Washington State Assessment Conference
December 2008

Aligning Assessments to Monitor Growth in Math Achievement: A Validity Study

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

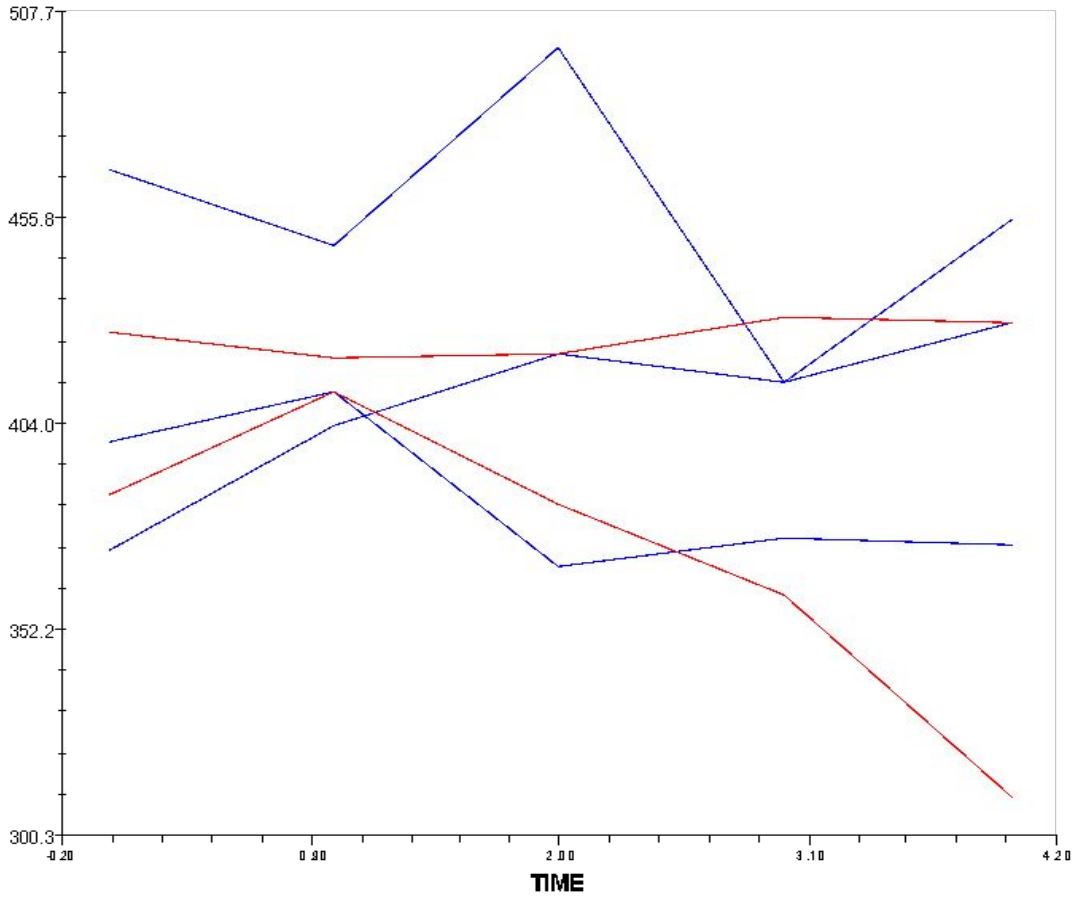
SHORELINE PUBLIC SCHOOLS

Core Themes

1. Technical quality of locally developed district assessments as outcome measure for program evaluation
   - Growth modeling in HLM

2. Fidelity of local district instrumentation to state instrumentation
   - Sampling same domain
   - Built from same maps and item specs
   - Measuring same thing?

Evaluating Program with Growth Curves



Possible Questions:

1. ***What does growth in math look like over the course of one year? Do the data match our expectations and theory of action?***

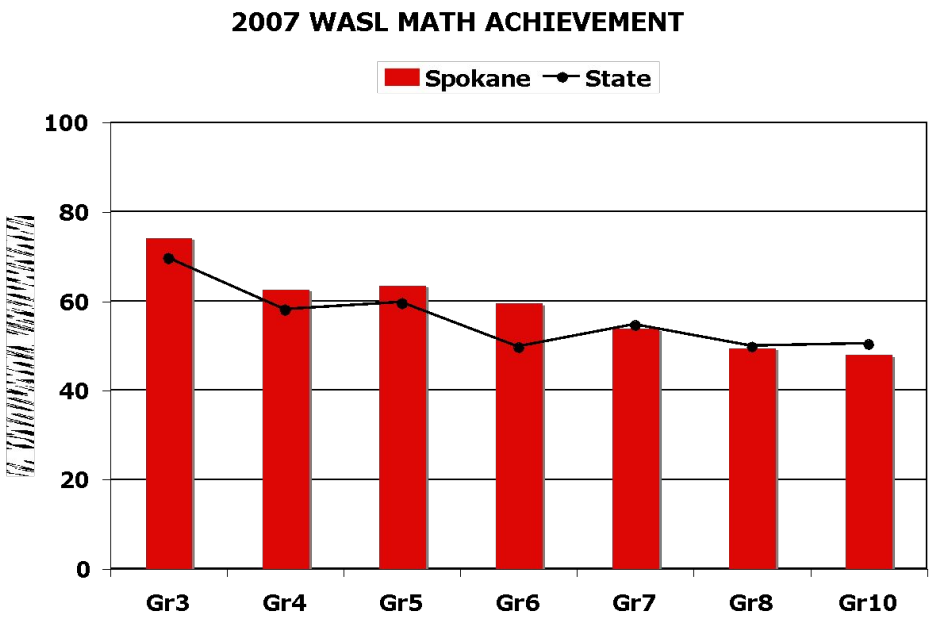
2. ***Why do some students have steeper growth curves than others?***

3. ***What causes growth? Does math coaching explain variation in the slopes of students' growth curves?***

Background of this Project

Issue #1:

WASL math achievement seems to decline after 3rd grade



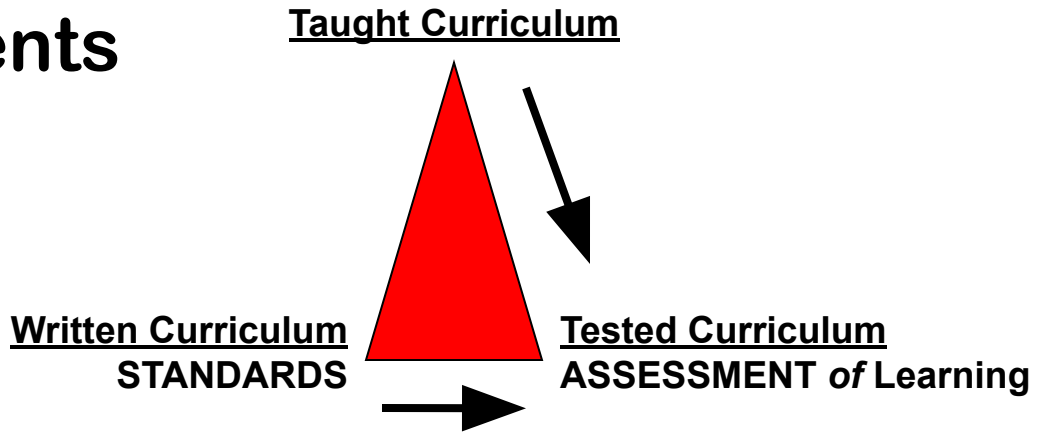
Working Explanations:

1. **Reflection of curriculum and instruction?**
2. **Artifact of WASL standard setting**

Background of this Project

Issue #2:

District assessments ready for action



EVOLOVING PURPOSES (AND VALIDATION) OF ASSESSMENTS

- ***Monitor learning across the system***
- ***Inferences about fidelity to curriculum, instructional effectiveness***
- ***Curriculum development, professional development***
- ***Predictive of WASL performance***

Design for Exploratory Study

What's happening to our students from Grade 5 to Grade 7?

Longitudinal: Follow the same students from Grade 5 to Grade 7

Stratified: Approximately 10 students per classroom

Repeated measurements: What can we learn from these data sets?.....

- 2007 Grade 4 WASL
- Fall SASL
- Winter SASL
- 5 End-of-unit district assessments
- 2008 Grade 5 WASL

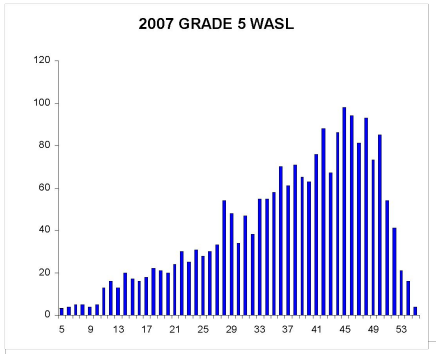
10 waves

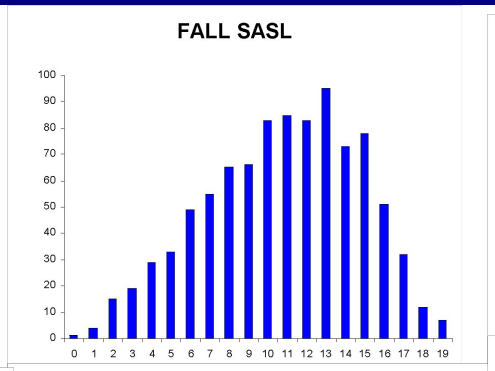
Research Questions

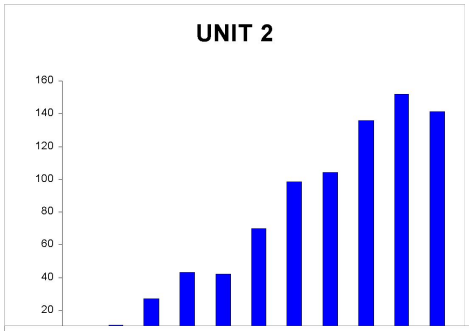
What can we validly infer from WASL-like district benchmark assessments about….

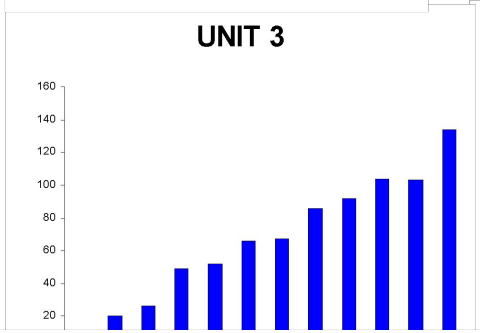
1. **Fidelity to curriculum: Are teachers teaching the district curriculum?**
2. **Rigor of curriculum: Is it rigorous enough? Are we preparing our kids adequately?**
3. **Growth toward state standard: Are our district assessments showing us true achievement of state standards?**

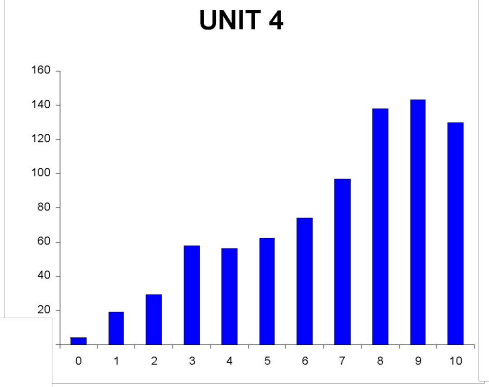
Results: Performance on District Assessments

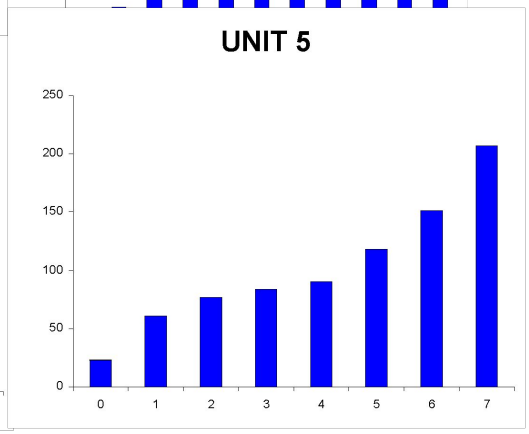


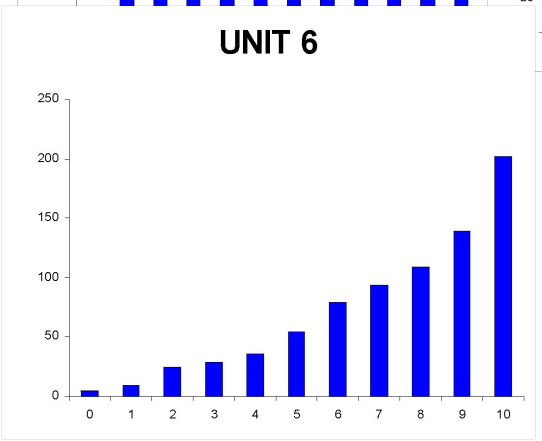


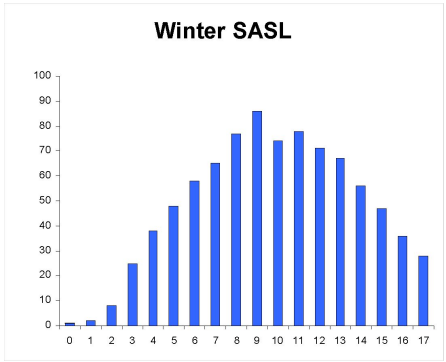












QuickTime™ and a
TIFF (Uncompressed) decompressor
are needed to see this picture.

Research Questions

Further Questions about Growth toward State Standard:

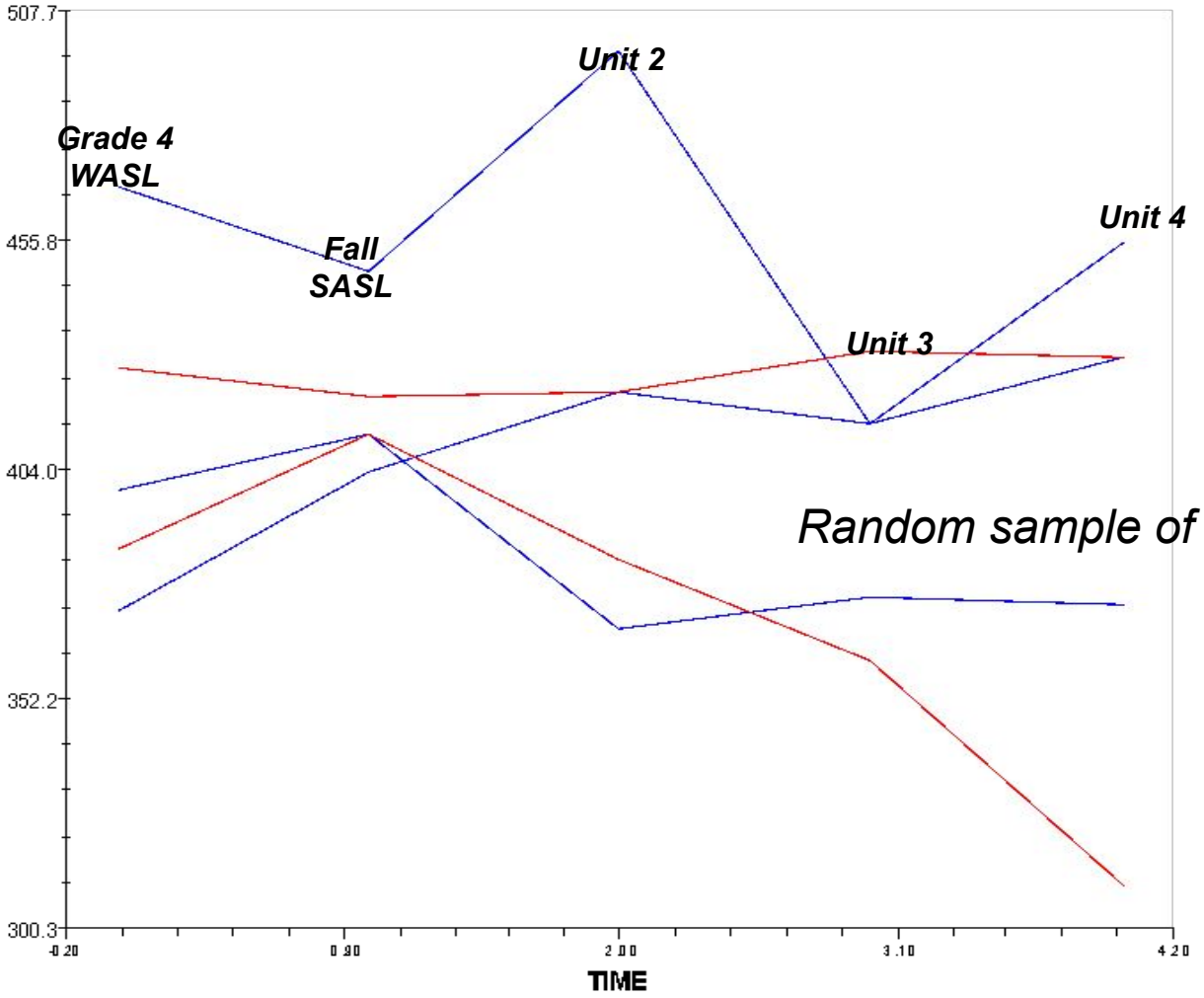
- **Are our district assessments showing us true achievement of state standards?**
- **Are our districts assessments aligned to WASL and measuring the same thing?**
- **How can I "link" scores from district assessments to WASL so that they all sit on the same scale?**
- **Could difficulty parameters for local items help teachers see the relative difficulty of different skills?**

Linking District Assessments to WASL

Equating, Scaling and Linking Literature

- Linear Equating
- Equipercentile Equating
- Item Response Theory (IRT)
  - Rasch Model (Winsteps) places item difficulties and person abilities on the same scale
  - "Common person" single group design
  - Concurrent calibration

Results: Tentative Growth Curves



Random sample of 5 students

Two Challenges

Reliability

- A function of test length and redundancy of items
- District assessments fairly short
- HLM raised doubts about reliability of growth curves

Dimensionality

- Rasch model assumes one dimension
- Each district assessment measures somewhat different content

Broader Issues for District Assessments

Locally developed WASL-like district benchmark assessments

- Advantages: Local capacity building, content validity
- Challenge: Time and expense of development, technical quality, scaling and equating

Outside Instruments, e.g., Measures of Academic Progress (MAP)

- Advantage: Technical quality, growth scale
- Challenge: Content validity