items made the item coverage more balanced.

2013 Math Common Items and Tests

• The number and percentage of common items for 2013 forms meet psychometric requirements.
• Operational common items represent content domains for most grades.
• After adding the linked FT items, common items represent content domains well for all grades.
### Original Number and Percentage of Common Items in 2013 Math

<table>
<thead>
<tr>
<th>Grade</th>
<th>Total Points on the Form</th>
<th># Common Items</th>
<th>% Common Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>54</td>
<td>21</td>
<td>39%</td>
</tr>
<tr>
<td>4</td>
<td>76</td>
<td>27</td>
<td>36%</td>
</tr>
<tr>
<td>5</td>
<td>58</td>
<td>16</td>
<td>28%</td>
</tr>
<tr>
<td>6</td>
<td>68</td>
<td>32</td>
<td>47%</td>
</tr>
<tr>
<td>7</td>
<td>68</td>
<td>17</td>
<td>25%</td>
</tr>
<tr>
<td>8</td>
<td>76</td>
<td>29</td>
<td>38%</td>
</tr>
</tbody>
</table>

### Number and Percentage of Anchors with Additional Data Collection

<table>
<thead>
<tr>
<th>Grade</th>
<th>Total Points on the Form</th>
<th># Anchor Points</th>
<th>% Anchor Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>54</td>
<td>44</td>
<td>81%</td>
</tr>
<tr>
<td>4</td>
<td>76</td>
<td>60</td>
<td>79%</td>
</tr>
<tr>
<td>5</td>
<td>58</td>
<td>40</td>
<td>69%</td>
</tr>
<tr>
<td>6</td>
<td>68</td>
<td>46</td>
<td>68%</td>
</tr>
<tr>
<td>7</td>
<td>68</td>
<td>42</td>
<td>62%</td>
</tr>
<tr>
<td>8</td>
<td>76</td>
<td>57</td>
<td>75%</td>
</tr>
</tbody>
</table>

### Content Representativeness of Common Items

<table>
<thead>
<tr>
<th>Grade</th>
<th>Content Strands</th>
<th>Possible Anchor Points</th>
<th>Form Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>Measurement Data and Geometry</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>Multiplication and Division</td>
<td>19</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>Number</td>
<td>14</td>
<td>24</td>
</tr>
<tr>
<td>5</td>
<td>Fractions</td>
<td>26</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>Measurement Data and Geometry</td>
<td>13</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>Number and Operations</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>Measurement Data and Geometry</td>
<td>13</td>
<td>16</td>
</tr>
<tr>
<td>6</td>
<td>Number System</td>
<td>24</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>Ratio Probability and Algebra</td>
<td>26</td>
<td>28</td>
</tr>
<tr>
<td>7</td>
<td>Measurement Data and Geometry</td>
<td>10</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>Number System</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>Ratio Probability and Algebra</td>
<td>37</td>
<td>40</td>
</tr>
</tbody>
</table>
2013 Linking and Verification

• The IRT model and the common-item linking design were used for 2013.
• The validity of the linking were evaluated carefully.
• The cut scores through linking procedures were evaluated to examine whether the cut scores “truthfully” or “close to truthfully” preserved.

Evaluation of Common Items

• Common items are sensible indicators for the direct comparison of results over the two years.
• Common items are representative in the standards in both years.
• The difficulty level of the common items are stable in both years.
• DIF analysis was used to ensure the item functioning was not affected by
  – Administration conditions in 2012 FT or 2013 OP
  – The other kinds of items being presented along with the items
Cross-Validation for Linking

• Equipercentile linking approach provides a reasonableness check
• The results from both IRT and equipercentile linking methods show consistency in the location of the cut scores and percentages of students in the performance levels.
• Results and impact data from other subjects that didn’t have transition, i.e., science and social studies, were used as references within a grade.
• Verify results and impact data across grades and subject
• Results show that the linking is reliable.
Inferences and Interpretations in 2013

• The most rigorous comparisons are the scores in Science and Social Studies.
• Scores in ELA may be more rigorously comparable as those in Math.
• The scores on writing may not mean the same in terms of writing competence.
• General trends on percentages of students at performance levels over years for the state, districts, and schools are possible.

Psychometric Strategies for 2014

• Due to various degrees of transition, psychometric approaches for 2013 and 2014 assessments are different:
  – In 2013, a variety of linking approaches were applied to 2013 transition assessments
  – In 2014, standard validation and additional research studies will be conducted to ensure the validity of assessment results.

New Components in 2014 Transition

• The State implemented CCSS in curriculum and instruction.
• The state assessment programs were aligned to CCSS.
• The test blueprints have been aligned to CCSS.
• The content coverage and weights changed.
• Items from previous operational administrations that did not align to CCSS were removed.
• Newly-developed field tested CCSS items based on CCSS are included.
• The rigor of the 2014 assessments changed.
Psychometric Challenges for 2014

- Instruction effect
- Different content standards, CCSS versus State’s previous standards
- Context effect
- Items representative and coverage of the standards

Without additional psychometric studies and standard validation:
- the validity of the assessment results will be questionable.

Psychometric Support for 2014

- Standard validation will be used to revisit the performance level descriptors and cut scores.
- Double linkages were used in 2013 CCSS item FT
- Double linkages were used in 2014 OP
- Common items will be selected into two sets:
  - Previous OP items that linked to the state baseline scale,
  - CCSS items developed in 2013, possibly linked to the state baseline scale.
- Two linking sets may produce two different cut scores.

Standard Validation

- Two linking results will be presented to standard validation committees.
- The prior achievement level and cuts could be used as “scale anchoring”
- Two different cuts will be reviewed
- Two different sets of items (previous OP and CCSS) will be reviewed at a single achievement level, i.e. Basic.
- This comparison will identify differences in terms of rigorous levels between previous standards and the CCSS and help revise the description of the performance descriptors and the cut scores.
Possible Achievement Trends over Years Statewide (Hypothetical for 2014)

We hope that standards validation and additional psychometric analysis can be helpful for the State to develop an “Aspirational Performance Package” for schools:

- The blue line shows where the state was based on the state old standards
- The red line shows where the state was based on the actual CCSS standards

References

- Lazer, Mazzeo, Way, Twing, Camara, & Sweeney, 2010. Thoughts on Assessment of Common Core Standards.