FCAT 2.0 Science, Biology 1, and Geometry EOC Assessments Standard Setting

Pam Stewart
Interim Commissioner,
Florida Department of Education

Stephen Murphy
Pearson

September 18, 2012
Welcome & Introductions

- Florida Educators
- Florida Department of Education
- Pearson
Overview – What are you doing here?

• To review thoroughly the content requirements of the FCAT 2.0 Science, Biology 1 and Geometry EOC Assessments

• To provide your individual judgments about that content

• To help the State of Florida establish achievement level standards for these assessments
Why have standards?

• To define what students should know and be able to do

• To identify clear expectations for students, parents, and teachers

• To improve teaching and learning

• To develop a society able to compete in a global economy
Types of Standards

- **Content Standards:** Define desired student knowledge and skills (the “what”)
  - Next Generation Sunshine State Standards
  - Common Core State Standards

- **Performance Standards:** Describe how much content knowledge a student is required to demonstrate
  - **Achievement-Level Standards**
  - **Graduation Requirement**

- **Accountability Standards**
  - School Grading Criteria
  - Adequate Yearly Progress
Past Experience - FCAT Science Standards

FCAT Science by Achievement Level
Grade 5

Achievement Level 3 and Above (On Grade Level and Above)
Past Experience - FCAT Science Standards

![Graph showing FCAT Science by Achievement Level for Grade 8 from 2003 to 2011. The graph indicates a trend where the percentage of students achieving Level 3 and above has generally increased, while the percentage at Level 1 has decreased.](image-url)
External Data: NAEP Science 2009 – Grade 4 (Mean Scale Score)
External Data: NAEP Science 2011 – Grade 8 (Mean Scale Score)
NAEP Results: Grade 4 and Grade 8 Science (Percentage at or above Proficient)
Partnership for Assessment of Readiness for College and Careers (PARCC)
PARCC Assessments

- English, Language Arts/Literacy assessments in Grades 3-10
- Mathematics assessments in Grades 3-8;
- End-of-Course assessments (EOCs) in Algebra 1, Geometry, Algebra 2.
PARCC Timeline

**SY 2010-11**
Launch and design phase

**SY 2011-12**
Development begins

**SY 2012-13**
First year pilot/field testing and related research and data collection

**SY 2013-14**
Second year pilot/field testing and related research and data collection

**SY 2014-15**
Full administration of PARCC assessments

**Summer 2015**
Set achievement levels, including college-ready performance levels
# Transition Schedule

<table>
<thead>
<tr>
<th>Type of Assessment</th>
<th>Assessment Area</th>
<th>Year Administered to Students</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>FCAT</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>FCAT Writing</td>
<td>Gr 4, 8, 10</td>
</tr>
<tr>
<td><strong>FCAT 2.0</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>FCAT 2.0 Writing</td>
<td>Gr 4, 8, 10</td>
</tr>
<tr>
<td></td>
<td>FCAT 2.0 Reading</td>
<td>Gr 3-10</td>
</tr>
<tr>
<td></td>
<td>FCAT 2.0 Mathematics</td>
<td>Gr 3-8</td>
</tr>
<tr>
<td></td>
<td>FCAT 2.0 Science</td>
<td>Gr 5, 8</td>
</tr>
<tr>
<td><strong>End-of-Course Assessments</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Algebra 1</td>
<td>In Course</td>
</tr>
<tr>
<td></td>
<td><strong>Geometry</strong></td>
<td>In Course</td>
</tr>
<tr>
<td></td>
<td>Biology 1</td>
<td>In Course</td>
</tr>
<tr>
<td></td>
<td>US History</td>
<td>In Course</td>
</tr>
<tr>
<td></td>
<td>Civics (Middle School)</td>
<td>In Course</td>
</tr>
<tr>
<td><strong>Partnership for Assessment of Readiness for College and Careers (PARCC)</strong></td>
<td>English language Arts</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mathematics</td>
<td></td>
</tr>
<tr>
<td></td>
<td>High School Math EOCs</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(Algebra, <strong>Geometry</strong>, Algebra 2)</td>
<td></td>
</tr>
</tbody>
</table>
Your Mission

• Separate these concepts
  – Achievement Levels
  – Accountability

• Focus on student achievement
  – Levels of success with the challenging content of the Next Generation Sunshine State Standards

• Set high standards for Florida’s students
Overview of the Process

• Educator judgments

• Superintendent/Business Community reactions

• Department of Education reactions

• State Board of Education Rule advertised

• Public hearings to provide reactions

• Adoption by the State Board of Education
Pearson Staff Roles

- **Lead Facilitator**
  - Lead large-group sessions
  - Process oversight
  - Floater

- **Facilitator**
  - Provides training on standard setting procedure and leads breakout session activities

- **Content Specialist**
  - Available to answer questions related to item content

- **Data Analyst**
  - Computes feedback data between rounds

- **Program Team Staff**
  - Reimbursement questions and other logistical issues
Confidentiality

• **DO NOT**
  - Discuss the test items outside of this meeting
  - Remove any secure materials from the room on breaks or at end of day
  - Discuss judgments or cut scores (yours or others) with anyone outside of the meeting
  - Discuss secure materials with non-participants
  - Use cell phones in the meeting rooms (Please turn your cell phone ringer off.)

• “What happens in the meeting room stays in the meeting room”

• General conversations about the process and days’ events are okay, but participants should avoid discussing details, particularly those involving items, cut scores, and any other confidential information

• Notes should be taken using provided materials only

• The only materials allowed on the table are standard setting materials
Purpose of the Meeting

• Why you are here
  – Standard setting relies on expert judgments from individuals who are knowledgeable about the test content and the population of test-takers
  – This is one step in a larger process

• What we will do
  – Over the next four days, you will provide expert judgments that will be used to form recommended cut scores

• How we will set standards
  – We will use a popular technique that is widely used to set standards for large-scale assessments
When is Standard Setting Necessary?

• Standard setting becomes necessary whenever any of the following occur
  – New test
  – Curriculum updates
  – Blueprint changes
  – Achievement Level Description changes

• Next Generation Sunshine State Standards – content standards
Setting Standards is a Multi-Stage Process

Achievement Level Descriptions → Educator Panel → Reactor Panel

State Board of Education ← Public Input ← Commissioner’s Recommendations
What is Standard Setting?

- A process of deriving levels of performance on educational or professional assessments, by which decisions or classifications of persons will be made (Cizek, 2006)

- Test scores can be used to group students into meaningful achievement levels

- Standard setting is the process whereby we “draw the lines” that separate the test scores into various achievement levels
Setting Performance Standards

- ALDs
- Content Standards
- Test
- Student Knowledge / Expertise

Cut scores that match students to their appropriate performance categories
Standard-Setting Vocabulary

- **Content Standards**: The content that students are expected to know
- **Achievement Levels**: Levels of student achievement based on observed scale scores
- **Achievement Level Descriptions (ALDs)**: Descriptions of the competencies associated with each level of achievement
- **Cut Scores (Standards)**: Scores on an assessment that separate one achievement level from another
- **Panelists (Judges/Raters)**: Those who participate in the standard setting process (stakeholders, educators, professionals – must understand the content assessed)
Standard-Setting Vocabulary (Continued)

- **Feedback Data**: Data provided to panelists to help them assess the validity and reasonableness of the standards they are recommending (e.g. median/mean cut score ratings, table agreement, etc.)

- **Impact Data (Normative Feedback)**: Data that summarize the consequences of a proposed set of cut scores (e.g., How many students’ scores will be classified at Level 3?)
Standard-Setting Method

• One of a number of approaches available for setting standards
  – Judgmental procedure
• The standard-setting approach we will be following is one of the most widely-used methods for setting standards
• Judges consider characteristics of each item and expectations of test-takers to render item-level judgments that can be aggregated into overall cut scores for the test form
• Multiple rounds of judgments and delivery of information is designed to optimize decision making
Standard-Setting Procedure

• Review and discuss achievement level descriptions (ALDs)
  – Developed prior to standard setting
  – Defines expectations of students at each level
  – Distinguishes students at adjacent levels
  – Panelists discuss characteristics that define students who are just barely at a given achievement level

• Panelists “take the test”
  – Panelists think about the test experience as if they were students
  – Consider the knowledge/skills required to answer each item correctly
Standard-Setting Procedure (Continued)

• Standard-setting procedure training
  – Provided by facilitators in breakout sessions
  – Details of the standard-setting method

• Practice Round
  – Opportunity for panelists to practice rendering judgments for 6-12 items
  – Practice entering judgments using Qwizdom data entry remotes

• Panelist Judgments
  – Panelists review each item
  – Estimate and record the percentage of students just barely at the achievement level who should correctly answer the item
  – Example: What percentage of students just barely at Achievement Level 3 should answer item 1 correctly?
The ‘Just-Barely’ Test Taker

- Borderline in terms of achievement level
- Just barely meets criteria to be classified into the achievement level
Standard-Setting Procedure (Continued)

• Recording judgments
  – Panelists provide separate judgments for each achievement level for each item (i.e., if there are four cut scores, the judge will rate each item four times)
  – Judgments recorded using electronic input device (clicker) and also recorded using paper and pencil
    ▪ Clicker facilitates fast feedback data turnarounds
    ▪ Paper/pencil recorded ratings help panelists track their judgments across achievement levels

• Multiple rounds of judgment
  – Panelists will render judgments across five rounds
  – Feedback provided between rounds
  – Feedback provided as a “reality check”
  – Discussion with peers between rounds of judgments
Standard-Setting Procedure (Continued)

- Feedback between rounds
  - Discuss items with greatest variability in judgments
    - Table discussion
    - Committee discussion
  - Review empirical item difficulty
    - Items classified into three difficulty categories (low, medium, high difficulty)
  - Impact data – percentages of students grouped into achievement levels based on recommended cut scores

- Following Round 4 (final) judgments, final recommended cuts and impact data provided
Future Steps

• Separate Reactor Panel reviews outcomes from this meeting and makes recommendations for adopting cut scores

• State Board of Education makes decision to implement final cut scores
Questions?

• About the tests?

• About the achievement levels?

• About the standard-setting activity?
Time for a break!

Please reconvene in your committee room.

Please be in your room so we can begin promptly at 2:45 PM.