

# How Common Standards Might Support Improved State Assessments

*Presentation to:*  
Board on Testing and Assessment  
Workshop on Best Practices in State Assessments

December 11, 2009  
Washington, DC

*Presenter:*  
Laurel L. Wise, HumRRO

# Presentation Outline

1. What needs to be supported?
  - Increased value of state assessments
    - Wider range of uses
    - Better validity, reliability, fairness
  - Feasibility of improvements
  - Sustainability of improvements
2. How can common standards support improved state assessments?
  - Increased funding for development of common assessments
  - Greater opportunity to study the content to be assessed
3. How might state assessments be improved?
  - More meaningful reporting scales
  - Timely results more closely tied to instruction
  - Richer, more diagnostic feedback
  - Greater teacher engagement
  - Validity studies to support wider interpretations and uses

# What needs to be supported?

- Need to support a wider range of uses
  - Most state assessments are adequate for identifying children being “left behind”, but
  - Have limited diagnostic value for individual students
    - Highly correlated subscale scores with low reliabilities
    - Information not available until after school year is finished
  - Are not great as school accountability measures
    - Covers only a limited range of what is taught (reading, math, science at some grades) particularly at the high school level
  - Are limited as indicators of instructional effectiveness
    - Again, only a limited range of what is taught is covered
  - Are not well connected to interim assessments used by many districts and schools for mid-year tracking

# What needs to be supported?

- Need improved validity, reliability, and fairness to overcome current:
  - Limited demonstrations of score validity
    - As indicators of instructional effectiveness
    - As predictors of readiness for subsequent instruction
  - Limited score reliabilities
    - Too many tests, too little time (given limited uses)
    - Targeted to very broad content specifications (mile wide)
  - Limited support for ensuring fair access for all
    - Universal design not yet well understood
    - Limited research on validity of alternate accommodations
    - Current focus on what is known more than what is taught gives advantage to students from higher SES families

# What needs to be supported?

- Feasibility
  - Of computer-based testing for richer content, adaptability to improve precision, and reduced costs
  - Of automated scoring methods for timely reporting and reduced ongoing costs
  - Of deeper content analyses to improve diagnostics
- Sustainability
  - Affordability of administration and processing costs (particularly if extensive human scoring is required)
  - Affordability of on-going development costs
    - To support high quality in a typically “low-bid” environment
  - Of teacher involvement in and benefit from assessment development and results

# Increased Funding

- What states are spending now
  - Modest survey of state assessment costs
    - Only covers activities that are contracted out
    - Attempted to assess costs of different types of assessments
    - CCSSO is conducting a more definitive cost study
- Increased funding by pooling development costs across states with common standards
  - Positive NECAP example – this actually works
  - No savings on administration and scoring costs
- Race to the Top Assessment Grants
  - Serious federal funding for developing and implementing common assessments

# Current Assessment Cost Survey

- Survey of 15 state testing programs and 2 test development contractors
- Attempt to separate:
  - fixed development costs (per year for each state) and
  - variable administration, scoring, and reporting costs (per test administered)
- Separated regular and alternate assessments
- For regular assessments, separated costs for strictly multiple choice versus extended constructed response (CR) tests

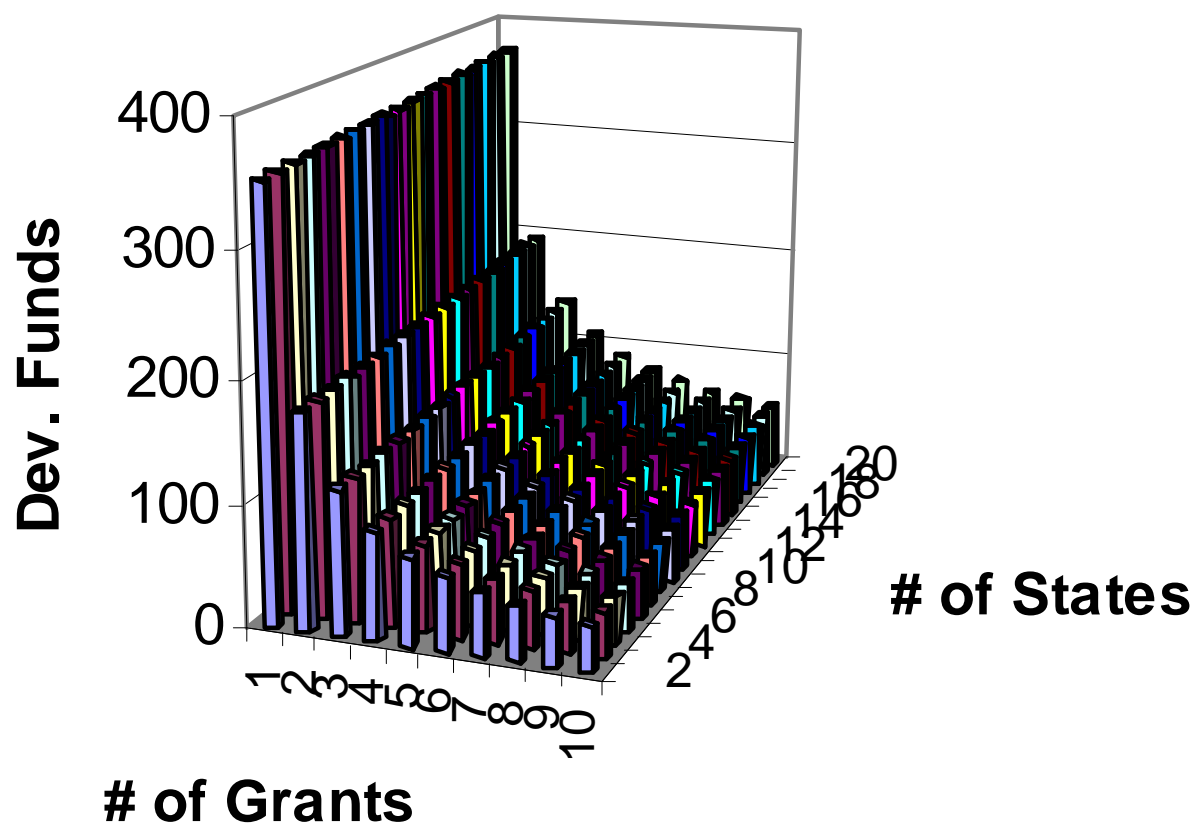
# Current Assessment Cost Estimates

- **Development Cost (annual / state, about 20%)**
  - Regular MC: \$900K (\$200K – \$2,000K)
  - Regular CR: \$1,300K (\$300K - \$3,600K)
  - Alternate: \$400K (\$100K - \$700K)
- **Variable Costs (per student tested, about 80%)**
  - Regular MC: \$3 (\$1-\$10)
  - Regular CR: \$27 (\$14-\$42)
  - Alternate: \$350 (\$50 - \$850)
- **Summary**
  - About \$1M/state for on-going development costs
  - Need to keep variable costs under \$25 / student

## Increased Funding with Common Standards

- Greatly increased funds for development
  - \$350M / n from RTT Assessment Grants
  - About \$1M / state for ongoing development
  - Example: If 10 states work together and get one of 5 Assessment Grants:
    - \$70M + \$10M = \$80M for initial development
    - \$10M for ongoing development from pooling
    - With no increase in current development costs
- Need to keep administration, scoring, and reporting costs at or below current levels
  - Unless costs of interim assessments can also be reduced

## Development Funds by Number of Grants and Number of States



# Deeper Study of Test Content

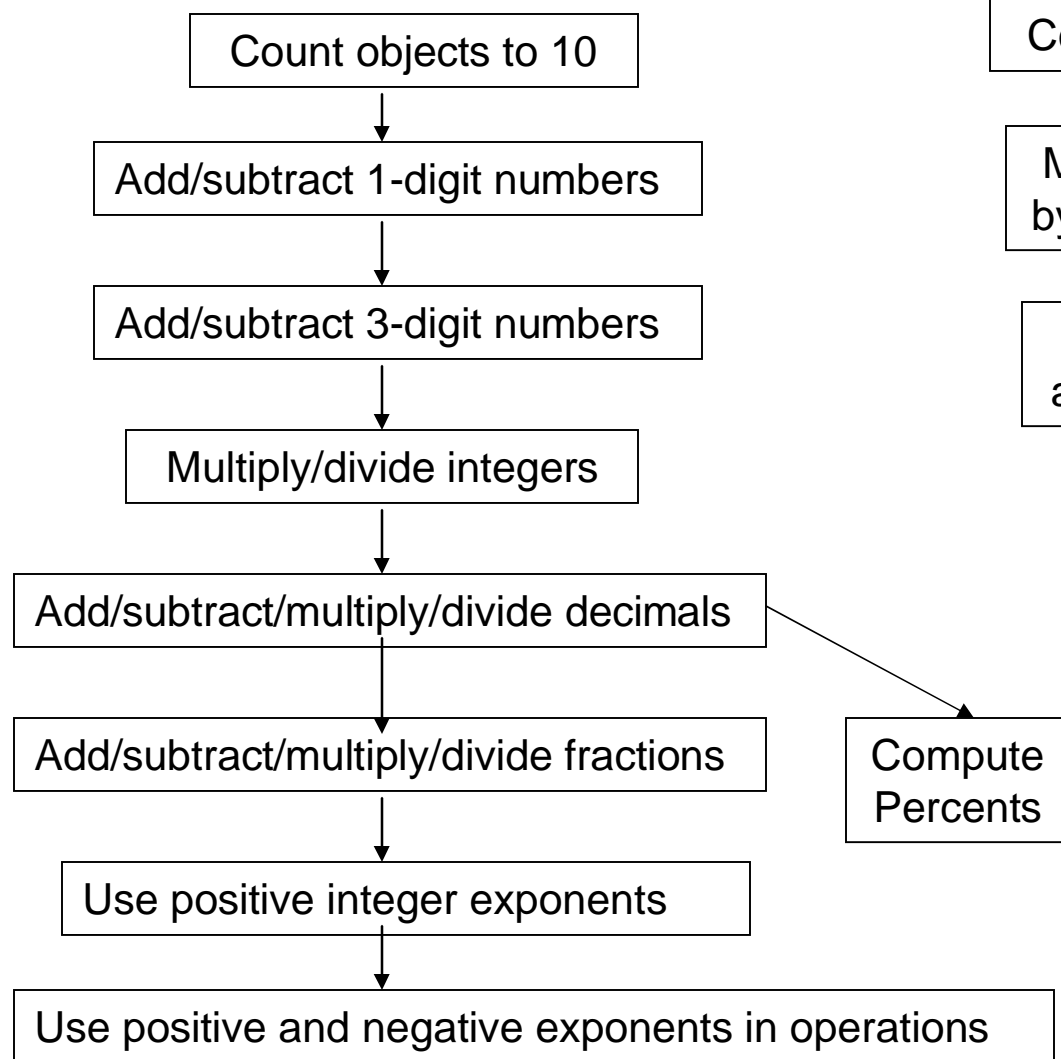
- To improve instruction as well as assessment
  - Cognitive analyses of skills to be taught and assessed (*Knowing What Students Know, pages 92-102*)
    - E.g., what does it mean to read at grade level?
  - Deeper understanding of barriers to skill acquisition
    - E.g., understanding “carrying” in multi-digit addition
    - How to detect key barriers as well as overcome them
    - Note: diagnosis is more about finding what is wrong as through careful construction and analysis of incorrect answers
  - Evaluate alternatives for sequencing coverage of different objectives
    - Might support integrated interim assessments – test as taught rather than wait for the end of the year

## Content Analysis: Learning Trajectories

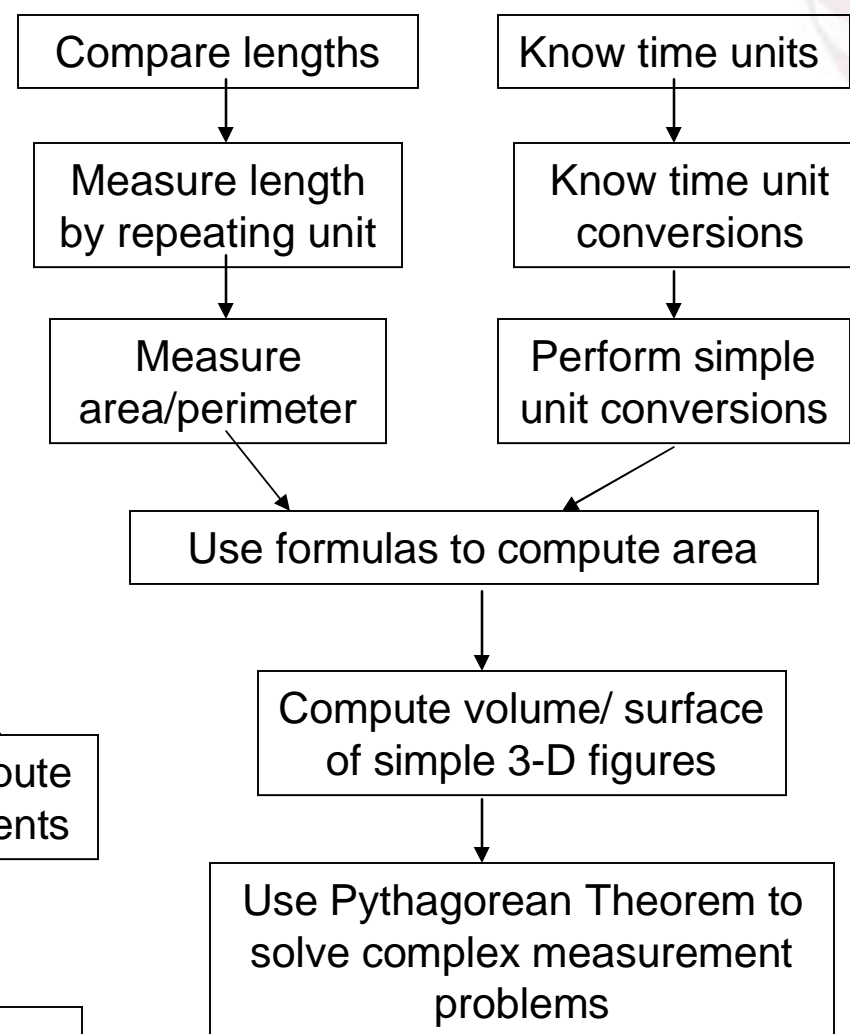
- Need content/skill model that is well articulated across grades
  - Focus on what new content is to be added each year
  - Macro-level growth trajectory targets
  - **Should be** evident in common core standards
- Need a good model of within-year sequencing
  - Micro-level growth trajectory targets
  - May be (somewhat) multivariate (e.g., separate tracks for content and skill)
  - **May not be** evident in initial version of core content standards

# Example of Learning Trajectories – Elementary/Middle Math

## Numbers and Operations



## Measurement



# Improvements to be Supported

- More meaningful reporting scales
  - Not just a number of an arbitrary scale or the percent above an arbitrary cut on this scale
- Timely results to support use in instruction
  - Not sometime next year
- Richer diagnostic feedback
  - What the student does not yet know or cannot yet do
- Greater teacher engagement
- Validity studies that support extended uses
  - Linking student achievement to instruction
  - Showing readiness for further instruction

# Improved Reporting Scale

- Not grade-by-grade performance levels
  - With broad, general descriptions of content mastered
  - That may lead to uneven expectations across grades
- Not an arbitrary numerical scale
  - Interval properties of vertical growth scales are suspect
- Instead, a Learning Trajectory Scale
  - Points on the trajectory related to mastery of specific objectives
  - Also points to next objectives to be mastered
  - Requires a careful and validated sequencing of objectives
  - Goes beyond “grade-level” testing restrictions

## More Timely Results

- Multiple assessments focused on smaller sets of content (mid-terms or quarterly assessments)
- Adaptive tests that assess mastery of specific skills or areas of content
  - Place students accurately along learning trajectories
  - Machine scored with *immediate* feedback
- Followed by 1 or 2 extended answer questions
  - Targeted to estimated point on the trajectory
  - Teacher-scored, with some external audits
  - Providing diagnostic information (why student is at this point and not the next)
  - Used to confirm or modify initial score level estimate

# Richer Diagnostic Information

- From learning trajectory scales
  - Analogous to the Bookmark approach to standard setting
    - Provides information on the next several pages to be studied
- From multiple assessments per year
  - Focused on target sets of objectives
  - Supports mid-course corrections
  - Within-year progress provides a better measure of program and teacher effectiveness

# Better Summative Information

- End-of-year estimate of progress toward overall readiness
  - Need to map within-year trajectories onto larger models (K-12) of learning progressions
  - Absolute level to report to students and parents
    - Is the student on track to meet post-secondary readiness standards by the 12<sup>th</sup> grade?
  - Year-to-year growth for school accountability and feedback on programs and teachers
- Multiple assessments support greater reliability compared to a single end-of-year test

# Greater Teacher Engagement

- How can teachers be more fully engaged in improvements to assessment and instruction?
  - Involvement in test development
    - Creating and/or reviewing test questions or exercises
  - Scoring and interpreting responses to extended constructed response questions
    - Must be perceived as useful; not just another demand
  - Using improved assessment information to monitor and improve instructional effectiveness
  - Assisting in ongoing enhancement to learning trajectory models

# Validity Evidence

- **Alignment Studies**
  - Internal consistency across items assessing specific content clusters
  - Mapping to assure coverage of all objectives at appropriate depths of knowledge
  - Evidence of common patterns of progression
- **Convergent Validity Studies**
  - Agreement with teacher judgments of placement within learning trajectories
  - Prediction of subsequent progress toward overall readiness targets

# Summary

- Common standards can support significant improvements in state assessments
  - Increased funding by pooling assessment development efforts
    - As well as from RTT Common Assessment grants
  - Cognitive analyses of common, well-articulated standards to support improved instruction as well as improved assessment
    - To support assessing and preventing barriers to mastering specific objectives
    - To improve sequencing of instruction and testing

## Summary (Continued)

- Increased funding and better content analyses would support:
  - More meaningful reporting scales
    - Including measures of short-term and long-term growth
  - More timely results
    - Immediate feedback from machine score sections
    - From mid-year as well as end-of-year assessments
  - Richer diagnostic information
    - Prescriptive as well as descriptive
  - Greater teacher engagement
  - A wider range of validity studies
    - Supporting a wider range of interpretations and uses