



Santa Clara County
Office of Education

Protocol for Analyzing Math Benchmark Assessments
April 25th, 2012

Math Team:

Grade Level:

Math Subject:

Introduction:

Our team seeks to improve its upcoming Math Benchmark Assessments in June. In order to accomplish this work, grade level teams of Math teachers will participate in an instructional session where they will learn how to use item analysis, classical statistics, and Wright Maps to analyze the June, 2011 Math Benchmark Assessments. Teacher teams will use the tools and reports found on the SChoolPlan system to conduct the analyses, generate finding, and make recommendations for the improvement of the June, 2012 Math Benchmark Assessments.

Purpose:

To improve the upcoming June, 2012 Math Benchmark Assessments through a careful analysis of the previous June, 2011 Math Benchmark Assessments using item analyses, classical statistics and Wright Maps.

Deliverables:

- Key findings about the June, 2011 Math Benchmark Assessments based on item analysis, classical statistics, and Wright Maps.
- 5-10 Recommendations for improving the June 2012 Math Benchmark Assessments
- 3-4 Action Steps the team will take to implement the recommendations.

Description:

Each team of teachers will receive copies of the appropriate June, 2011 Math Benchmark Assessment. Teams will review the actual assessments and answer the following questions.

Theory of Action:

Step 1: After reviewing the assessment items, does there appear to be a theory of action that underlies the overall assessment?

- One theory of action might be that the assessment uses items aligned to standards and built on Bloom's taxonomy starting with items based on knowledge, comprehension, and moving to items based on analysis.
- A second theory of action might include items that are based on student recognition of key math ideas moving to a conceptual understanding and ultimately leading to application of the concepts to the solution of problems.
- A third theory of action might include items that involve single step, two step and multiple step solutions.

- A fourth theory of action, might state that some items are aligned to more challenging standards than others.

Findings:

Step 2: 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 the team identifies.

Findings:

Step 3: Review the items from the perspective of item quality. Review the document that describes the quality characteristics for selected response items. Do any of the items reflect quality that do not meet the criteria and may be difficult for students?

Findings:

Summary:

Based on the development of findings from a review of the actual assessment, identify items on the test that may be more challenging for students to solve based on the criteria that your team developed or based on quality issues with the items.

Challenging Items Based on Key Criteria	Challenging Items Based on Quality

Based on this initial review of the items, identify several recommendations for improving the assessment.

- Recommendation 1:
- Recommendation 2:
- Recommendation 3:

Item Analysis:

Your packet will include an Answer Frequency Report that is an analysis of items that includes the percentage of students who selected the correct response and the percentage of students who selected the distractors for each item on the test. The items will be organized from the items with

the lowest percentage of correct responses to the items with the highest percentage of correct responses.

Step 4: Review the items with the least percent correct. Do these items reflect possible quality issues? Do the items reflect the challenging criteria previously described in steps 1 and 2?

Findings:

Step 5: Review and circle distractors with greater than 40% response rates. Do these distractors represent potential student misconceptions or errors in solving the problem?

Findings:

Summary:

Based on the development of findings from the item analysis, identify items that may be challenging for students or represent quality challenges.

Challenging Items Based on the Item Analyses	Challenging Items Based on Quality

Based on this item analysis review of the items, identify several recommendations for improving the assessment.

- Recommendation 1:
- Recommendation 2:
- Recommendation 3:

Classical Statistics Analysis:

The Answer Frequency Report also provides classical statistics data on the items. Please use these classical statistics analyses to generate findings and recommendations below.

Step 6: Review the classical statistics document from the perspective of the Point Biserial statistic. This statistic differentiates items based on a consistent level of difficulty. Students who perform well on the overall test should get easier items correct a high percent of the time and more challenging items correct 50% of the time. Students who do not do well overall on the test may get the easier items correct 50% of the time, but will get the more challenging items correct at a much lower percent. A positive point biserial number indicates that this rule is generally followed throughout the test for the test takers. A negative point biserial means that there is

inconsistency in this rule and that the item should be reviewed for quality. A positive point biserial of at least .015 is recommended. Good items are considered to be above 0.25.

Identify items that have a negative point biserial and review them for quality and then record your findings about these items.

Findings:

Step 7: Review the Answer Frequency Report from the perspective of Item 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

Identify items that have very high difficulty scale rankings or very low difficulty scale score rankings and record your findings about these items.

Findings:

Step 8: Review the Answer Frequency Report from the perspective of 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.

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

Findings:

Summary:

Based on the development of findings from the classical statistics analysis, identify items that may be challenging for students.

Challenging Items Based on the Classical Statistics Analyses	Challenging Items Based on Quality

Based on this Classical Statistics review of the items, identify several recommendations for improving the assessment.

- Recommendation 1:
- Recommendation 2:
- Recommendation 3:

Wright Map Analysis

Your packet contains a Wright Map of the assessment items. A 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 students on the Map. There should be items within performance bands that also align with students at that band. Students will have a 50% probability of getting items correct that are in their band. Please use the Power Point Presentation that describes 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”.

Step 9: Based on your review of the Wright Maps, does the distribution of students match the distribution of items based on the level of item difficulty? Please describe.

Step 10: Based on your review of the Wright Maps, does the Map corroborate or not corroborate hypotheses that you previously made about item complexity or quality? Please describe.

Finding:

Step 11: Does the Map identify new items that need to be reviewed because they are easier or more challenging than expected?

Findings:

Summary:

Based on the development of findings from the Wright Map analysis, identify items that may be challenging for students.

Challenging Items Based on the Wright Map Analyses	Challenging Items Based on Quality

Based on this Wright Map review of the items, identify several recommendations for improving the assessment.

- Recommendation 1:
- Recommendation 2:
- Recommendation 3:

Based on this review of the 2011 7th Grade 2nd Quarter Math Benchmark Assessment, the team developed the following Action Steps to improve the quality of the assessment.

Action Step 1:

Action Step 2:

Action Step 3:

Action Step 4:

References

Dan Mason, Mark Moulton, Ph.D., Dale Russell, Ed.D., Diana Wilmot, Ph.D., 2009. Three Facets of Formative Assessment. Santa Clara County Office of Education. San Jose, CA.