Connecticut SDE

February, 04, 2010

Bureau of Student Assessment

Gilbert Andrada
State Education Consultant

Connecticut SDE

Educational Assessment in CT

- Connecticut Mastery Test
- Connecticut Academic Performance Test
- College Entrance Examinations
- National Assessment of Educational Progress
- Advance Placement Tests
- Connecticut Technical High School Tests

All of these are **summative tests**
### The Power of Numbers

- Connecticut statewide testing has been around for the past 25 years!
- Comparisons of one educational entity to another use aggregate scores from statewide testing
  - Legislation
  - Buying/Selling a house
  - Evaluating students, teachers, programs, administrators, politicians, effect of lead poisoning

### Comprehensive Assessment System

- Well-defined, well-coordinated, well-purposed
  - Interim Tests
    - Formative
    - Diagnostic
    - Benchmark
  - Summative Tests

*Designed to work together to inform educators*
Formative Assessment

- Formative assessment is anything that teachers and students do to find out what the effects of an instructional program are on learning in order to make adjustments in the teaching and learning process that will lead to improved student learning.
- For assessments to be formative, therefore, they must be administered during the course of instruction, the results must be analyzed and the findings must be used to direct further action on the part of the teacher and/or student.

Diagnostic Assessment

- Yields measures of highly related abilities underlying achievement in a subject.
- Designed to identify particular strengths and weaknesses within reasonable limits to make inferences about underlying causes

Any test can be diagnostic to some extent, but diagnostic tests serve other purposes poorly.
Benchmark Assessment

- Levels of academic performance used as checkpoints to monitor progress toward performance goals and/or academic standards.

- Connecticut Benchmark Assessment System (CBAS)

A Review of Literature

Implementing Formative Assessment at the District Level
An Annotated Bibliography

Developed by:

Maria Paz R. Avery, Ph.D.
Chad Fogelman
Steve Hamilton
Nick Hardy
Karon Labs, Ph.D.
Connecticut SDE

Guidelines Document

Connecticut’s Initiative to Support a Comprehensive Assessment System:

Guidelines for Implementing Formative Assessment at the District Level

- Barbara Q. Beaudin, Ed. D., CSDE
- Maria-Paz B. Avery, Ph.D., State Liaison NECC

<table>
<thead>
<tr>
<th>CBAS Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>To provide a resource to aid in increasing student achievement in Mathematics and Reading Comprehension by:</td>
</tr>
<tr>
<td>- Providing assessments based on the CSDE Mathematics and Language Arts Standards</td>
</tr>
<tr>
<td>- Providing assessments covering a school year using items that are keyed to Connecticut standards by CSDE Curriculum Consultants</td>
</tr>
<tr>
<td>- To develop on-line large-scale assessments</td>
</tr>
</tbody>
</table>

Bureau of Student Assessment
## What Is Being Tested?

- **Mathematics**
  - Grade Level Expectations (GLEs)
  - NOTE: These are different than what will be on the CMT which tests **Mastery Objectives**.
- **Reading Comprehension**
  - CMT Reading Comprehension strands using grade-level passages.
- More on these later

## Mastery vs Benchmark

- The CMT assesses objectives that entered into instruction a year or two ago and have been supported in instruction until they are mastered and then tested on the CMT.

- The CBAS assesses performance on objectives that would be taught in days or weeks prior to CBAS testing, *if the state pacing guides were being followed.*
Does CBAS Predict CMT?

- **NO**
  - It was not designed for this purpose
  - Interim tests should not *predict* summative tests
    - Correlations should be zero
    - Educators work to falsify a prediction line.

- **YES**
  - If, by *prediction*, you mean that students are unlikely to do tomorrow what you find they cannot do today.

Can Growth Be Measured?

- Not currently
- Each season’s test addresses different objectives as compared to other tests in the same year.
- (Winter Score) minus (Fall Score) has no interpretable meaning.
- In the future when we have sets of equivalent or parallel forms, YES.
### How Are Objectives Tested?

- **Multiple Choice Items**
  - No open-ended items at this time
- **On-line administration**
  - No paper and pencil version
- **Pilot districts reported that testing times ranged from 30 to 75 minutes with the vast majority completing testing in 40-45 minutes.**
  - Extended time is allowed
  - Tests can be paused and restarted by the proctor

### Reports

- **Student Reports** – available to teachers immediately after testing
- **School and District Data sets**
  - Rolled up every other week
  - Downloadable as an EXCEL file
- **Total Raw Scores**
- **Raw Sub-scores by Math GLE or Reading Strand**
- **No standard setting**
Partners

- CSDE
  - Bureau of Student Assessment
  - Bureau of Teaching and Learning
  - Bureau of Accountability & Improvement
- Measurement Incorporated
- Connecticut Pilot Districts

Pilot Overview

- Mathematics and Reading Comprehension
- Online testing via Measurement Incorporated Secure Test (MIST)
- Three Administrations
  - October (November)
  - February
  - May
- Over 12,000 respondents in 6 grades
### Item Analyses

- **Item quality**
  - Difficulty
  - Discrimination
  - Options Analyses
- **Test and Item statistics**
  - Form the basis of baseline performance
    - Generally speaking pilot test scores were about 67%
  - Provide indicators for edits to items

---

### Critical Feedback

- Teacher and student interface with MIST
- Logistics of large-scale computer-based testing at schools
- Reporting – emphasis on immediacy
  - Student
  - Classroom
  - School
  - District
We are currently using the winter benchmark assessments.

Current windows for testing:
- Fall (October – Late November)
- Winter (January 16 – End of February)
- Spring (May – End of school year)
- 2010-2011 – Open 24/7 with access to all 18 tests in a subject.

Operational Benchmark Assessment

To provide an interim measure of grade-level achievement.
- "Have students learned the important and testable material that the State Curriculum and Pacing Guides identified to have been taught by a particular season of a particular grade?"

Additional forms
- Irregularities during testing
- Test-retest purposes (SRBI purposes)
- PSIS-based registration will allow for a student’s scores to follow during moves.
Most reading passages were selected from retired CMT forms, while other passages were newly written by Measurement Incorporated Reading Specialists. Test items were written by CSDE Curriculum and Assessment Consultants and Measurement Incorporated Reading Specialists. Grade Level Expectations (GLE’s) were considered when writing test questions. Benchmark items are assessed and reported by the following content strands:

- **A** Forming a General Understanding
- **B** Developing Interpretation
- **D** Examining Content and Structure
### Connecticut SDE

#### Reading Comprehension

**Test Features**

- Benchmarks are available for grades 3-8
- Reading passages include a combination of text: literary, non-fiction and task (grades 5-8)
- The pilot tests comprise of 3 reading passages per grade
- The pilot tests are comprised of 14 multiple-choice questions per passage; 6 items per objective
- The reading benchmark should involve approximately 60 minutes of test taking

---

#### Sample Blueprint – Grade 3

<table>
<thead>
<tr>
<th>G</th>
<th>BA</th>
<th>A1</th>
<th>A2</th>
<th>A3</th>
<th>A5</th>
<th>B1</th>
<th>B2</th>
<th>B3</th>
<th>D1</th>
<th>D2</th>
<th>D3</th>
<th>Passage</th>
<th>Form</th>
<th>Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Alpha</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>12</td>
<td>36</td>
<td>120</td>
</tr>
<tr>
<td>2</td>
<td>Alpha</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>12</td>
<td>36</td>
<td>120</td>
</tr>
<tr>
<td>3</td>
<td>Alpha</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>12</td>
<td>36</td>
<td>120</td>
</tr>
<tr>
<td>1</td>
<td>Beta</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>14</td>
<td>42</td>
<td>120</td>
</tr>
<tr>
<td>2</td>
<td>Beta</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>14</td>
<td>42</td>
<td>120</td>
</tr>
<tr>
<td>3</td>
<td>Beta</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>14</td>
<td>42</td>
<td>120</td>
</tr>
<tr>
<td>1</td>
<td>Gamma</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>14</td>
<td>42</td>
<td>120</td>
</tr>
<tr>
<td>2</td>
<td>Gamma</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>14</td>
<td>42</td>
<td>120</td>
</tr>
<tr>
<td>3</td>
<td>Gamma</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>14</td>
<td>42</td>
<td>120</td>
</tr>
</tbody>
</table>

**Substrand Coverage**

| 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 |

**Strand Coverage**

| 48 | 36 | 36 |

**Total**

| 120 |
Reading Comprehension Pacing Guide

Grade Level Expectation (GRADE 3)  | Strand | Sub | F | W | S
--- | --- | --- | --- | --- | ---
Express opinions about texts and the reasons why (e.g., I liked ... I did not like ...). | C | 1 | 2 | 2 |
Summarize information with a beginning, middle, and end. | A | 3 | 2 | 2 |
Make text-to-self, text-to-text, and text-to-world connections. | C | 1 | 2 | 2 |
Identify rhythm, rhyme, alliteration, and assonance in poetry. | D | 1 | 2 | 2 |
Describe characters’ physical and personality traits. | A | 2 | 2 | 2 |
Draw conclusions based on implicit or explicit evidence from text. | B | 3 | 2 | 2 |
Describe the conflict faced by a character in a story. | A | 2 | 2 | 2 |
Explain similarities and differences in a story. | B | 3 | 2 | 2 |
State the main idea with supporting details in informational text. | A | 1 | 2 |
Identify why a text is preferred, using evidence to support opinion. | C | 2 | 2 |
Identify most surprising/interesting/important part of a text and explain why. | C | 2 | 2 |
Develop a new title that best fits a text. | A | 1 | 2 |
Identify and explain text structures (e.g., sequence, main idea/details, compare/contrast, cause-and-effect). | B | 1 | 2 |
Connecticut SDE

Test Format

- Model for Mathematics Curriculum
  - Grades 3 – 5 Sequenced GLEs
  - Grades 6 – 8 Pacing Guides
- 7 or 8 GLEs are assessed on each test
- 6 multiple-choice items per GLE
- Approximately 60 minutes of testing time

Bureau of Student Assessment

Calculator and Non-calculator Sessions

- Grades 5 – 8
  - Session 1 – Calculator Session
  - Session 2 – Non-Calculator Session
Connecticut SDE

Test Blueprint

<table>
<thead>
<tr>
<th>Season</th>
<th>Grade</th>
<th>Points</th>
<th>Assessed GLEs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fall</td>
<td>G3</td>
<td>48</td>
<td>1.1.1 2.1.1 2.2.19 3.3.10 3.3.11 4.1.2</td>
</tr>
<tr>
<td>Fall</td>
<td>G4</td>
<td>48</td>
<td>1.1.1 2.1.1 2.1.2 2.1.9 2.2.15 3.3.6 4.2.3 4.3.5</td>
</tr>
<tr>
<td>Fall</td>
<td>G5</td>
<td>48</td>
<td>1.1.1 2.1.1 2.1.2 2.1.2 2.2.15 3.3.8 3.3.10 4.2.4</td>
</tr>
<tr>
<td>Fall</td>
<td>G6</td>
<td>42</td>
<td>1.1.1 2.1.2 2.1.3 2.1.1 2.1.3 2.2.20</td>
</tr>
<tr>
<td>Fall</td>
<td>G7</td>
<td>48</td>
<td>1.3.7 2.1.2 2.1.3 2.1.6 2.2.7 2.2.8 2.2.9</td>
</tr>
<tr>
<td>Fall</td>
<td>G8</td>
<td>36</td>
<td>1.1.1 1.3.12 2.1.2 2.1.3 2.2.5 2.2.11</td>
</tr>
</tbody>
</table>

Bureau of Student Assessment

Pre K – 8 Curriculum Standards

GRADE 3

Algebraic Reasoning: Patterns and Functions
Patterns and functions relationships can be represented and analyzed using a variety of strategies, tools, and technologies.

State Framework | Grade-Level Expectations | CMT Correlations |
---|---|---|
1. Descriptive understanding of laws of exponents and exponential notation and rules of exponents.
2. Identify and demonstrate an understanding of equations using the methods.
3. Demonstrate an understanding of patterns and functions relationships using the methods.
4. Demonstrate an understanding of patterns and functions relationships using the methods.
5. Demonstrate an understanding of patterns and functions relationships using the methods.
6. Demonstrate an understanding of patterns and functions relationships using the methods.
7. Demonstrate an understanding of patterns and functions relationships using the methods.
8. Demonstrate an understanding of patterns and functions relationships using the methods.
9. Demonstrate an understanding of patterns and functions relationships using the methods.
10. Demonstrate an understanding of patterns and functions relationships using the methods.

Bureau of Student Assessment
Continued Development

- **Mathematics**
  - Additional sets of benchmark assessments for current GLEs
  - Further coverage of more GLEs is planned

- **Reading**
  - Utilization of current reading passages to develop additional forms

- **Open-Ended Items**
  - Development of items and scoring materials
Future Directions

- Research into automated content scorers
  - may allow us to include open-ended items with the same turn-around time for results
  - can give us the ability to offer assessments that cover the curriculum more-fully

- Item bank development
  - will eventually allow us to offer teacher-customized assessments that focus on narrower sets of content.

Future Directions (continued)

- Automated essay scoring already exists so we are exploring the inclusion of a machine scored direct assessment of writing component.

- Current department research projects in to computer-based tests accommodations contribute to the conceptualization large-scale computer-based testing for some future generation of statewide testing.
### Research: Classroom Teachers

- How do Benchmark Assessments help to
  - Inform instructional change
  - Individualize instruction
  - Serve special populations
- How responsive are CBAS scores to the effects of instruction?

### Research: Curriculum Specialists

- Can CBAS items be cross-classified to provide information relative to
  - International Standards?
  - National Standards?
  - Regional Standards?
- Standard setting
**Connecticut SDE**

**Research: Measurement/ Evaluation**

- What equating designs should CBAS employ?
  - Horizontal – for parallel forms
  - Vertical – for tests 3 – 8
- Can item parameters be used to develop GLE or strand level cut scores to make scores more interpretable?
- What are the critical elements of the item bank that will allow for adaptive testing?

---

**Connecticut SDE**

**Research: Assessment**

- What can we know about Learning Progressions using CBAS?
- To what extent and in what manner can we use automated scoring for
  - Essays
  - Short answer text and short answer math items
Research: Education Technology

- What are the properties of new, computer-based item types?
- What types of on-line tools need to be developed to allow for computerized benchmark testing?
- What accommodations can be given using computers?
- What are the critical specifications for technology/test development?

Additional Information

- Gilbert Andrada
  - Gilbert.Andrada@ct.gov
  - (860) 713-6883
  - CBAS@ct.gov