Empirical Validation of Benchmark Assessments Webinar

Bill Conrad and Dan Mason
Assessment and Accountability Department
408-453-4332

bill_conrad@sccoe.org
Dan_mason@sccoe.org
The Moral Purpose

To support every student at our district/school as if he/she were our own child.
Webinar Logistics

1. Please put your phone on mute until you want to speak.
2. You can use the chat or question function to ask questions during the session.
3. We will use the polling function today. When a poll comes up, just click your answer!
4. Remember, no emailing during the webinar! The system is watching! :>)
5. Have fun!
What Problems Are We Trying to Fix?

School districts do a good job of building and using benchmark assessments to regularly measure how well students achieve key learning standards throughout the year.

School districts do not have good systems to let them know in an empirical way how well these assessments actually achieved the intended outcomes for the benchmark assessments.
Webinar Outcomes

- Understand the key elements for empirically validating benchmark assessments after they have been administered.
  - Review the role of Item Analysis for Formative Purposes
  - Understand and check the theory of action of the assessment
  - Review the classical statistics for items
  - Using a Wright Map to understand the quality of the assessments
- Understand the Role of Leadership
- Plan a District Assessment Review Process
Our Students are in Need of a More Balanced Assessment System

- Minute-to-Minute and Hour-to-Hour Classroom Assessments
- Teacher developed assessments for learning
- Questioning Strategies
- Whiteboard activities
- Teacher Team developed common assessments
- Item Analysis focused on Distractor selection
- Weekly Quizzes
- Student Journals
- Daily Exit Tickets
- Math Benchmark Assessments
- Grade Level Team developed End of Unit Assessments
- NWEA Math Assessments
- Teacher Team developed common assessments
- California Standards Test (CST)
- California High School Exit Exam (CAHSEE)
- Report Cards
- AP Math Exams
Active Listening

- Teaching the dog to whistle
Active Listening

Learning, rather than teaching, becomes the central issue.
Active Listening

- Solving the Problem \(20 - 9 = ?\)

I would use my fingers.
I’d start with 20 and count backwards with my fingers. Then when I got to 9 I’d see how many finger I used. 11
Active Listening

- Solving the Problem $20 - 9 = ?$

I would make 20 tally marks then mark 9 off. Then I would count the remainder of 11.
Active Listening

- Solving the Problem 20 - 9 = ?

\[ I \text{ would write } 20 \text{ then countdown starting with } 20 \text{ then countdown by 1 until I took away } q. \]
Active Listening

- Solving the Problem 20 - 9 = ?

I would solve the simpler problem of 20 - 10 to get 10 and then just add 1 to solve the problem 20 - 9 and get 11.
Active Listening

- Solving the Problem $20 - 9 = ?$

I would set the problem up and borrow from the tens place and do $10 - 9$ and then add ten to it. This gives the correct answer of $11$. 
Active Listening

- Solving the Problem 20 - 9 = ?

I would add nine to twenty and add nine to nine then I would subtract twenty-nine and eighteen, so I don’t have to borrow from the tens.
Active Listening

What role does active listening play in an effective formative assessment process?
Math Reasoning Inventory

Find out what your students really understand about math

- Focus on how students think and reason
- Uncover students' strategies, understandings, and misconceptions
- Learn how students respond to questions the Common Core expects all middle school students to answer successfully

MRI is FREE for all teachers!

https://mathreasoninginventory.com/
By listening to students, we can identify error patterns in student thinking and then build assessment items that incorporate distractors that attract student responses where students may have these error patterns in thinking.
15. “A function has x-intercept 3 and y-intercept 2. Which of the functions below could be this function?

A 4 + 3x = 2y
B 2x - 3y = -6
C 2y + 3x = 4
D 3y - 6 = -2x
Item Analysis Example

Item 15 aligns with a component within Algebra I Standard 6.0, in which “students graph a linear equation and compute the x-and y-intercepts (e.g., $2x + 6y = 4$). They are also able to sketch the region defined by linear inequality (e.g., they sketch the region defined by $2x + 6y < 4$).”

Source: Three Facets of Formative Assessment: How to Revolutionize (and actually use locally developed tests. Dan Mason et al. Santa Clara County Office of Education
## Answer Frequency
*sorted by % Correct*

![Answer Frequency Table](Image)

<table>
<thead>
<tr>
<th>Item No</th>
<th>Strand/Standard Aligned</th>
<th>Correct Response</th>
<th>Percent Correct</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>No Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>35</td>
<td>GR 08 Algebra I 15.0</td>
<td>D</td>
<td>11%</td>
<td>15%</td>
<td>57%</td>
<td>13%</td>
<td>11%</td>
<td>4%</td>
</tr>
<tr>
<td>32</td>
<td>GR 08 Algebra I 22.0</td>
<td>D</td>
<td>14%</td>
<td>26%</td>
<td>24%</td>
<td>31%</td>
<td>14%</td>
<td>5%</td>
</tr>
<tr>
<td>39</td>
<td>GR 08 Algebra II 22.0</td>
<td>B</td>
<td>14%</td>
<td>27%</td>
<td>14%</td>
<td>44%</td>
<td>9%</td>
<td>7%</td>
</tr>
<tr>
<td>14</td>
<td>GR 08 Algebra I 22.0</td>
<td>B</td>
<td>15%</td>
<td>47%</td>
<td>15%</td>
<td>20%</td>
<td>15%</td>
<td>3%</td>
</tr>
<tr>
<td>15</td>
<td>GR 08 Algebra I 6.0</td>
<td>D</td>
<td>16%</td>
<td>24%</td>
<td>19%</td>
<td>39%</td>
<td>16%</td>
<td>3%</td>
</tr>
</tbody>
</table>

Item Analysis Example

It is evident from the data that students who chose distractor “C: 2y + 3x = 4” may have a misconception that may be related to their understanding of variables, in general, since they do not recognize x- and y- intercepts as points on a coordinate plane. These students may not know that they can substitute in values for variables, x and y. If they understood this concept, they may have computed the y-intercept by substituting 0 for x and computed 2 for y, and then substituted 0 for y and computed 4/3 for x. (Slide 22). Instead of finding the equation that satisfied the true points (0,2) and (3,0), the students who chose distractor “C” simply place the 2 next to the y in the given equation as the y-intercept and 3 next to the x in the given equation as the x-intercept.
## Item Analysis Example

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>y-intercept is 2</td>
<td>4 + 3x = 2y</td>
<td>2x - 3y = -6</td>
<td>2y + 3x = 4</td>
<td>3y - 6 = -2x</td>
</tr>
<tr>
<td>X=0</td>
<td>4+3(0)=2y (0, 3/2)</td>
<td>2(0) - 3y = -6 6</td>
<td>2y + 3(0) = 4 4</td>
<td>3y - 6 = -2(0) (0,2)</td>
</tr>
<tr>
<td>x-intercept is 3</td>
<td>4 + 3x = 0 (-4/3, 0)</td>
<td>2x - 3(0) = -6 6</td>
<td>2(0) + 3x = 4 4</td>
<td>3(0) - 6 = -2x (3,0)</td>
</tr>
</tbody>
</table>

Source: Three Facets of Formative Assessment: How to Revolutionize (and actually use) locally developed tests. Dan Mason et al. Santa Clara County Office of Education
Review the 7th grade Math Benchmark Answer Frequency document. Do you find any assessment items where students selected incorrect distractors at a percentage rate greater than 40%?

Assessment Items with incorrect distractor selection greater than 40%

Describe how the selected distractor may identify a student misconception, error pattern, or problem with the quality of the item.

What are your recommendations to improve student learning or to improve the item?
Webinar Outcomes

- Understand the key elements for empirically validating benchmark assessments after they have been administered.
  - Review the role of Item Analysis for Formative Purposes
  - Understand and check the theory of action of the assessment
  - Review the classical statistics for items
  - Using a Wright Map to understand the quality of the assessments
- Understand the Role of Leadership
- Plan a District Assessment Review Process
Benchmark Assessment Possible
Theories of Action

1. Items aligned to standards and built on Bloom’s taxonomy starting with items based on knowledge, comprehension, and moving to items based on analysis.
2. Student recognition of key math ideas moving to a conceptual understanding and ultimately leading to application of the concepts to the solution of problems.
3. Include items that involve single step, two step and multiple step solutions.
4. Some items are aligned to more challenging standards than others.
Benchmark Assessment Possible

Theories of Action

After reviewing the 7th grade math assessment items, does there appear to be a theory of action that underlies the overall assessment? Please describe the theory of action below.
Benchmark Assessment Possible Theories of Action

Possible findings of a Theory of Action on the 7th Grade Benchmark Assessment

There are several potential theories of action working within this assessment. The assessment is aligned to three math areas including Algebra and Functions, Measurement and Geometry, and Number Sense. A prerequisite for students to do well in Algebra is built upon their Number Sense abilities. Thus it would be expected that assessment items aligned to Number Sense would be less challenging to students than items aligned to Algebra I.

The assessment includes items that require students to use to solve purely mathematical problems (problems 1-8 are examples) and there are also problems that require students to apply their knowledge of math skills to solve problems. (Problems # 11 and #12). It would be expected that application problems would be more challenging for students than pure math problems.
Benchmark Assessment Possible
Theories of Action

If there is no explicit theory of action that underlies the assessment, review the items and identify those items that may be more challenging for students to solve based on Bloom’s taxonomy or other criteria for complexity that you identify.
Benchmark Assessment Possible Theories of Action

Some Potential Findings from the 7th Grade Math Benchmark Assessment:

• Some of the problems on the test require students to use multiple steps to solve them. (Problems 11, 12, 22).

• The test includes items aligned to Number Senses and Algebra I. It would be expected that Algebra I items would be more challenging than items aligned to the topic of Number Sense.
Webinar Outcomes

- Understand the key elements for empirically validating benchmark assessments after they have been administered.
  - Review the role of Item Analysis for Formative Purposes
  - Understand and check the theory of action of the assessment
  - Review the classical statistics for items
  - Using a Wright Map to understand the quality of the assessments
- Understand the Role of Leadership
- Plan a District Assessment Review Process
The Point Biserial Correlation statistic differentiates items based on whether they classify students in the same way as the test as a whole. Students who perform well on the overall test should also perform well on the item (have a higher probability of success). Students who do not perform well on the overall test should have a lower probability of success on the item. A strong positive point biserial correlation (> 0.25) means that this expectation is met. A low or negative point biserial correlation (< 0.15) means that this expectation was not met, that high-performing students are getting the item wrong while low-performing students are getting it right. In this case, the item should be reviewed for quality.
Point Biserial Analysis

Identify items that have a very low or negative point biserial and review them for quality and then record your findings about these items.
### Point Biserial Analysis

**Source:** SCoolPlan Data Decision Support System; [http://www.sccoe.org/schoolplan/](http://www.sccoe.org/schoolplan/)

#### Answer Frequency

**Total Number of Students:** 760  
**Total Students in Calculation:** 760

<table>
<thead>
<tr>
<th>Item No.</th>
<th>Item Summary</th>
<th>Strand/Standards Aligned</th>
<th>Correct Response</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>NR</th>
<th>MR</th>
<th>Point Biserial</th>
</tr>
</thead>
<tbody>
<tr>
<td>17</td>
<td></td>
<td>Mathematical Reasoning 2...</td>
<td>E</td>
<td>386 51%</td>
<td>313 41%</td>
<td>47 6%</td>
<td>0 0%</td>
<td>11 1%</td>
<td>3 0%</td>
<td>41% 0.13</td>
</tr>
<tr>
<td>20</td>
<td></td>
<td>Mathematical Reasoning 2...</td>
<td>D</td>
<td>69 9%</td>
<td>180 24%</td>
<td>88 12%</td>
<td>409 54%</td>
<td>11 1%</td>
<td>3 0%</td>
<td>54% 0.24</td>
</tr>
<tr>
<td>11</td>
<td></td>
<td>Algebra and Functions 1.3</td>
<td>A</td>
<td>708 93%</td>
<td>19 3%</td>
<td>17 2%</td>
<td>13 2%</td>
<td>2 0%</td>
<td>1 0%</td>
<td>93% 0.32</td>
</tr>
<tr>
<td>16</td>
<td></td>
<td>Algebra and Functions 1.1</td>
<td>B</td>
<td>112 15%</td>
<td>471 62%</td>
<td>67 9%</td>
<td>101 13%</td>
<td>6 1%</td>
<td>3 0%</td>
<td>62% 0.33</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>Algebra and Functions 1.1</td>
<td>A</td>
<td>387 51%</td>
<td>85 11%</td>
<td>168 22%</td>
<td>114 15%</td>
<td>6 1%</td>
<td>0 0%</td>
<td>51% 0.35</td>
</tr>
<tr>
<td>21</td>
<td></td>
<td>Number Sense 1.1</td>
<td>A</td>
<td>707 93%</td>
<td>19 3%</td>
<td>21 3%</td>
<td>8 1%</td>
<td>5 1%</td>
<td>0 0%</td>
<td>93% 0.36</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>Algebra and Functions 1.1</td>
<td>B</td>
<td>15 2%</td>
<td>670 88%</td>
<td>60 8%</td>
<td>11 1%</td>
<td>3 0%</td>
<td>1 0%</td>
<td>88% 0.38</td>
</tr>
</tbody>
</table>
Finding:

None of the items demonstrated a negative point biserial value. However, item 17 demonstrated a 0.13 point biserial value for all 7th graders who took the test.
Level of Difficulty

The Scale Score Difficulty statistic reflects the probability of getting an item wrong converted to a CST scale score metric. Items with high scale scores are very difficult; there is a high probability of getting them wrong. Items with low scale scores are easy, with a low probability of getting them wrong. Items that have very high difficulty scores or very low difficulty scores should be checked to determine if it is the quality of the item that is making them difficult or the theory of action that drives the assessment.
Level of Difficulty

Identify items that have high difficulty rankings or very low difficulty rankings and record your findings about these items below.
Level of Difficulty

Answer Frequency Report for 7th Gr Math 2nd Quart EDS 10-11 Math

Total Number of Students: 760
Total Students in Calculation: 760

| Item No. | Item Summary | Strand/Standards Aligned | Correct Response | Total | A | # | % | B | # | % | C | # | % | D | # | % | NR | # | % | MR | # | % | % | Point Biserial | Scale Score Difficulty |
| 17      | Mathematical Reasoning 2... | B | 366 | 51% | 313 | 41% | 47 | 6% | 0 | 0% | 11 | 1% | 3 | 0% | 41% | 0.13 | 402 |
| 10      | Algebra and Functions 1.1 | A | 367 | 51% | 85 | 11% | 168 | 22% | 114 | 15% | 6 | 1% | 0 | 0% | 51% | 0.35 | 337 |
| 22      | Algebra and Functions 1.1 | D | 93 | 12% | 104 | 14% | 86 | 11% | 464 | 61% | 8 | 1% | 5 | 1% | 61% | 0.5 | 322 |
| 20      | Mathematical Reasoning 2... | D | 69 | 9% | 180 | 24% | 88 | 12% | 409 | 54% | 11 | 1% | 3 | 0% | 54% | 0.24 | 319 |
| 15      | Algebra and Functions 4.1 | C | 96 | 13% | 74 | 10% | 464 | 61% | 119 | 16% | 5 | 1% | 2 | 0% | 61% | 0.48 | 319 |
| 14      | Algebra and Functions 1.1 | D | 134 | 18% | 83 | 11% | 83 | 11% | 455 | 60% | 3 | 0% | 2 | 0% | 60% | 0.41 | 316 |
| 18      | Algebra and Functions 1.2 | A | 486 | 64% | 109 | 14% | 115 | 15% | 41 | 5% | 6 | 1% | 3 | 0% | 64% | 0.46 | 306 |

Level of Difficulty

Findings:

Item 17 demonstrated a level of item difficulty of 410 which is a very high scale score. The next highest scale score level of difficulty was item 10 with a scale score difficulty level of 337.
Item Reliability

The Item Reliability statistic ranges from 0.0 to 1.0, where 1.0 means "perfectly reliable". It is a measure of how well the item is able to reveal differences between high performing students and low performing students. Two quantities go into item reliability -- a) the spread of the students on the item (as measured by their differing probabilities of success on the item); and b) the average margin of error around each student's probability of success. When the students are well-spread out and the average margin of error is very low, the item reliability approaches 1.0. When the students are clumped together in the middle of the scale, or when the average margin of error is very high, the item reliability approaches 0.0. When items are given an item reliability of 0.0 or near 0.0, it means that the average margin of error is the same as or greater than the spread of the students along the scale. In other words, viewed through the lens of that item, the students are one big blur and it would be difficult to attribute their performance on the item to the learning target to which it was aligned.
Item Reliability

Identify items that demonstrate reliabilities at 0.0 or near 0.0 and record your findings about these items.
# Item Reliability

## Total Number of Students: 760
## Total Students in Calculation: 760

<table>
<thead>
<tr>
<th>Item No.</th>
<th>Item Summary</th>
<th>Strand/Standards Aligned</th>
<th>Correct Response</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>NR</th>
<th>MR</th>
<th>%</th>
<th>Point Biserial</th>
<th>Scale Score Difficulty</th>
<th>Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>17</td>
<td>Mathematical Reasoning 2</td>
<td>B</td>
<td></td>
<td>386</td>
<td>51%</td>
<td>313</td>
<td>41%</td>
<td>47</td>
<td>6%</td>
<td>0</td>
<td>0%</td>
<td>11</td>
<td>1%</td>
</tr>
<tr>
<td>20</td>
<td>Mathematical Reasoning 2</td>
<td>D</td>
<td></td>
<td>69</td>
<td>9%</td>
<td>180</td>
<td>24%</td>
<td>88</td>
<td>12%</td>
<td>409</td>
<td>54%</td>
<td>11</td>
<td>1%</td>
</tr>
<tr>
<td>16</td>
<td>Algebra and Functions 1</td>
<td>B</td>
<td></td>
<td>112</td>
<td>15%</td>
<td>471</td>
<td>52%</td>
<td>67</td>
<td>9%</td>
<td>101</td>
<td>13%</td>
<td>6</td>
<td>1%</td>
</tr>
<tr>
<td>11</td>
<td>Algebra and Functions 1</td>
<td>A</td>
<td></td>
<td>708</td>
<td>93%</td>
<td>19</td>
<td>3%</td>
<td>17</td>
<td>2%</td>
<td>13</td>
<td>2%</td>
<td>2</td>
<td>0%</td>
</tr>
<tr>
<td>21</td>
<td>Number Sense 1</td>
<td>A</td>
<td></td>
<td>707</td>
<td>93%</td>
<td>19</td>
<td>3%</td>
<td>21</td>
<td>3%</td>
<td>8</td>
<td>1%</td>
<td>5</td>
<td>1%</td>
</tr>
<tr>
<td>10</td>
<td>Algebra and Functions 1</td>
<td>A</td>
<td></td>
<td>387</td>
<td>51%</td>
<td>85</td>
<td>11%</td>
<td>168</td>
<td>22%</td>
<td>114</td>
<td>15%</td>
<td>6</td>
<td>1%</td>
</tr>
<tr>
<td>19</td>
<td>Mathematical Reasoning 2</td>
<td>D</td>
<td></td>
<td>46</td>
<td>6%</td>
<td>139</td>
<td>18%</td>
<td>61</td>
<td>8%</td>
<td>506</td>
<td>67%</td>
<td>7</td>
<td>1%</td>
</tr>
<tr>
<td>3</td>
<td>Algebra and Functions 1</td>
<td>B</td>
<td></td>
<td>15</td>
<td>2%</td>
<td>670</td>
<td>88%</td>
<td>60</td>
<td>8%</td>
<td>11</td>
<td>1%</td>
<td>3</td>
<td>0%</td>
</tr>
<tr>
<td>4</td>
<td>Algebra and Functions 1</td>
<td>D</td>
<td></td>
<td>18</td>
<td>2%</td>
<td>5</td>
<td>1%</td>
<td>23</td>
<td>3%</td>
<td>713</td>
<td>94%</td>
<td>1</td>
<td>0%</td>
</tr>
</tbody>
</table>

Findings:

Item 17 demonstrated an item reliability of 0.0.

The next lowest level of reliability was item 20 with a reliability value of 0.33.
Possible Recommendations for Item Improvement based on Classical Statistics Analyses

Investigate both the quality of item 17 in terms of the level of difficulty of numbers that it presents to students. It may be appropriate to include smaller single and double digit numbers for students to evaluate as prime or composite.
Webinar Outcomes

- Understand the key elements for empirically validating benchmark assessments after they have been administered.
  - Review the role of Item Analysis for Formative Purposes
  - Understand and check the theory of action of the assessment
  - Review the classical statistics for items
  - Using a Wright Map to understand the quality of the assessments
- Understand the Role of Leadership
- Plan a District Assessment Review Process
Wright Maps

Your packet contains a Wright Map of the 2nd Quarter 7th grade Math assessment items. The Wright Map places the items on the same CST scale as the students. There is an expectation that the spread of students should more or less match the spread of items on the Map. There should be items within performance band that also align with students at that band. Please use the slides that follow to learn how to use the Wright Map to analyze assessment items. You can also find detailed descriptions of how to use Wright Maps to analyze assessment items in the book called *Three Facets of Formative Assessments*. 
Analyzing the Performance of a Benchmark Assessment – Theory of Action

Understanding the level of complexity of the content that will form the basis for a Benchmark assessment will help in the analysis of how items aligned to the levels perform on the test.

For example, if solving simpler linear equations is more complex than solving Rational Radical Polynomials than items aligned to these levels should perform at different levels of difficulty on the test.

<table>
<thead>
<tr>
<th>Level of Complexity</th>
<th>What the Student Knows</th>
<th>Response to items (repeats at every level)</th>
</tr>
</thead>
</table>
| 6 - Trigonometric Polar Parametric | Student understands trigonometric, polar and parametric functions | Responses indicate that a student can:  
- generalize this type of functions with a rule,  
- recognize/create/describe patterns from this type of function,  
- create and extend patterns from this function with a rule, |
| 5 - Exponential Logarithmic Recursive | Student understands exponential, logarithmic and recursive functions |                                                                                                          |
| 4 - Rational Radical Polynomial | Student understands rational, radical and polynomial functions |                                                                                                          |
| 3 - Absolute Value Piecewise Quadratic | Student understands absolute value, piecewise and quadratic functions |                                                                                                          |
| 2 - Multi-step Linear Inequalities | Student understands multi-step linear functions and inequalities. |                                                                                                          |
| 1 - Simple Linear | Student understands simple linear functions |                                                                                                          |

Figure 12: “Complexity of Functions” progress map based on the Springboard’s CBSCS

Source: Three Facets of Formative Assessment: How to Revolutionize (and actually use) locally developed tests. Dan Mason et al. Santa Clara County Office of Education
Analyzing the Performance of a Benchmark Assessment

Source: Three Facets of Formative Assessment: How to Revolutionize (and actually use) locally developed tests. Dan Mason et al. Santa Clara County Office of Education
Analyzing the Performance of a Benchmark Assessment

Even without a theory of action for a benchmark assessment, the Wright Map can be useful in evaluating the performance of items aligned to student performance. Students in any given row have a 50% probability of getting the item in that row correct and higher probability for items below the row.

Reviewing the distribution of items can tell the test developers if there is a balanced range of difficulty of items on the test.

Source: Three Facets of Formative Assessment: How to Revolutionize (and actually use) locally developed tests. Dan Mason et al. Santa Clara County Office of Education
The Wright Map can also be used to evaluate the ability of a Placement test to effectively identify students who are prepared to take the next level of mathematics.

This district uses 80% correct on a placement test to identify students ready for the next level of math but there are few items at this cut point on the Wright Map.
# Wright Map


**Student Roster:** SES 11-12 Dec  
**Assessment:** CSP-Equivalent Scale Scores and Performance Bands

<table>
<thead>
<tr>
<th>Course</th>
<th>Subject</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>Math</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Grade</th>
<th>Grade Level Tested</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>All</td>
</tr>
</tbody>
</table>

| My Group     | All Students       |

## Grade 7

### Overall Item Distribution

#### Levels

<table>
<thead>
<tr>
<th>Levels</th>
<th>Scale Score</th>
<th>Distribution of students (n=760)</th>
<th>Algebra and Functions</th>
<th>Mathematical Reasoning</th>
<th>Measurement and Geometry</th>
<th>Number Sense</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advanced</td>
<td>801-999</td>
<td>XXXXXXXXXXXX</td>
<td>1.1 1.2 1.3 4.1</td>
<td>1.1 2.5 2.9 2.8</td>
<td>1.1 1.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>610-800</td>
<td>X XXXXXXXXXXXX</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>41-609</td>
<td>XXXXXXXXXXXX</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proficient</td>
<td>330-342</td>
<td>XXXXXX</td>
<td>17</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>320-329</td>
<td>X XXXXXXXX</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>306-319</td>
<td>XXXXXXXXXXXX</td>
<td>16 14 22 18</td>
<td>15 13</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Basic</td>
<td>333-349</td>
<td>XXXXXXXXXXXX</td>
<td>16 14 22 18</td>
<td>15 13</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td></td>
<td>317-322</td>
<td>X XXXXXXXX</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>300-316</td>
<td>XXXXXXXXXXXX</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Below Basic</td>
<td>296-299</td>
<td>XXXXXX</td>
<td>16</td>
<td>10 12</td>
<td>12</td>
<td>25 23 24</td>
</tr>
<tr>
<td></td>
<td>272-285</td>
<td>X XXXXXXX</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>257-271</td>
<td>XXXXXXX</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Far Below</td>
<td>221-256</td>
<td>XXXXXX</td>
<td>3,1</td>
<td>2,6 11</td>
<td>6,7</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>196-220</td>
<td>XXXXX</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>166-185</td>
<td>XXX</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Each X represents <= 10 student(s)

----- Average Scale Score
Wright Map

Based on your review of the Wright Maps for the 2nd Quarter 7th grade Math Benchmark Assessment, does the Wright Map corroborate or not corroborate hypotheses that you previously made about item complexity or quality?
Wright Map Findings

Most of the items on the Wright Map do not match the distribution of student performance on the assessment indicating that many of the items are not challenging enough for students.

Some of the number sense items such as 17 and 20 score at higher levels of difficulty than would be predicted based upon the theory of action. Item 17 which is categorized as a Number Sense Item scores in the proficient range higher than any other items on the test.
Wright Map Sample Recommendations

Improve the rigor of items that require application of math content and skills. Possibly consider that they all involve multi-step solutions.

Review the quality of item 17 and ensure that there are more less challenging opportunities for students to identify prime and composite numbers.
Webinar Outcomes

- Understand the key elements for empirically validating benchmark assessments after they have been administered.
  - Review the role of Item Analysis for Formative Purposes
  - Understand and check the theory of action of the assessment
  - Review the classical statistics for items
  - Using a Wright Map to understand the quality of the assessments
- Understand the Role of Leadership
- Plan a District Assessment Review Process
What is the Role of Leadership?

• Support a vision for Assessment that includes the regular validity review of Benchmark Assessment
• Build the structural supports necessary to conduct an annual Benchmark Assessment Review in your District
• Ensure that there Board policy supports the regular review of Benchmark assessments
• Ensure that you have the human resource to carry out the Benchmark Assessment Review
• Build the political support for the work
• Establish symbolic artifacts and processes that celebrate the success of the Benchmark Assessment Review process
Webinar Outcomes

- Understand the key elements for empirically validating benchmark assessments after they have been administered.
  - Review the role of Item Analysis for Formative Purposes
  - Understand and check the theory of action of the assessment
  - Review the classical statistics for items
  - Using a Wright Map to understand the quality of the assessments
- Understand the Role of Leadership
- Plan a District Assessment Review Process
How can I plan a Benchmark Assessment Review Process in My District?

- If you are interested in planning and conducting a Benchmark Assessment Review Process in your school district, you can contact Bill Conrad or Dan Mason for the resources, tools, and protocols to conduct the review.

Bill Conrad
408-453-4332 (Office)
Bill_conrad@scccoe.org

Dan Mason
408-453-4346
Dan_mason@scccoe.org
Webinar Evaluation

Please go to the site below to complete an evaluation of this webinar. Thank you.

http://www.surveymonkey.com/s/7XKZVRZ
Empirical Validation of Benchmark Assessments Webinar

Bill Conrad and Dan Mason
Assessment and Accountability Department
408-453-4332

bill_conrad@sccoe.org
Dan_mason@sccoe.org