

# *Using Assessments to Determine Teacher Performance*

Mariann Lemke

November 14, 2013

Center on  
**GREAT TEACHERS & LEADERS**

at American Institutes for Research ■

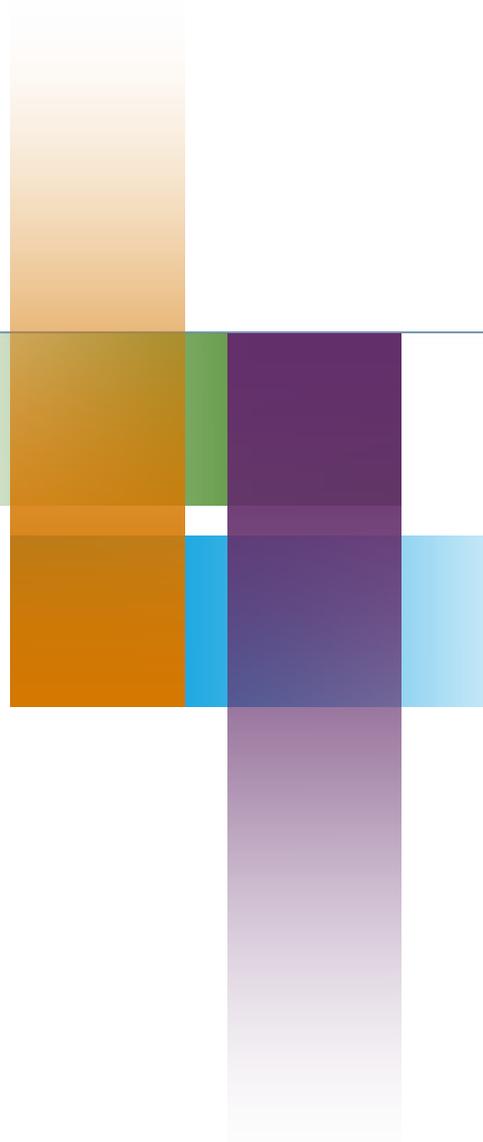
 **AIR**<sup>®</sup>  
AMERICAN INSTITUTES FOR RESEARCH<sup>®</sup>



# Mission of GTL Center

---

The mission of the Center on Great Teachers and Leaders (GTL Center) is to foster the capacity of vibrant networks of practitioners, researchers, innovators, and experts to build and sustain a seamless system of support for great teachers and leaders for every school in every state in the nation.



# About AIR

---

- AIR's mission is to conduct and apply the best behavioral and social science research and evaluation toward improving peoples' lives, with a special emphasis on the disadvantaged.
- One of the largest not-for-profit behavioral and social science research organizations in the world.
- Headquartered in Washington, DC, with approximately 1,600 employees around the world working in education research, assessment, evaluation, and technical assistance.

# Overview

---

- Give an overview of the national landscape related to the use of student achievement data for teacher evaluation, including various state approaches
- Summarize benefits and limitations related to the use of student achievement data

# Why This Topic?

---

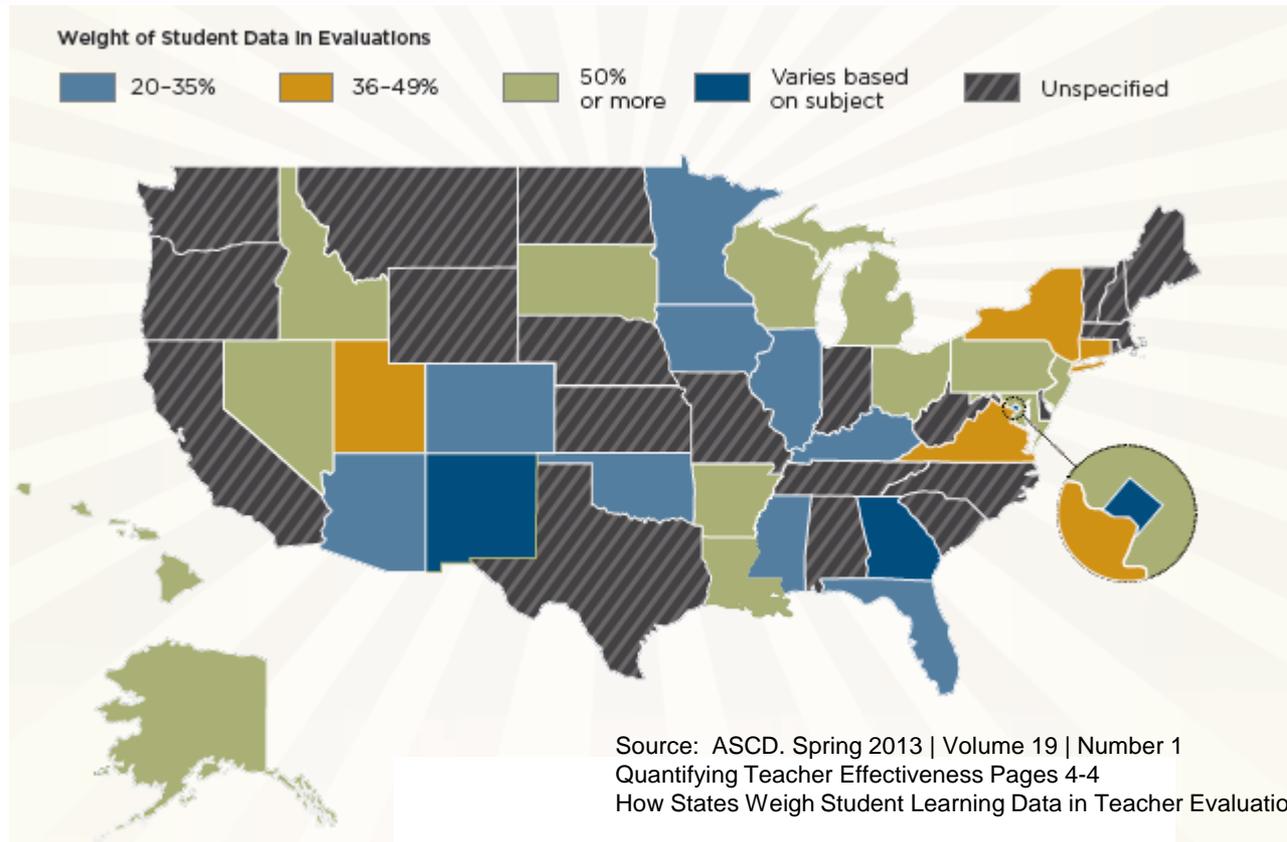
- Ultimate goal is to improve teaching and learning.
- One mechanism to improve teaching and learning is improved evaluation systems that provide more information than previous systems.
- Desire to implement evaluation systems that provide information on what educators do (professional practice) and on how well students learn (outcomes).

# National Landscape

---

- Forty-four states and the District of Columbia have updated educator evaluation legislation and rules in the last few years (and several others have changes pending).
- Changes include mandates or recommendations to incorporate student achievement data into educator evaluation, with variation in the specificity of requirements.
  - Focus on student growth, taking into account student starting points rather than a measure at a single point in time.

# National Landscape on Use of Student Achievement for Educator Evaluation



# Using Assessments as a Teacher Evaluation Measure (State Level)

---

- Teachers of tested grades and subjects (e.g., 4th grade reading)
  - Approximately 35 states use or will implement a statewide value-added or student growth percentile measure
- Teachers of non-tested grades and subjects (e.g., art, music, kindergarten)
  - Approaches include:
    - Collective measures based on tested grades and subjects (e.g., schoolwide value-added)
    - Building or buying assessments to measure growth
    - Student learning objectives/goal-setting

# General Considerations in Using Assessments for Teacher Evaluation

---

## Does the assessment or measure:

- Align with what students are expected to learn and teachers are expected to teach
  - Address content standards and/or course content and complexity of content
- Measure growth and fairly assess all students
  - Capture student learning at different levels
  - Address a progression of skills and content
  - Contain clear directions and questions that do not advantage some students
- Have evidence of reliability
  - Use administration and scoring procedures that produce consistent and comparable results and contain sufficient items to assess desired content

# Considerations in Using Value-Added Measures

---

- Value-added measures can provide useful information.
  - Positive correlation between student growth measures and other measures of teacher performance (e.g., instructional practice, principal evaluations)
  - Evidence that teachers with high value-added scores do something different (as measured through observations) than teachers with low value-added scores
  - Evidence that teachers with high value-added scores have a positive effect on future student achievement and other long-term outcomes
- Value-added or growth measures should not be used alone for high-stakes decisions.
- Communication and stakeholder engagement are critical for new or complex measures.



# Considerations in Using Value-Added Measures

---

**Value-added measures are not perfectly precise or reliable.**

- Value-added measures have limited ability to differentiate among teachers.
  - Measures should report precision of scores and use precision/confidence intervals in assigning teacher ratings.
- Student growth measures depend on test data, which is itself an imperfect measure for a variety of reasons.
  - Models should account for measurement error.



# Considerations in Using Value-Added Measures

---

**Value-added measures are not perfectly precise or reliable.**

- Quality and availability of data linking students, teachers, and courses can be a challenge; even with great data, questions of attribution are difficult to address.
  - Test timing, co-teaching, push-in/pull-out situations, non-random assignment of students, and other real-life issues can make it hard to determine who (or what) accounts for student learning. Consider mechanisms to verify data.
- Small numbers of students can lead to imprecision and instability.
  - Develop appropriate business rules and consider multiple years of data.

# Teachers in Untested Grades and Subjects

---

Approaches include:

- Collective measures based on tested grades and subjects (e.g., schoolwide value-added)
- Building or buying assessments to measure growth
- Student learning objectives/goal-setting

# Examples and Considerations: Collective Measures

---

Example: Use of schoolwide value-added scores from other subjects for teachers in untested grades and subjects in Tennessee, Nevada

- May be attractive option because these scores build on existing data, but there are strong validity concerns with assigning scores from one subject to teachers in other subjects

# Examples and Considerations: Building and Buying Assessments

---

## ■ Examples:

- Creating lists of approved or recommended assessments
  - Colorado, New York, Ohio
- Developing new statewide assessments using internal teams or vendors
  - North Carolina, Delaware, Florida
- Providing guidance for local assessment development
  - Massachusetts, Colorado

## ■ Considerations:

- May be difficult and expensive to develop or purchase assessments that cover all subject areas
- Providing guidance alone may be insufficient to ensure assessment quality

# Colorado Department of Education

<http://www.coloradopl.org/assessment/assessments>

cde CDE Resource Bank

Home
Posts
Announcements
Events
Assessments
Documents
Members

**CONTENT KNOWLEDGE KEYWORD(S)**

**CONTENT AREA**

- Career & Technical Education
- Comprehensive Health
- Dance
- Drama
- Math
- Music
- Physical Education
- Reading, Writing, and Communicating

**CAREER AND TECHNICAL EDUCATION**

**EDUCATION**

<Any>

**GRADE LEVEL**

Pre-K  K  1

2  3  4

5  6  7

8  9  10

## Assessments

To effectively understand why assessments are recommended or partially recommended and the process determine this, please review [Determining High-Quality Content Assessment](#) and the [Assessment Review](#). Colorado Academic standards have recommended assessments. To find what assessments were reviewed the gaps are in each content area, go to the Posts tab and choose the category of "Assessment Inventory Reports".

<u>Title</u>	<u>Grade Level</u>	<u>Content Area</u>	<u>Level of Recommendation</u> ▼
<a href="#">Auralia</a>	4, 5, 6, 7, 8, 9, 10, 11, 12	Music	Recommended
<a href="#">Connecticut Arts Assessment: 2nd Grade Solo Improvising</a>	2	Music	Recommended
<a href="#">Connecticut Arts Assessment: 2nd Grade Solo Singing</a>	2	Music	Recommended
<a href="#">Connecticut Arts Assessment: 5th Grade Composing and Self Evaluation</a>	5	Music	Recommended

# New York State Education Department

<http://usny.nysed.gov/rttt/teachers-leaders/assessments/approved-list.html>

**Note that regardless of what the Service Summary states, ONLY the assessment products explicitly listed in the Table below are approved for the purposes of the APPR.**

**UPDATE – 11/26/12:** All reviews have been completed and the lists of approved assessments (CTE and non-CTE) are now finalized. Please see our [RFQ](#) page for information on future Assessment Request for Qualifications for 3rd party assessments to be used in teacher and principal evaluation.

Assessment Name / Vendor Name	Teacher(s) of:	Can this assessment be used for the growth subcomponent?	<b>Service Summary*</b> <small>*Please be sure to check the service summary for detailed information related to each approved assessment.</small>
<b>Explore</b> ACT Inc.	<b>Grades: 8-9</b> ELA, Math, Science	No	<a href="#">Service Summary</a>  (45 KB)
<b>PLAN</b> ACT Inc.	<b>Grade: 10</b> ELA, Math, Science	No	<a href="#">Service Summary</a>  (45 KB)
<b>The ACT</b>	<b>Grades: 11-12</b>		<a href="#">Service Summary</a>

# Examples and Considerations: Student Learning Objectives (SLOs)

---

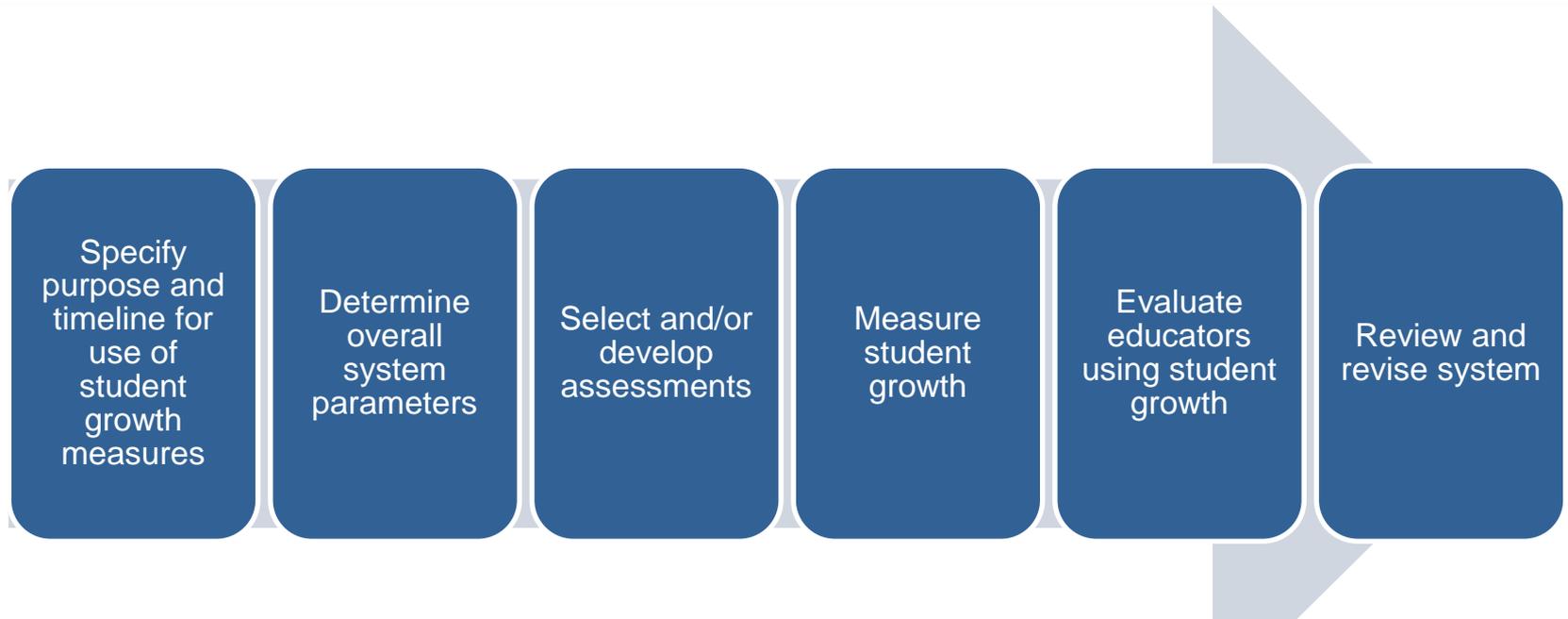
- Method in which teachers set goals for student performance on a particular assessment using baseline data; teachers are evaluated based on how many students meet their targets
- Examples
  - Rhode Island, New York, Ohio
- Considerations
  - SLOs provide opportunities for teacher professional growth and align well to teacher responsibilities
  - It may be difficult to implement this approach consistently across classrooms and schools, and it is labor intensive.

# Using Student Assessment and Other Data: Putting It All Together

- Combining multiple measures
- Two approaches:
  - Matrix (conjunctive approach)
  - Numeric (compensatory approach)
    - $.35 \times \text{student growth score} + .15 \times \text{student survey score} + .50 \times \text{observational score}$
  - Can also blend approaches
- Considerations:
  - Numeric approach appears best in terms of minimizing error and bias

	SG1-High	SG1-Low
SG2-High	*****	***
SG2-Low	***	*

# Final Thoughts: A Process for Designing and Implementing Student Growth Measures



**State role:** What can and will the state do? What will districts, schools, or teachers do?  
**Communication and engagement:** Who will provide input and make decisions at each step? How will information be communicated to stakeholders?

**Mariann Lemke**

202-570-6677

mlemke@air.org

1000 Thomas Jefferson Street NW

Washington, DC 20007-3835

877-322-8700

www.gtlcenter.org

gtlcenter@air.org

▶ *Advancing state efforts to grow, respect, and retain great teachers and leaders for all students*

Center on

**GREAT TEACHERS & LEADERS**

at American Institutes for Research ■

 **AIR**<sup>®</sup>  
AMERICAN INSTITUTES FOR RESEARCH<sup>®</sup>

# Sources and Resources

---

Braun, H. (2005). *Using student progress to evaluate teachers: A primer on value-added models*. Princeton, NJ: Educational Testing Service. Retrieved from <http://www.ets.org/Media/Research/pdf/PICVAM.pdf>

Braun, H., Chudowsky, N., & Koenig, J. (Eds.), for National Research Council & National Academy of Education. (2010). *Getting value out of value-added: Report of a workshop*. Washington, DC: The National Academies Press.

Goldschmidt, P., Roschewski, P., Choi, K., Auty, W., Hebbler, S., Blank, R., et al. (2005). *Policymakers' guide to growth models for school accountability: How do accountability models differ?* Washington, DC: Council of Chief State School Officers.

Harris, D. N. (2011). *Value-added measures in education: What every educator needs to know*. Boston, MA: Harvard Education Press.

McCaffrey, D. F., Koretz, D., Lockwood, J. R., & Hamilton, L. S. (2004). *The promise and peril of using value-added modeling to measure teacher effectiveness* (Research Brief No. RB-9050-EDU). Santa Monica, CA: RAND Corporation. Retrieved from [http://www.rand.org/pubs/research\\_briefs/2005/RAND\\_RB9050.pdf](http://www.rand.org/pubs/research_briefs/2005/RAND_RB9050.pdf)

McCaffrey, D. F., Lockwood, J. R., Koretz, D., & Hamilton, L. S. (2004). *Evaluating value-added models for teacher accountability*. Santa Monica, CA: RAND Corporation. Retrieved from [http://www.rand.org/pubs/monographs/2004/RAND\\_MG158.pdf](http://www.rand.org/pubs/monographs/2004/RAND_MG158.pdf)

Steele, J. L., Hamilton, L. S., & Stecher, B. M. (2010). *Incorporating student performance measures into teacher evaluation systems* (Technical Report No. TR-917-CAP). Santa Monica, CA: RAND Corporation. Retrieved from [http://www.rand.org/pubs/technical\\_reports/TR917.html](http://www.rand.org/pubs/technical_reports/TR917.html)

# Sources and Resources

---

Goldhaber, D. D. & Hansen, M. (2009). *Assessing the potential of using value-added estimates of teacher job performance for making tenure decisions. Working Paper 2009-2*. Seattle, WA: Center on Reinventing Public Education.

Jacob, B. & Lefgren, L. (2008). Can principals identify effective teachers? Evidence on subjective performance evaluation in education. *Journal of Labor Economics*. 26(1), 101–36.

Glazerman, S., Loeb, S., Goldhaber, D., Staiger, D., Raudenbush, S., Whitehurst, R (2010). *Evaluating Teachers: The Important Role of Value-Added*. Washington, DC. The Brookings Brown Center Task Group on Teacher Quality.

Glazerman, S., Goldhaber, D., Loeb, S., Raudenbush, S., Staiger, D., Whitehurst, R., Croft, M. (2011). *Passing Muster: Evaluating Teacher Evaluation Systems*. Washington, DC. The Brookings Brown Center Task Group on Teacher Quality.

Herman, J. L., Heritage, M., & Goldschmidt, P. (2011). *Developing and selecting assessments of student growth for use in teacher evaluation systems* (extended version). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

- [Available at http://www.cse.ucla.edu/products/policy/shortTermGrowthMeasures\\_v6.pdf](http://www.cse.ucla.edu/products/policy/shortTermGrowthMeasures_v6.pdf)
- [Related PowerPoint at http://www.aacompcenter.org/cs/aacc/print/htdocs/aacc/home.htm](http://www.aacompcenter.org/cs/aacc/print/htdocs/aacc/home.htm)

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association. Retrieved from <http://teststandards.org>.

National Council on Measurement in Education. (2013). Resource Center. Retrieved from <http://ncme.org/resource-center/>