



Santa Clara County  
Office of Education

# Empirical Validation of Benchmark Assessments Webinar

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

[bill\\_conrad@sccoe.org](mailto: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.



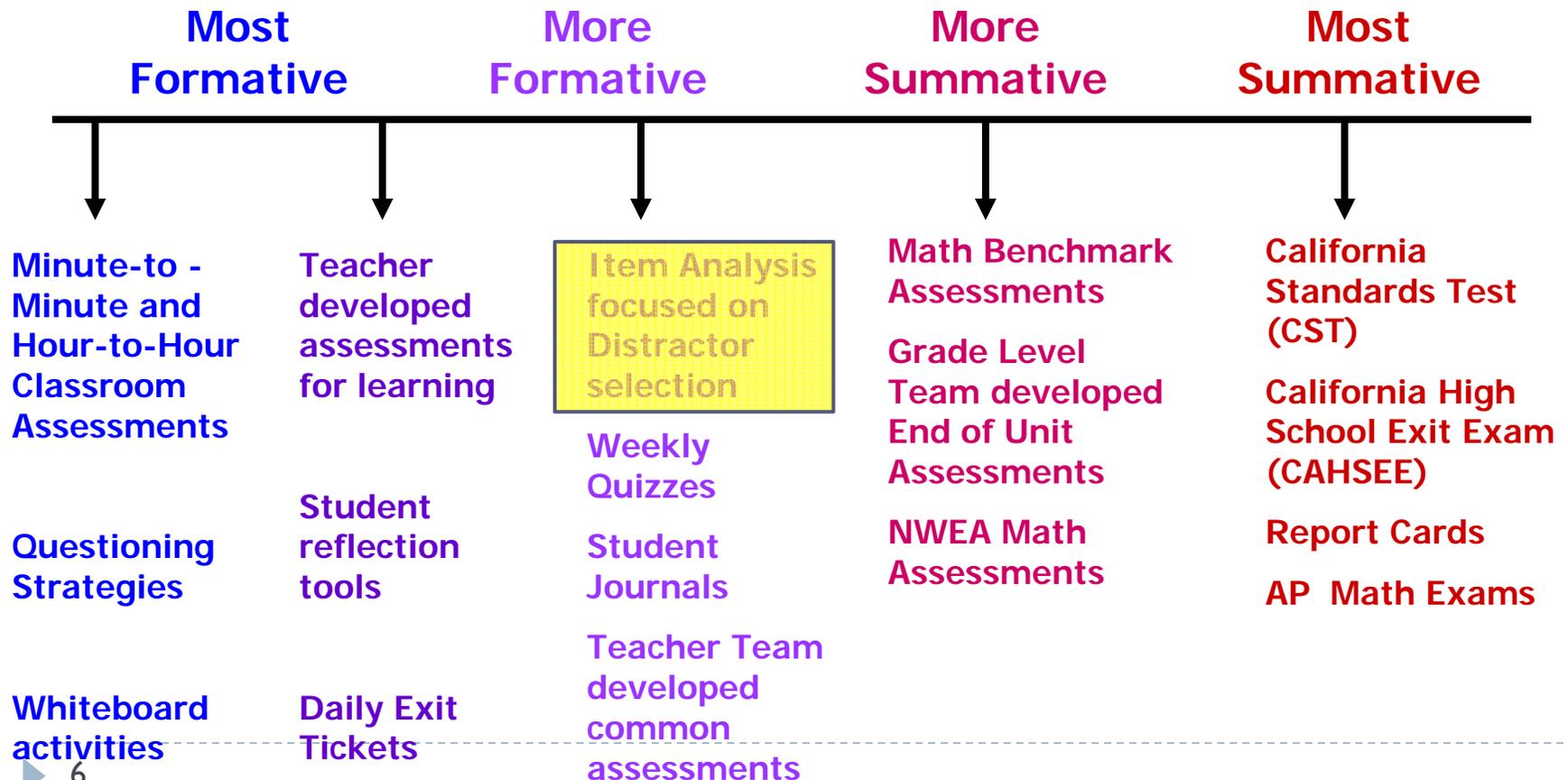
# Webinar Outcomes

---

- ▶ 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



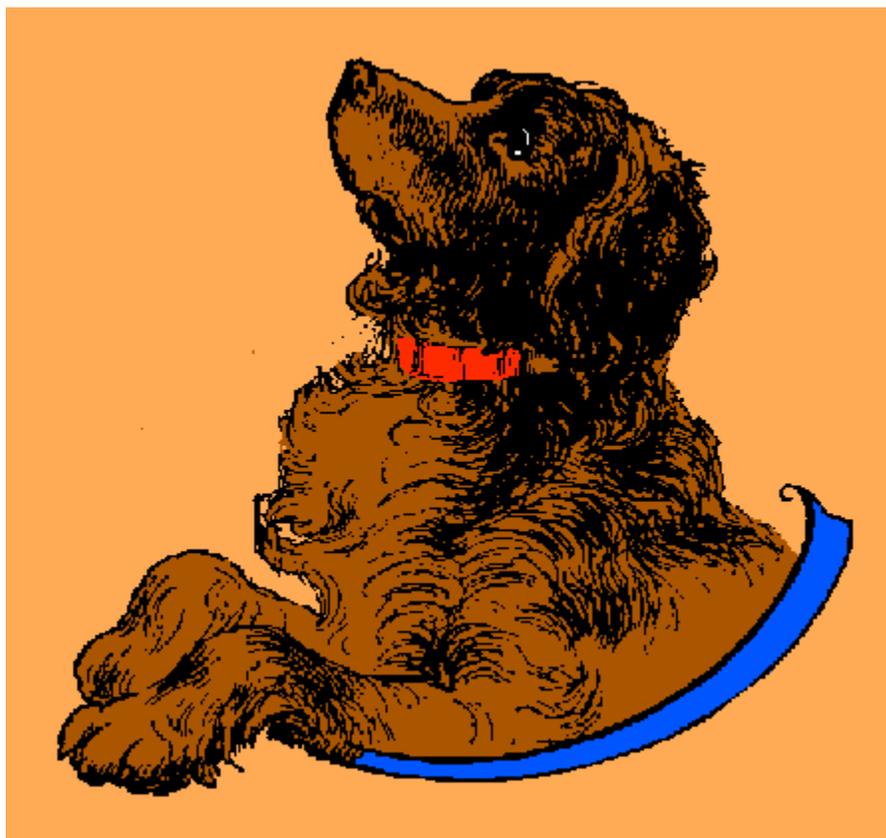
# Our Students are in Need of a More Balanced Assessment System



# Active Listening

---

- ▶ Teaching the dog to whistle



# Active Listening

---

**Learning,  
rather than teaching,  
becomes the central issue.**

# Active Listening

---

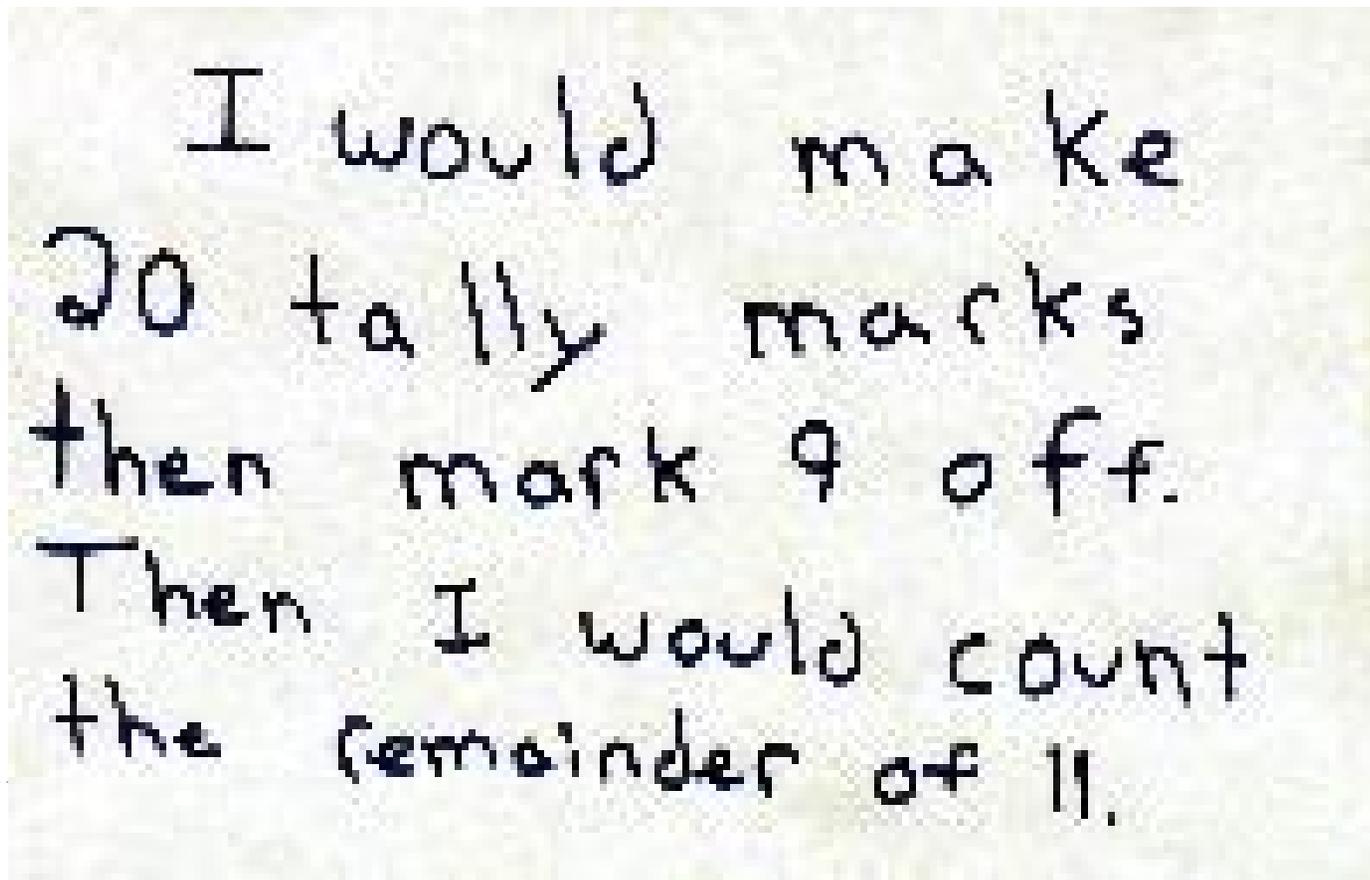
- ▶ Solving the Problem  $20 - 9 = ?$

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  $20 - 9 = ?$

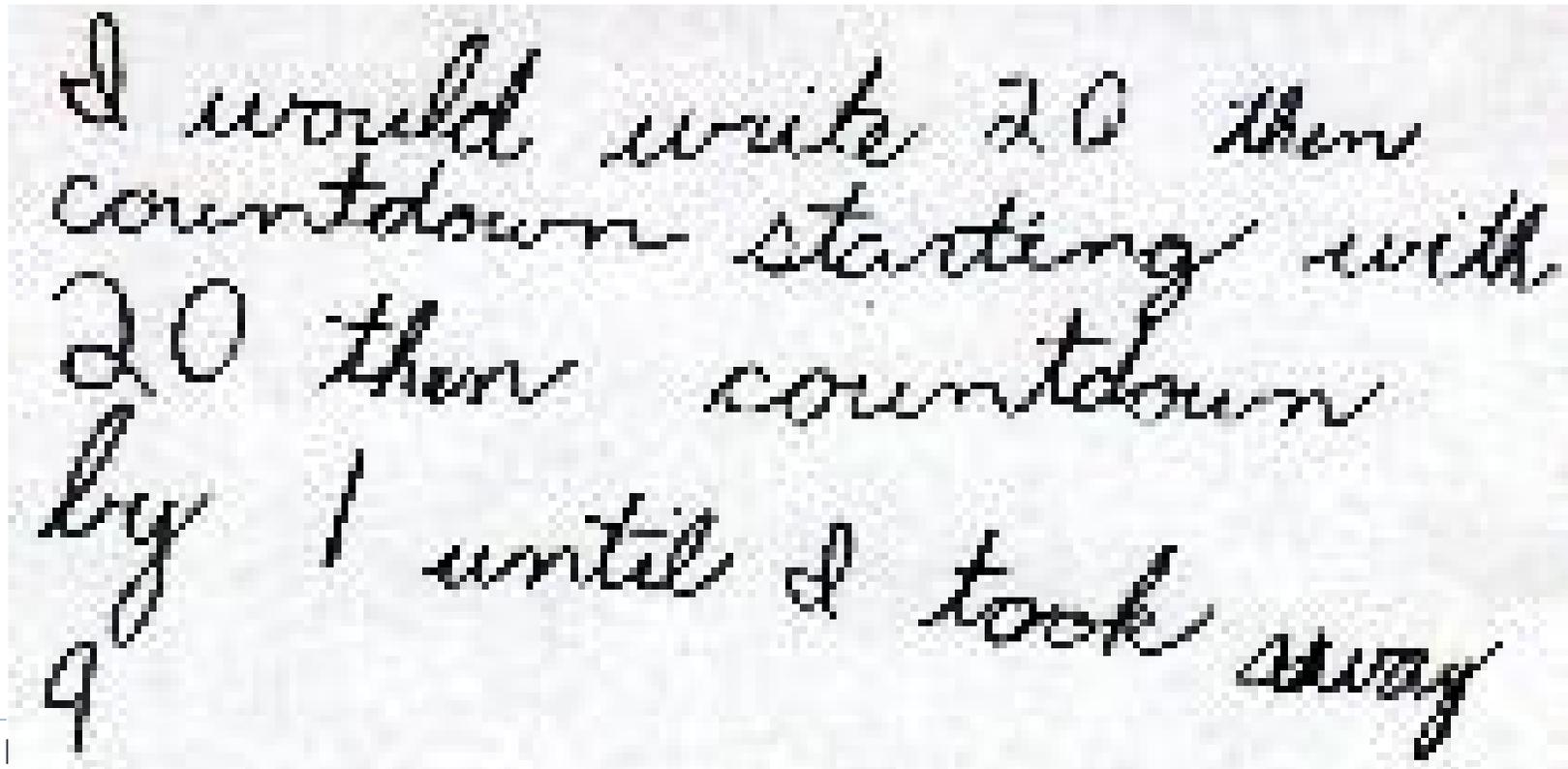


I would make  
20 tally marks  
then mark 9 off.  
Then I would count  
the remainder of 11.

# Active Listening

---

- ▶ Solving the Problem  $20 - 9 = ?$

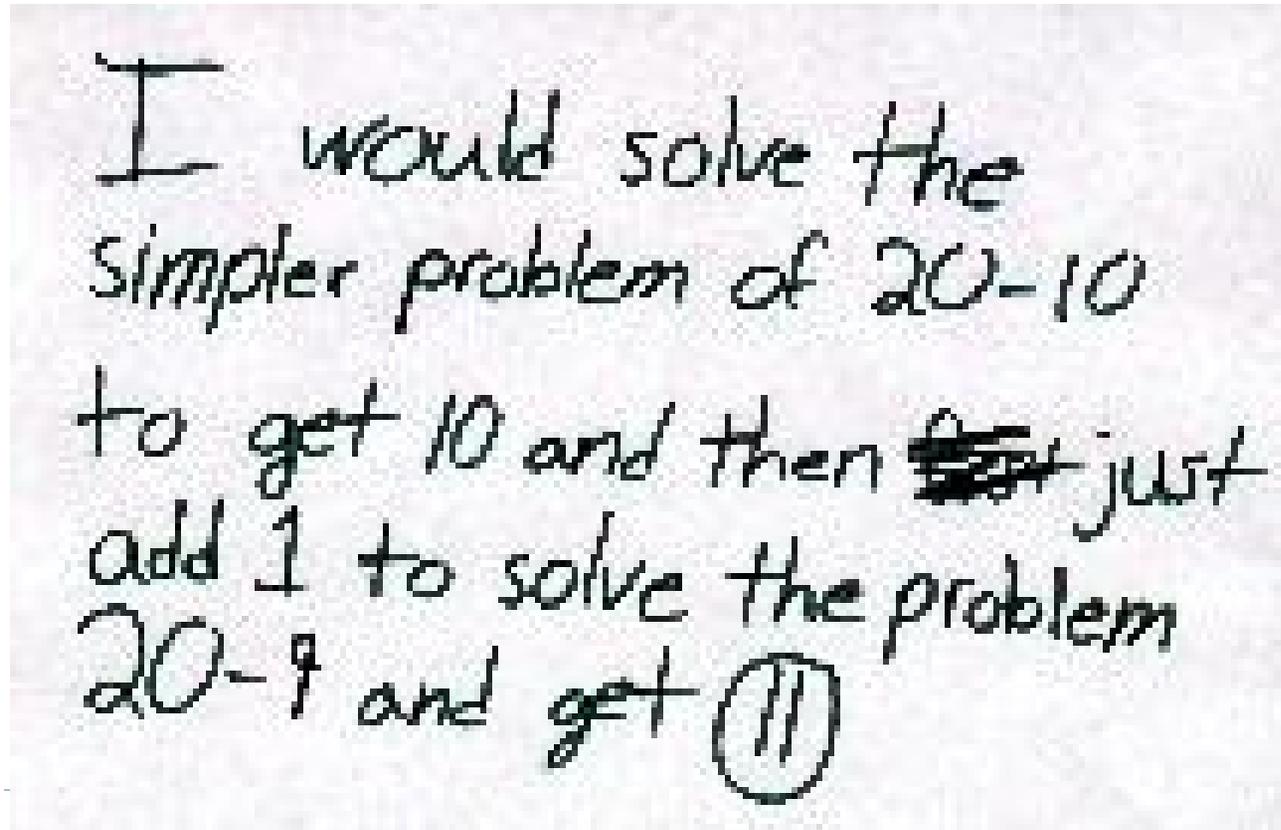


I would write 20 then  
countdown starting with  
20 then countdown  
by 1 until I took away  
9

# Active Listening

---

- ▶ Solving the Problem  $20 - 9 = ?$

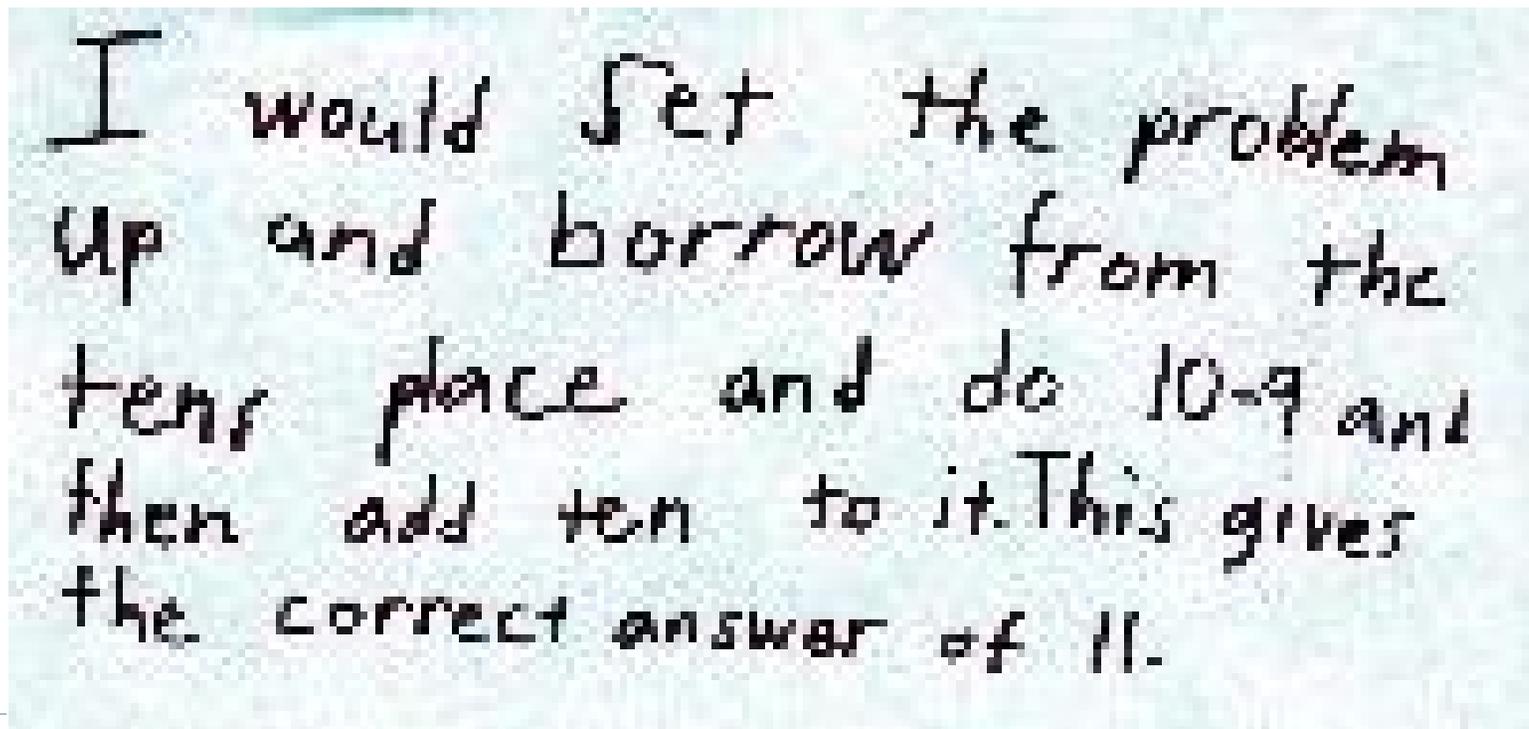


I would solve the simpler problem of  $20 - 10$  to get 10 and then ~~just~~ just add 1 to solve the problem  $20 - 9$  and get 11

# 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

The screenshot shows the homepage of the Math Reasoning Inventory website. At the top left is the logo, a blue starburst shape next to the text "Math Reasoning Inventory™". To the right of the logo are links for "Sign Up" and "Log In", along with social media icons for Facebook and Twitter. Below this is a navigation bar with five tabs: "Home" (highlighted in orange), "About the Assessment", "MRI Manager", "About Us", and "Support". The main content area has a light green background. On the left, there is a heading "Find out what your students really understand about math" followed by a bulleted list: "Focus on how students think and reason", "Uncover students' strategies, understandings, and misconceptions", and "Learn how students respond to questions the Common Core expects all middle school students to answer successfully". Below the list, it says "MRI is FREE for all teachers!". At the bottom left of this section are two buttons: "SIGN UP FOR FREE" (orange) and "LEARN MORE" (green). On the right side of the main content area, there is a video player. The video title is "Why is Marilyn Burns so excited about MRI?". The video shows an older woman with short white hair, wearing a green jacket and a colorful necklace, sitting in a red chair and gesturing with her hands. The video player has a play button in the center and a progress bar at the bottom showing "00:00".

**Math Reasoning Inventory™** Sign Up Log In f t

Home About the Assessment MRI Manager About Us Support

**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!

**SIGN UP FOR FREE** ▶ **LEARN MORE**

*Why is Marilyn Burns so excited about MRI?*

▶ ⏪ ⏩ 00:00 00:00 🔊 🗑️

<https://mathreasoninginventory.com/>

# Theory of Action

## Smarter Incorrect Distractors

---

**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**

# Item Analysis Example

---

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

Source: SSchoolPlan Data Decision Support System; <http://www.sccoe.org/schoolplan/>

# 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

---

		A	B	C	D
		$4 + 3x = 2y$	$2x - 3y = -6$	$2y + 3x = 4$	$3y - 6 = -2x$
y-intercept is 2	X=0	$4 + 3(0) = 2y$ $(0, 3/2)$	$2(0) - 3y = -6$ 6 $(0, 2)$	$2y + 3(0) = 4$ 4 $(0, 2)$	$3y - 6 = -2(0)$ $(0, 2)$
x-intercept is 3	Y=0	$4 + 3x = 0$ $(-4/3, 0)$	$2x - 3(0) = -6$ 6 $(-3, 0)$	$2(0) + 3x = 4$ 4 $(4/3, 0)$	$3(0) - 6 = -2x$ $(3, 0)$

# 7<sup>th</sup> Grade Benchmark Item Analysis Review

---

Review the 7<sup>th</sup> 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 7<sup>th</sup> 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 7<sup>th</sup> 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 7<sup>th</sup> 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



# Point Biserial Analysis

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

Grade 7 All Schools

**Answer Frequency** **Response Frequency: 76**

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

[Hide Statistical Analysis](#)

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

# Point Biserial Analysis

## 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 EOS 10-11 Math 

Grade 7 All Schools

Answer Frequency Response Frequency

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

Hide Statistical Analysis

Item No.	Item Summary	Strand / Standards Aligned	Correct Response	A		B		C		D		NR		MR		% ✓	Point Biserial	Scale Score Difficulty
				#	%	#	%	#	%	#	%	#	%	#	%			
17	 <a href="#">Mathematical Reasoning 2...</a>	B	386	51%	313	41%	47	6%	0	0%	11	1%	3	0%	41%	0.13	402	
10	 <a href="#">Algebra and Functions 1.1</a>	A	387	51%	85	11%	168	22%	114	15%	6	1%	0	0%	51%	0.35	337	
22	 <a href="#">Algebra and Functions 1.1</a>	D	93	12%	104	14%	86	11%	464	61%	8	1%	5	1%	61%	0.5	322	
20	 <a href="#">Mathematical Reasoning 2...</a>	D	69	9%	180	24%	88	12%	409	54%	11	1%	3	0%	54%	0.24	319	
15	 <a href="#">Algebra and Functions 4.1</a>	C	96	13%	74	10%	464	61%	119	16%	5	1%	2	0%	61%	0.48	319	
14	 <a href="#">Algebra and Functions 1.1</a>	D	134	18%	83	11%	83	11%	455	60%	3	0%	2	0%	60%	0.41	316	
18	 <a href="#">Algebra and Functions 1.2</a>	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				
																			Hide Statistical Analysis					
Item No.	Item Summary	Strand/Standards Aligned	Correct Response	A		B		C		D		NR		MR		% ✓	Point Biserial	Scale Score Difficulty	Reliability	25%	50%	75%	100%	
				#	%	#	%	#	%	#	%	#	%	#	%									#
17	Mathematical Reasoning 2....		B	386	51%	313	41%	47	6%	0	0%	11	1%	3	0%	41%	0.13	402						
20	Mathematical Reasoning 2....		D	69	9%	180	24%	88	12%	409	54%	11	1%	3	0%	54%	0.24	319	0.33					
16	Algebra and Functions 1.1		B	112	15%	471	62%	67	9%	101	13%	6	1%	3	0%	62%	0.33	299	0.73					
11	Algebra and Functions 1.3		A	708	93%	19	3%	17	2%	13	2%	2	0%	1	0%	93%	0.32	66	0.75					
21	Number Sense 1.1		A	707	93%	19	3%	21	3%	8	1%	5	1%	0	0%	93%	0.36	65	0.77					
10	Algebra and Functions 1.1		A	387	51%	85	11%	168	22%	114	15%	6	1%	0	0%	51%	0.35	337	0.79					
19	Mathematical Reasoning 2....		D	46	6%	139	18%	61	8%	506	67%	7	1%	1	0%	67%	0.42	287	0.8					
3	Algebra and Functions 1.1		B	15	2%	670	88%	60	8%	11	1%	3	0%	1	0%	88%	0.38	164	0.82					
4	Algebra and Functions 1.1		D	18	2%	5	1%	23	3%	713	94%	1	0%	0	0%	94%	0.4	103	0.85					

# Item Reliability

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 2<sup>nd</sup> Quarter 7<sup>th</sup> 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

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

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.

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

Algebra I Math Winter 2008				
Scaled Score	Distribution of students (n=200)		Distribution of items (n=40)	
	High Proficiency <sup>1</sup>		More difficult Items <sup>2</sup>	
YELLOW				
	2			
				35
				34 39
PINK			X	32
	1			31
			XXXX	15 24 28 40
GREEN			XXXX	6 17 19 29 30 33 36
			XXX	16 20 21 26 27 37 38
	0		X	11 14 23
			-----XXX	
			XX	3 4 25
ORANGE			XXXXXXXXXXXXXXXXXXXXXXX	12 22
	-1		X	5 8 9 13 18
			X	
			X	7 10
BLUE			X	2
	-2		X	
				1
			Less Proficiency	Less difficult items
			Each X represents 18 students	Cronbach's Alpha = .67
			(---) Average Proficiency	Person Separation Reliability = .65

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.

Figure 9: Wright Map for Algebra I Math Winter 2008 Benchmark Exam



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

# Wright Map: Placement Test

Performance Level	Ability estimate	Distribution of Students (n=174)	Distribution of Procedural items (n=40)	Distribution of Analytical items (n=25)
Advanced	4			
	3			
	2			
Proficient	1			
	0			
	-1			
Below Basic	-1			
	-2			
	-3			
		<b>80% Cut Point for Placement</b>		
Each X represents 2 students		Cronbach's Alpha = .92		
--- Average Proficiency		Person Separation Reliability = .90		

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

<b>Student Roster:</b>	SIS 11-12 Dec	<b>Assessment:</b>	 (CST-Equivalent Scale Scores and Performance Bands)
<b>Course:</b>	All	<b>Subject:</b>	Math
<b>Grade:</b>	All	<b>Grade Level Tested:</b>	All
<b>Group Year:</b>	All		
<b>My Group:</b>	All Students		

**Grade 7**

Overall Item Distribution		Distribution of Items by standards												
Levels	Scale Score	Distribution of students (n=760)	Distribution of Items (n=25)											
			Algebra and Functions				Mathematical Reasoning				Measurement and Geometry	Number Sense		
			1.1 	1.2 	1.3 	4.1 	1.1 	2.5 	2.6 	2.8 	1.1 	1.1 		
Advanced	805-999 610-804 414-609	X X X X X X X X X X												
Proficient	393-413 372-392 350-371	X X								17				
Basic	333-349 317-332 300-316	X X X X X X X X X ----- X X X X X X X X X X X X X X X X	10,14,22	18		15	13			20				
Below Basic	286-299 272-285 257-271	X X X X X X X X X X X X	16		1,8,9	12		19				25	23,24	
Far Below Basic	221-256 186-220 150-185	X X X X X X X X X X	3,4		2,5,11	6,7							21	

Each X represents <= 10 student(s)  
----- Average Scale Score

Source: SSchoolPlan Data Decision Support System: <http://www.sccoe.org/schoolplan/>

# Wright Map

---

Based on your review of the Wright Maps for the 2<sup>nd</sup> Quarter 7<sup>th</sup> 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](mailto:Bill_conrad@scccoe.org)

Dan Mason

408-453-4346

[Dan\\_mason@scccoe.org](mailto: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>





Santa Clara County  
Office of Education

# Empirical Validation of Benchmark Assessments Webinar

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

[bill\\_conrad@sccoe.org](mailto:bill_conrad@sccoe.org)  
Dan\_mason@sccoe.org