Using Data from Common District Assessments to Inform Decisions about Middle School Mathematics: Lessons from Shoreline Public Schools

Jack B. Monpas-Huber, Ph.D.
Director of Assessment & Student Information

Shereen Henry, M.Ed.
P-12 Mathematics Specialist

SHORELINE PUBLIC SCHOOLS

Overview of this Session

This session is a *case study* of our development and **common district end-of-unit math assessments** and use of data from those instruments to inform decisions in **middle school mathematics**

In this session we'll cover the following topics:

District Context for this Work
Why Common Assessments?
What Kind of Assessments?

Developing and Administering the Assessment
Sample Results

Issues Raised
Challenges
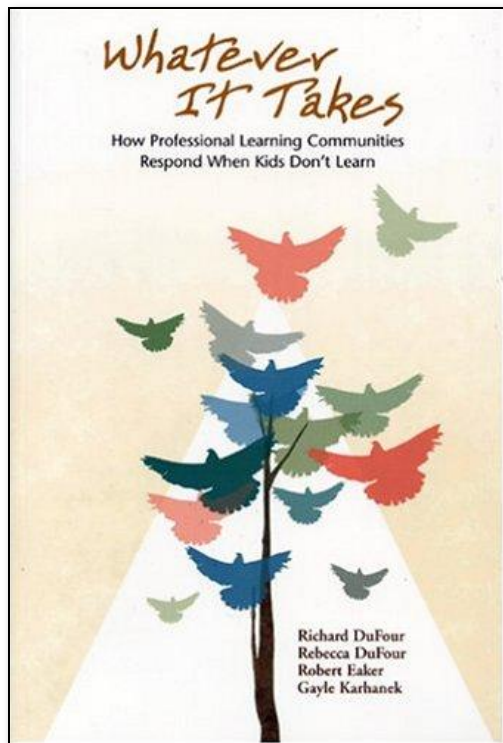What Worked

Context of this Work

District focus on math

- New 2008 math standards and high school curriculum adoption
- Strong community engagement in math
- Math Achievement Team (MAT)

District efforts to align programs

Program Alignment Team

Context of this Work

Professional Learning Communities

The Four Fundamental Questions

1. What exactly do we expect students to learn?
2. How do we know when they've learned it?
3. What do we do when they haven't learned it?
4. What do we do when they have learned it?

Why Common Assessments?

The site-based PLC argument:

Common assessments are an important component of teacher-based PLCs (Dufour, etc.)

- They ground teacher discussion in student learning data
- They provide common language and reference point for teachers
- They reinforce common performance and content expectations
- They surface effective teaching practices for sharing
- The data are a basis for differentiation, intervention
- Data for building- and team-based SMART goals

Why Common Assessments?

The system argument:

Establish a common base of information for Shoreline as a system

- "The Board wants district data"
- Provide common data source for administrative meta-PLC
- Reinforce common district expectations for content coverage and performance
- Data to inform professional development in content areas based on student needs

What Kind of Common Assessments?

Type	Primary Purpose/s	Examples
Large-scale summative	Accountability measure Comparing similar schools Program/curriculum evaluation Inequality data	WASL / MSP
Medium-scale interim/benchmark	Predict performance on state assessment Identify strengths/weaknesses Reinforce common expectations "System" formative	NWEA-MAP easyCBM / DIBELS District end-of-unit exams District end-of-grading period exams
Small-scale classroom formative	Inform daily instruction minute-to-minute, hour-to-hour, day-to-day	Quizzes, projects, informal, ungraded

What Kind of Common Assessments?

Type	Primary Purpose/s	Examples
Large-scale summative	Accountability measure Comparing similar schools Program/curriculum evaluation Inequality data	WASL / MSP
Medium-scale interim/benchmark	Predict performance on state assessment Identify strengths/weaknesses Reinforce common expectations "System" formative	NWEA-MAP easyCBM / DIBELS District end-of-unit exams District end-of-grading period exams
Small-scale classroom formative	Inform daily instruction minute-to-minute, hour-to-hour, day-to-day	Quizzes, projects, informal, ungraded

Our Goals Going into this Work

- Needed to be teacher-developed

- Make it useful, worthwhile to teachers in creating more effective instruction

- Collect some useful data for the district

- Proactive focus on validity / technical quality, consistency

- Build assessment literacy

Key Decisions

Focus on Middle School

1. High school was getting a new curriculum
2. Elementary needed more work with standards
3. Middle school was the only group whose curriculum (Connected Math Project (CMP)) had been aligned to the new math standards prior to the beginning of the year
4. Middle school was a small group
5. Middle school teachers expressed some interest in collaborating on common assessments

Teacher-developed end-of-unit assessments

II. Developing and Administering the Assessment

Assembling the Team

First Iteration

Began with 3-hour after-school class
- PowerPoint with purpose, vocabulary
- Designed to provide assessment literacy and build foundation for a common assessment

Second Iteration

Day-long class to deeply understand the process

Hands-on experience going through the steps of building a common assessment

Two 7th grade assessments, two 8th grade assessments

The Concept of a Test Map

PE	Description	Items		Points	
		N	%	N	%
8.1C	Represent a linear function with a verbal description, table, graph, or symbolic expression, and make connections among these representations.	6	40.0%	22	48.9%
8.1D	Determine the slope and y-intercept of a linear function described by a symbolic expression, table, or graph.	1	6.7%	5	11.1%
8.1E	Interpret the slope and y-intercept of the graph of a linear function representing a contextual situation.	1	6.7%	3	6.7%
8.1F	Solve single- and multi-step word problems involving linear functions and verify the solutions.	4	26.7%	12	26.7%
8.1G	Determine and justify whether a given verbal description, table, graph, or symbolic expression represents a linear relationship.	3	20.0%	3	6.7%
Total		15	100.0%	45	100.0%

Item Specifications

What are item specifications?
Presented teachers with WASL item specs

What do we already know about different item types and how can we build on that knowledge?
Listed different item types
What are the benefits of each item type? How could familiar item types be used in new ways?

What kind of item rules should we agree on?
Ultimately, no restrictions on types of questions and how many for each standard or exam

MSP Item Specs later were incorporated into pacing guides which now informs common assessments
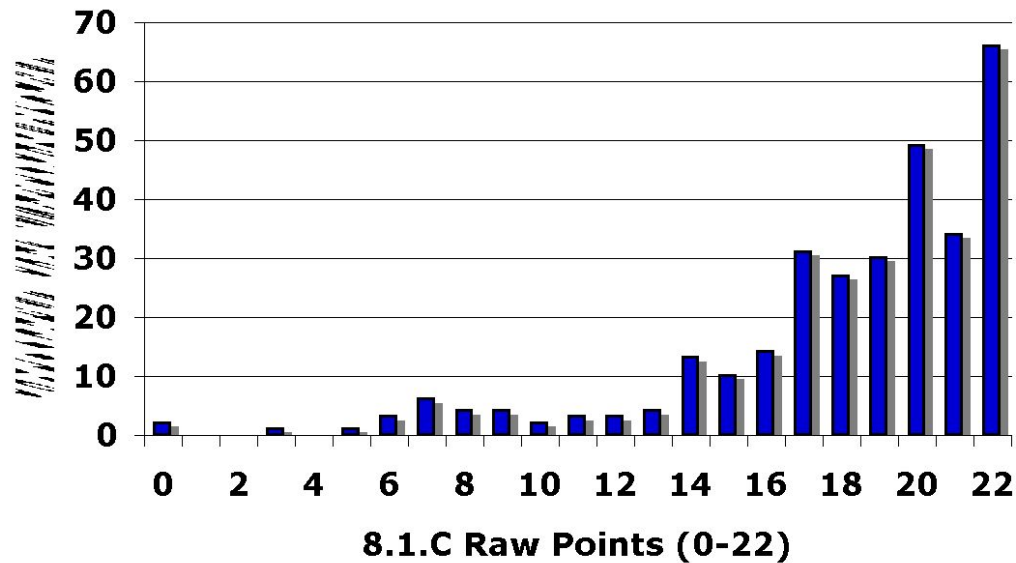
Sample Results

ALBERT EINSTEIN MIDDLE SCHOOL
Common Math Assessment - Grade 8
Thinking with Mathematical Models
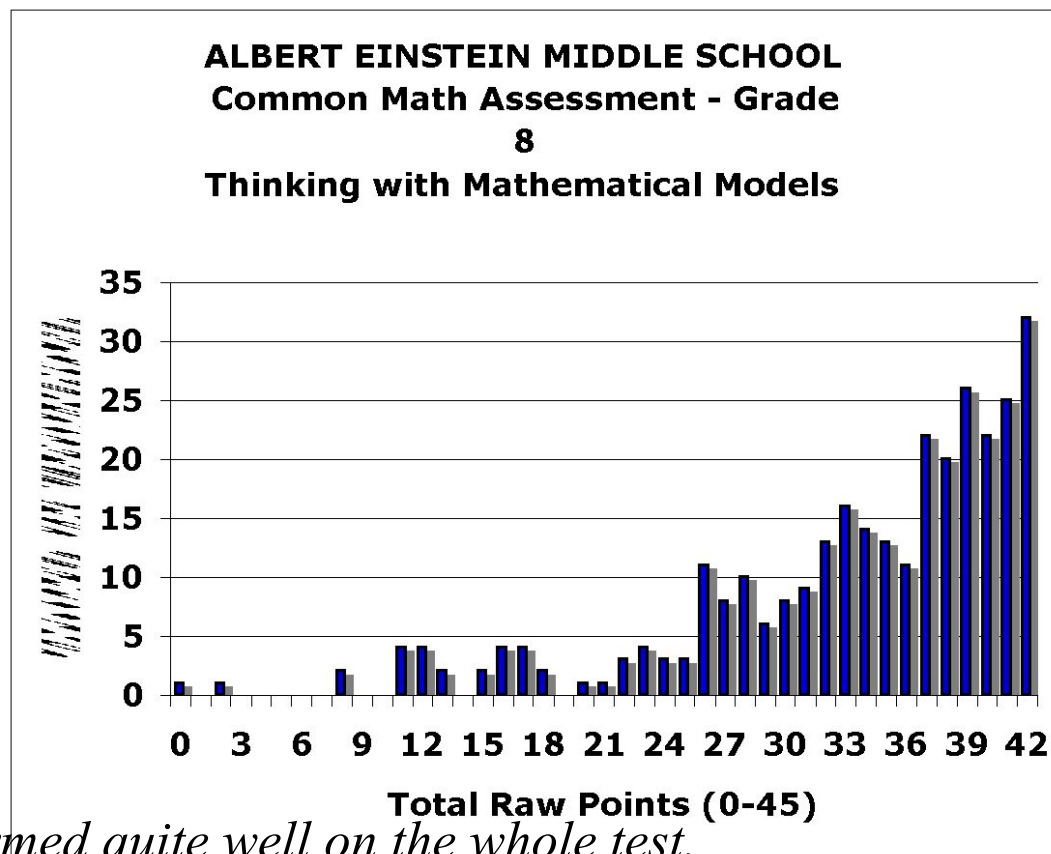
Students performed quite well on the whole test.
Where might we set the cut score for "meeting standard"?
Next step for PLC is how to intervene for the low-scoring students -- motivation and homework issues
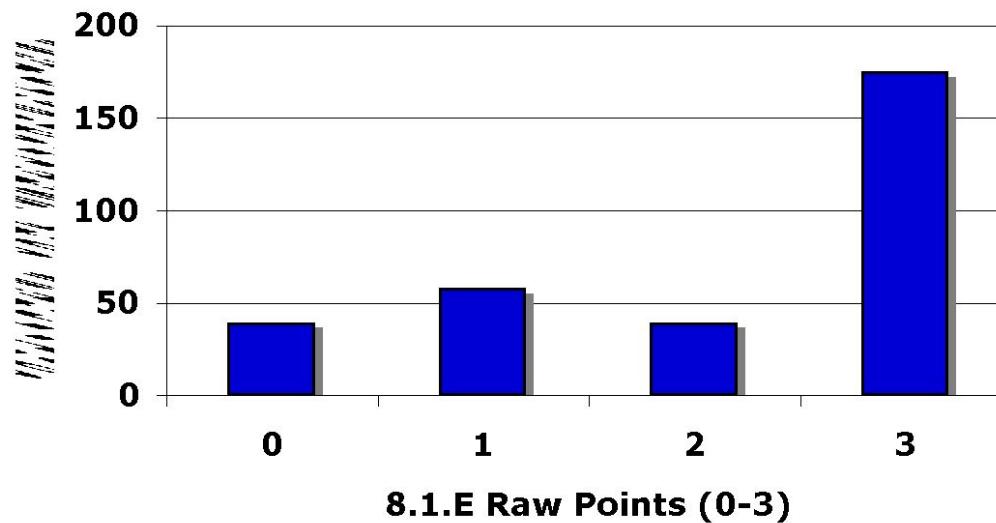
Sample Results

ALBERT EINSTEIN MIDDLE SCHOOL
Common Math Assessment - Grade 8
Thinking with Mathematical Models
PE 8.1.C: Representing Linear Functions

8.1.C Raw Points (0-22)

Sample Results

ALBERT EINSTEIN MIDDLE SCHOOL
Common Unit Math Assessment - Grade 8
Thinking with Mathematical Models
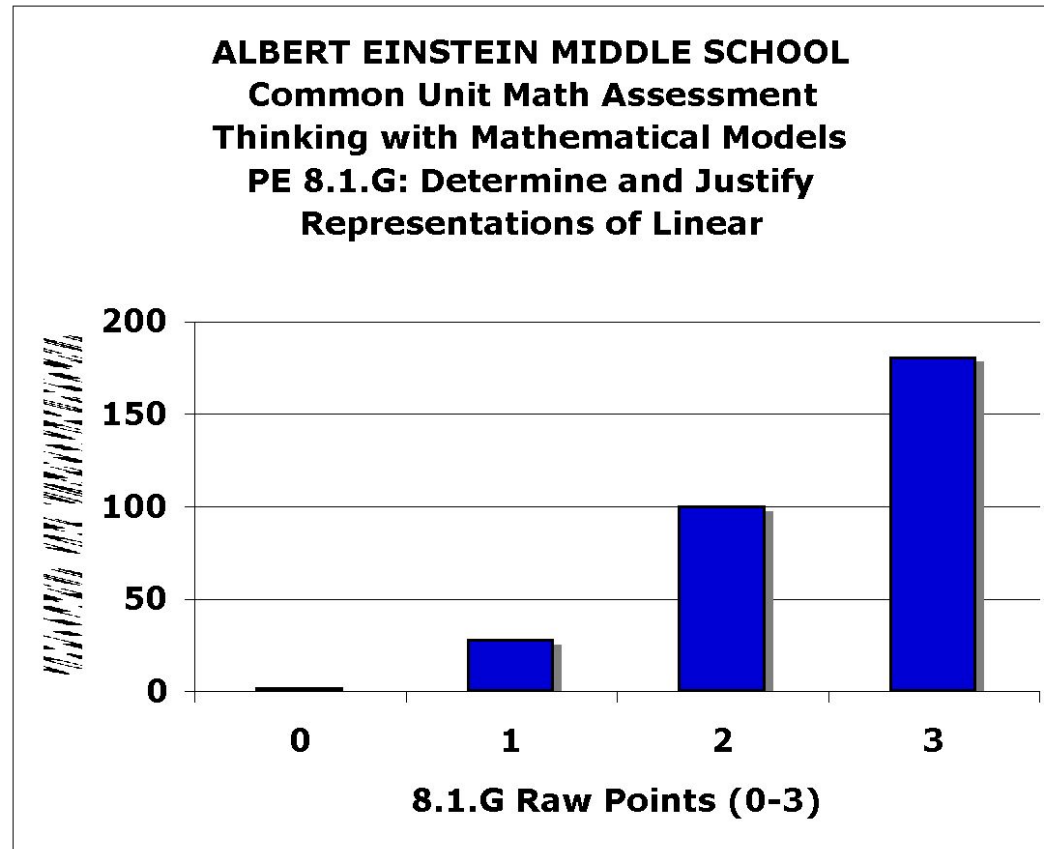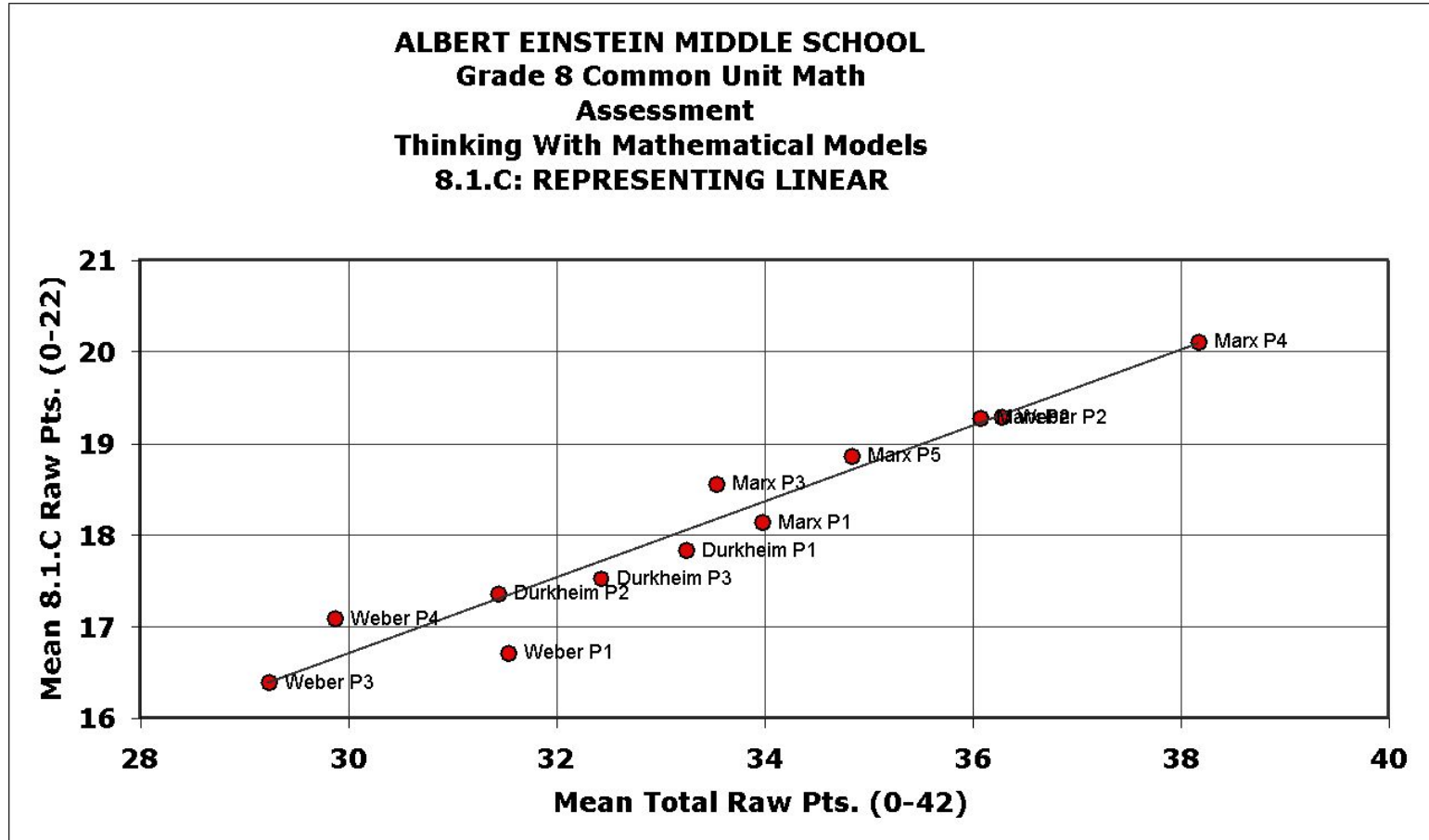PE 8.1.E: Interpret Slope and Y-Intercept
in Context

Sample Results

ALBERT EINSTEIN MIDDLE SCHOOL
Common Unit Math Assessment - Grade 8
Thinking with Mathematical Models
PE 8.1.D: Determine Slope and Y-Intercept
from Symbolic Expression

8.1.D Raw Points (0-5)

Sample Results

ALBERT EINSTEIN MIDDLE SCHOOL
Common Unit Math Assessment
Thinking with Mathematical Models
PE 8.1.G: Determine and Justify
Representations of Linear

8.1.G Raw Points (0-3)

Sample Results

ALBERT EINSTEIN MIDDLE SCHOOL
Grade 8 Common Unit Math
Assessment
Thinking With Mathematical Models
8.1.C: REPRESENTING LINEAR
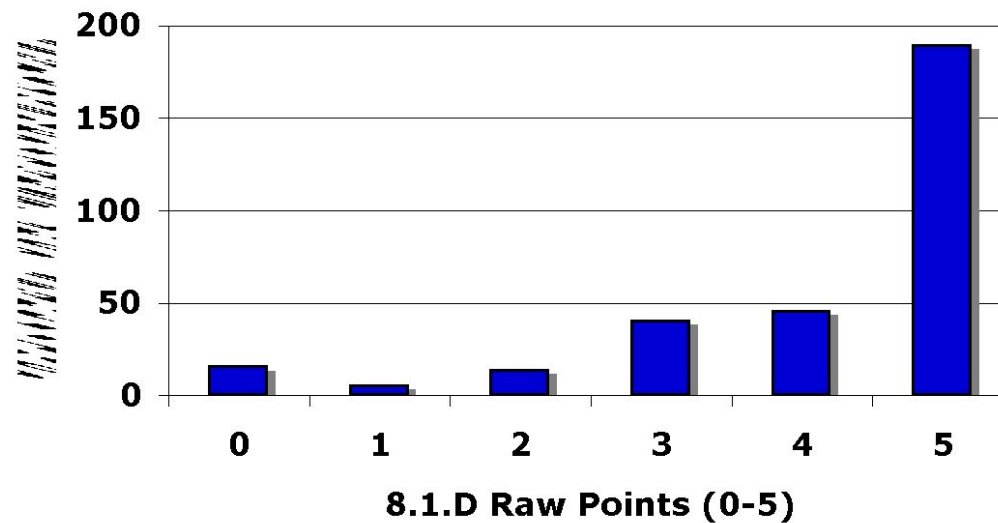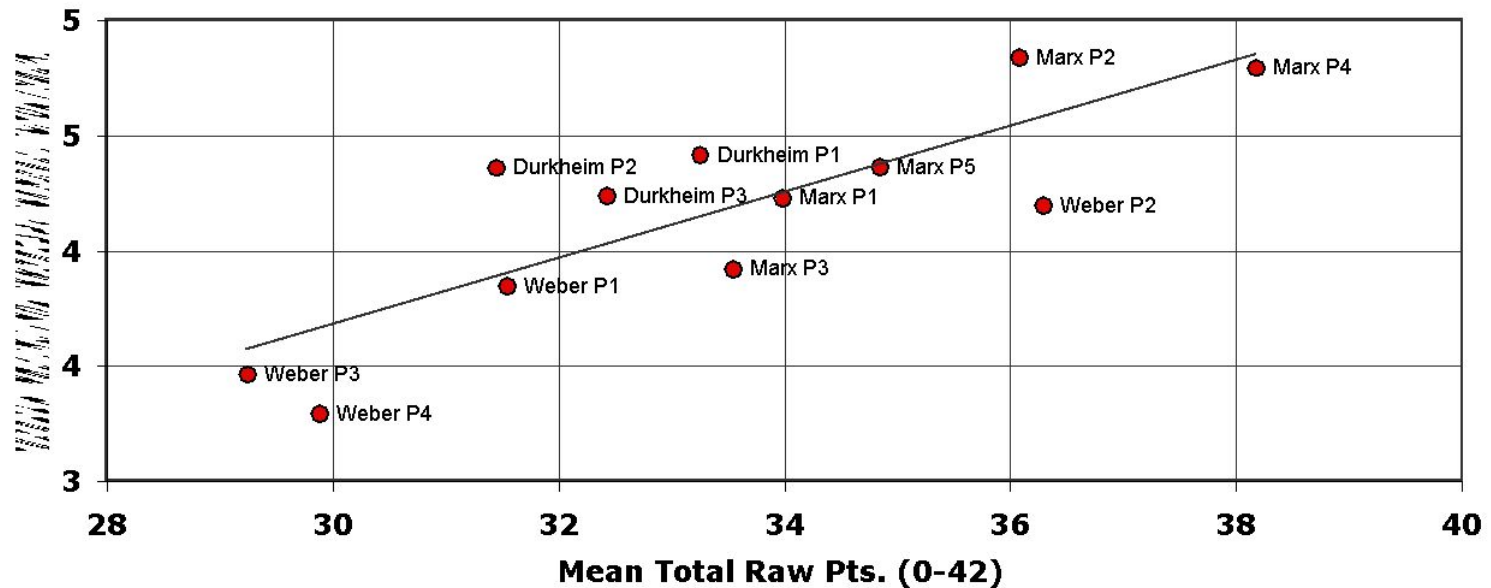
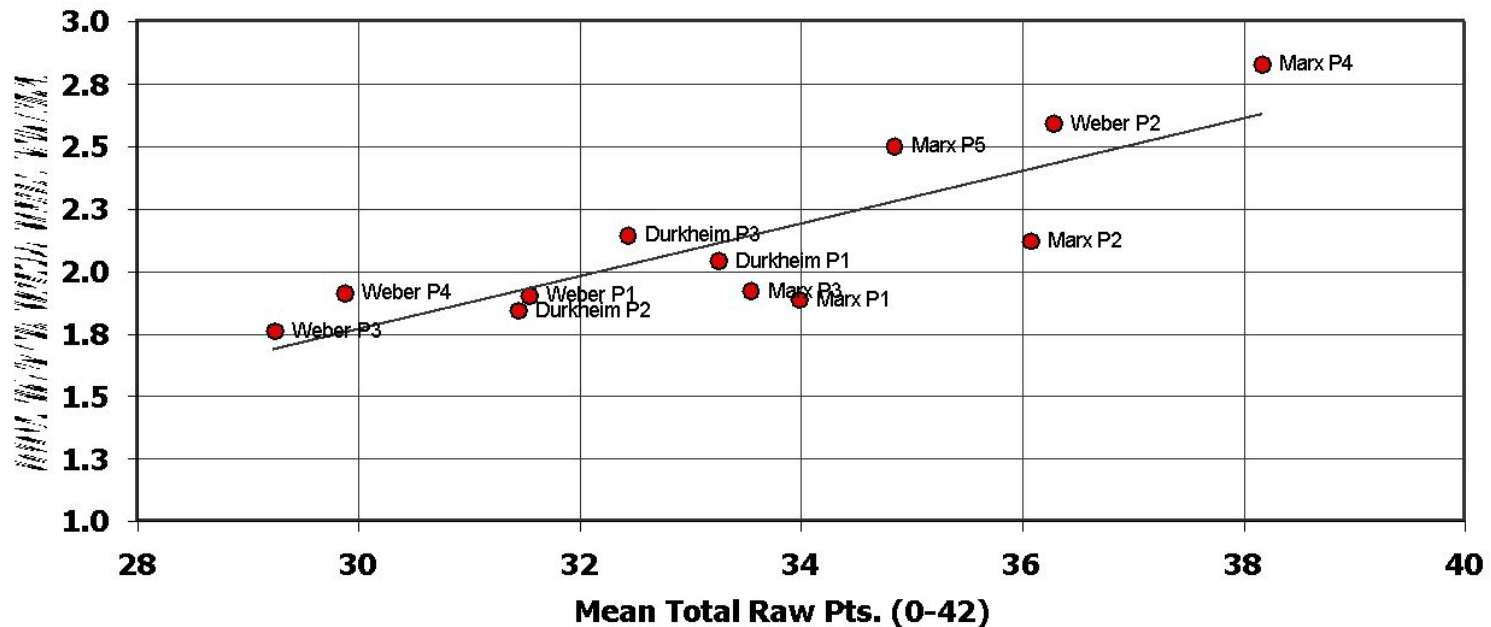**Use the Item-Total Correlation (Scatterplot) to (Fairly) Compare Teachers**

Sample Results

ALBERT EINSTEIN MIDDLE SCHOOL
Grade 8 Common Unit Math Assessment
Thinking With Mathematical Models
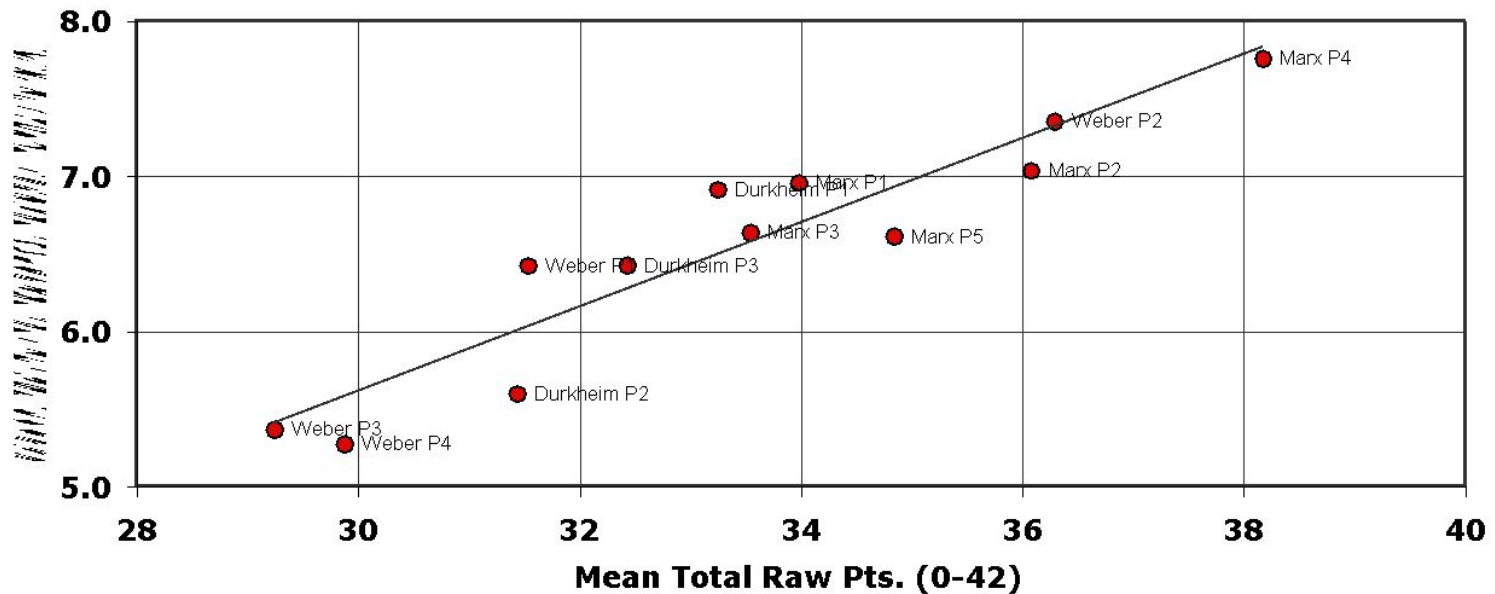8.1.D: DETERMINE SLOPE AND Y-INTERCEPT FROM
SYMBOLIC EXPRESSION
(N = 307)

Sample Results

ALBERT EINSTEIN MIDDLE SCHOOL
Grade 8 Common Unit Math Assessment
Thinking With Mathematical Models
8.1.E: INTERPRET SLOPE AND Y-INTERCEPT IN CONTEXT

Sample Results

ALBERT EINSTEIN MIDDLE SCHOOL
Grade 8 Common Unit Math Assessment
Thinking With Mathematical Models
8.1.F: SOLVE SINGLE- AND MULTI-STEP WORD PROBLEMS RE:
LINEAR FUNCTIONS
(N = 307)

Sample Results

ALBERT EINSTEIN MIDDLE SCHOOL
Grade 8 Common Unit Math Assessment
Thinking With Mathematical Models
8.1.G: DETERMINE AND JUSTIFY REPRESENTATIONS OF
LINEAR RELATIONSHIPS
(N = 307)

Mean Total Raw Pts. (0-42)
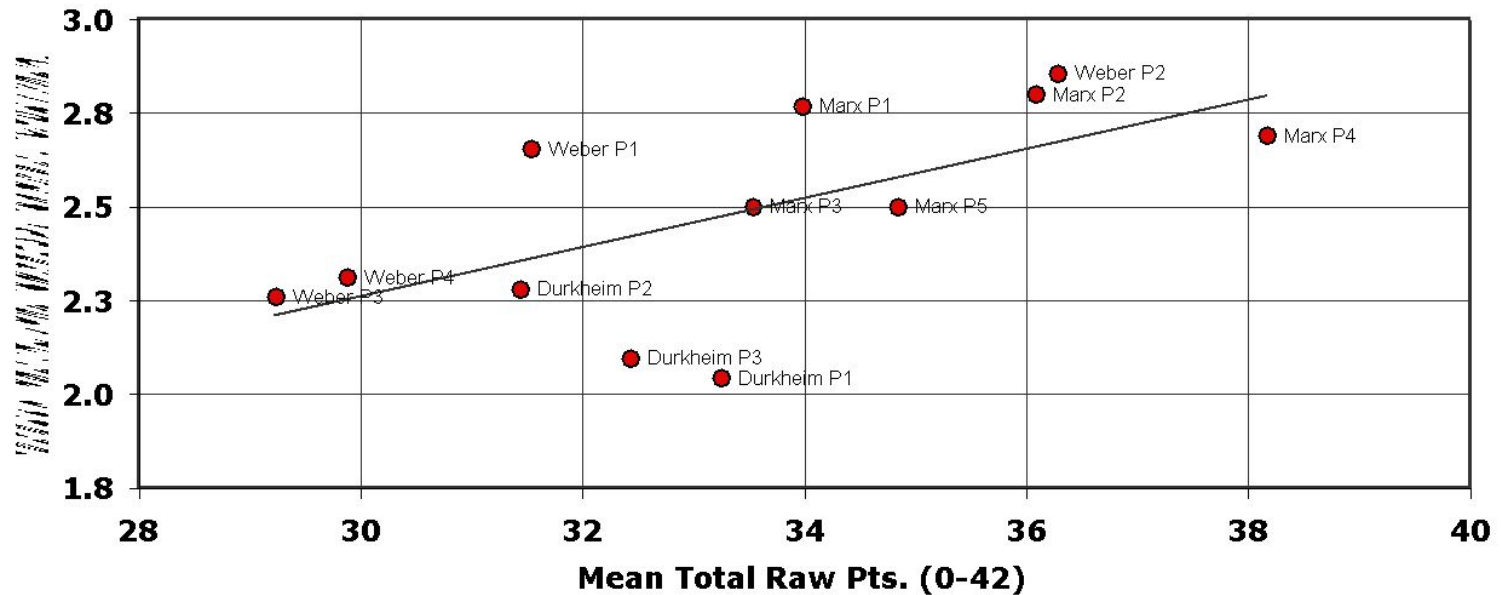
III. Making Sense of the Experience

What We Learned from the Results

- Students performed quite well on the assessments overall

- Teacher-level variation was quite minimal

- Teachers used the data to inform their SMART goals

Validity Issues

Item Analysis / Reliability: Were the scores reliable? Any items that didn't perform well?

Construct validity: Would a factor analysis confirm that items loaded on the constructs we would expect based on the test map?

Standard setting: What counts as proficient?

Instructional sensitivity: Are the results sensitive to real differences in teaching and learning?

Consequential validity: Did the results lead to more effective teaching and learning?

Issues Raised

District Vision

- What should be the district direction on this?

- Common district and/or common building assessments?

Challenges

Introducing a New Method

- Teachers had a personal system in place for creating a test.
- Asking them to do a different, more elaborate, process, was difficult.
- They needed rationale right away.

Challenges

<u>Competing Purposes</u>

- Teachers wanted specific instructional information from the assessment.
- The district wanted system-wide information.
- We had to build something to serve both sets of needs.
- This meant being flexible with the formatting of the assessment.

Challenges

New Vocabulary

- Teachers were not familiar with the terms "test map" and "item bank."
- These ideas had to be introduced and incorporated into the process.

Challenges

Collaborating on the Scoring

- The most difficult task was building the rubric and assigning points.
- Teachers had to agree on what they were truly assessing and what was most important.
- They often unintentionally weighted a topic with too many points which then had to be adjusted.

Challenges

The Item Bank

- The creation of an item bank has not yet been successful. There has yet to be enough time for them to
    (a) align existing questions to the standards or
    (b) write new questions that match the standards.
- Finding a common place to store questions was also a challenge.

Challenges

Challenge of District Level Teaming

- Once the test was implemented and graded, the learning-focused conversations happened within the PLC group at the building level

- This has created a desire to discontinue the district level work and just do building-level common assessments.

What Worked

<u>Step 1</u>
Clarify the purpose of the assessment. How will the information be used?

<u>Step 2</u>
Identify the essential learnings to be assessed. Sometimes these are part of a math performance expectation, sometimes two combined performance expectations, or sometimes the exact wording of a performance expectation.

What Worked

Step 3

Make a *rough* test map of the essential learnings that maps out number of questions and percentage of the overall points.

Note: This often changes in the process of making the rubric. Work on the rubric and work on the test map are interwoven.

What Worked

Step 4
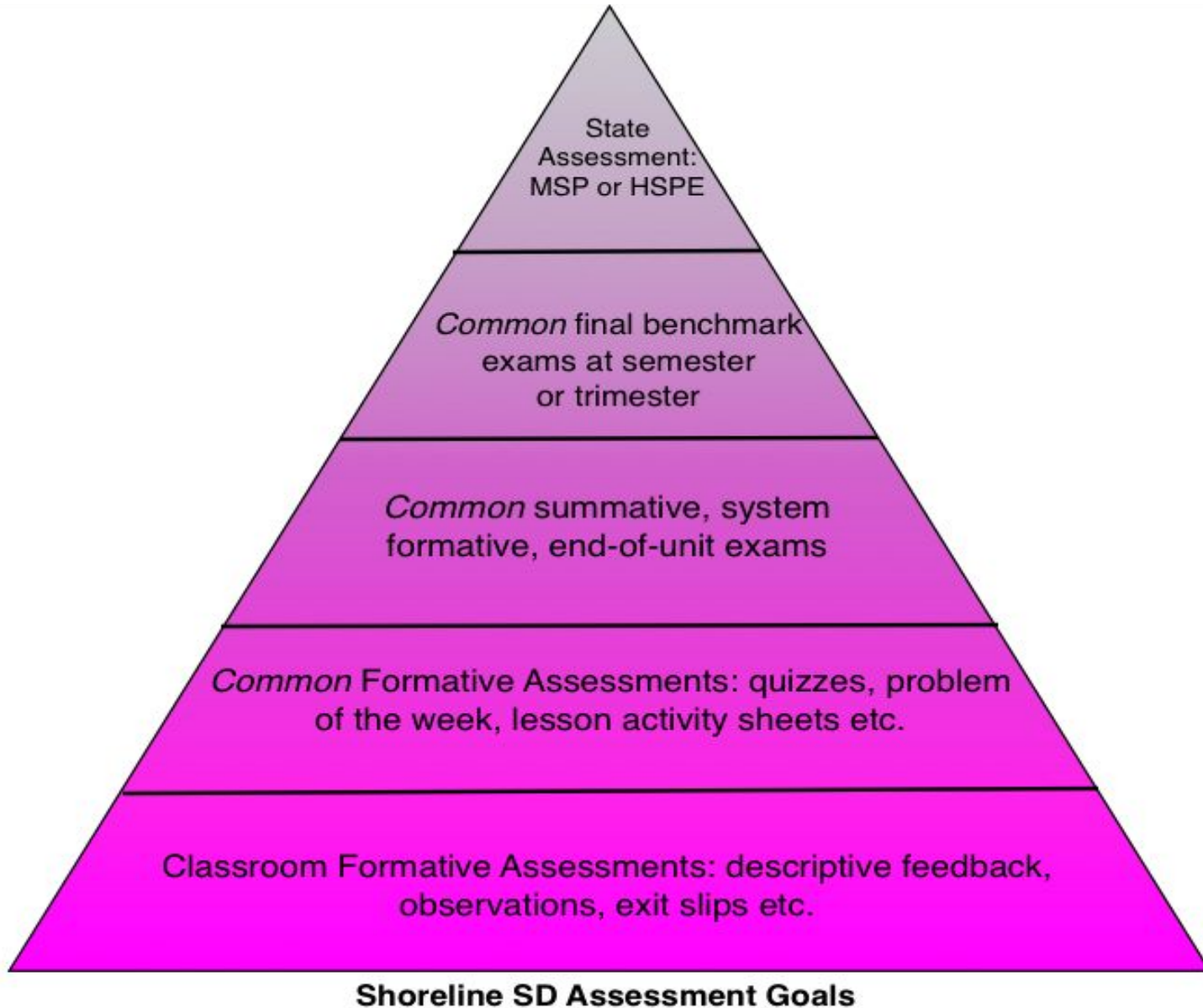Create the questions that match each essential learning.

Step 5
Then build a scoring rubric for each question specifying the evidence necessary to earn points. *This was a powerful discussion insofar as it surfaced real differences in philosophy and assessment practices.*

Step 6
Go back to the test map and compare it with the rubric. We had to revise the map or the rubric to get a match.

A Tentative Model for Assessment System

State Assessment: MSP or HSPE

Common final benchmark exams at semester or trimester

Common summative, system formative, end-of-unit exams

Common Formative Assessments: quizzes, problem of the week, lesson activity sheets etc.

Classroom Formative Assessments: descriptive feedback, observations, exit slips etc.

Shoreline SD Assessment Goals

Advice to Other Districts

- Purpose is paramount

- Small group

- Process that evolves

To Contact Us

Jack B. Monpas-Huber, Ph.D.
Director of Assessment & Student Information
jack.monpas.huber@shorelineschools.org
(206) 361-4774

Shereen Henry, M.Ed.
P-12 Mathematics Specialist
shereen.henry@shorelineschools.org
(206) 368-4709

SHORELINE
PUBLIC SCHOOLS