Welcome!
Making Connections: Evaluating Teachers of Untested Subjects and Grades
Hella Bel Hadj Amor
Senior Researcher
August 2012
Agenda

- Welcome and Introduction
- Framing the Conversation: Evaluating Teachers of Untested Subjects
- Panel Presentations: Implementing Various Measures
- Wrap Up and Closing Remarks
American Institutes for Research (AIR), through its merger with Learning Point Associates, has operated the Midwest’s regional educational laboratory for more than 25 years.
Our charge is to improve academic outcomes for students by

- Helping states, school districts, and schools systematically use data
- Conducting and supporting high-quality research and evaluation
- Promoting evidence-based decision making
Evaluating Teachers of Untested Subjects and Grades: Measurement Options and Innovations

Lisa Lachlan-Haché, Ed.D.
Senior Research and Policy Associate
August 29, 2012
Why Is This Topic Important?

- States and school districts are searching for new ways to build teacher evaluation systems that include multiple measures.
- Recent reforms have pushed states and school districts to consider student growth as a significant measure.
"Student growth is the change in student achievement for an individual student between two or more points in time."

About 69 percent of all teachers cannot be accurately assessed with models based on test scores, such as value-added models (Prince et al., 2009).
Three Emerging Approaches to Measuring Student Growth

- School wide value-added or student growth percentile scores
- Developing additional assessments to measure student growth
- Student learning objectives
School Wide Value-Added Measures

School wide value-added measures may assess the performance of a school, a grade level, an instructional department, teacher teams, or other groups of teachers.

These measures include a variety of forms, including schoolwide student growth measures, team-based collaborative achievement projects, and shared value-added scores for coteaching situations.
School Wide Value-Added Measures

How and where are collective value-added measures used in practice?

- 2011-12 Tennessee TEAM
- TAP: The System for Teacher and Student Advancement
- 2011-12 Washington DCPS IMPACT
- 2011-12 Delaware DPAS II
School Wide Value-Added Measures

What does the research say about collective value-added measures?

- Tennessee SCORE: Among teachers without individual value-added data who responded to the online questionnaire, 48 percent indicated that one of their top three concerns with the new system was that, “I am not confident that the available value-added student growth data…accurately reflect my effectiveness.”

- Fifty-six percent of principals and other evaluators also cited the accuracy of the value-added student growth data, as a measure of teacher effectiveness, as one of the top three things they think are most in need of improvement with the teacher evaluation system.
### School Wide Value-Added Measures

<table>
<thead>
<tr>
<th>Benefits</th>
<th>Implementation Challenges</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Address the variety of teaching assignments by using agreed-on measures for which schools or groups of teachers share responsibility.</td>
<td>• Measures of school wide performance mask high and low performers in the group and give little information about how individual teachers are doing in their classrooms.</td>
</tr>
<tr>
<td>• Build school wide or team-based effort around student achievement.</td>
<td>• May be perceived as unfair because teachers are held to a measure that they may have had limited ability to impact.</td>
</tr>
<tr>
<td></td>
<td>• May cause delayed evaluation scores, in some cases reducing the timeliness of decision making.</td>
</tr>
</tbody>
</table>
Developing Additional Assessments

Other measures of student growth for untested grades and subjects, such as end-of-course assessments, can be developed or adapted for use across schools or school districts. These measures may include early reading measures; standardized end-of-course assessments; formative assessments; benchmark, interim, or unit assessments; and standardized measures of English language proficiency.
Developing Additional Assessments

How and where are additional assessments used in practice?

- Colorado Content Collaboratives
- Delaware Department of Education
- Empowering Effective Teachers (Hillsborough County, Florida)
Developing Additional Assessments

What does the research say about the development of additional assessments?

- Additional assessments can be used to develop growth scores when used in pre-test/post-test, value-added, or student growth percentile models, which can provide useful evidence despite limitations.

- Lessons learned:
  - time and resource intensive
  - sharing the distribution of scores can support teacher reflection, projection and learning
  - scoring schedule can be problematic, but supported with multiple years of data
## Developing Additional Assessments

<table>
<thead>
<tr>
<th>Benefits</th>
<th>Implementation Challenges</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Create comparability within tested fields of study and can create similar rigor through multiple classrooms and schools.</td>
<td>• Time and resource intensive</td>
</tr>
<tr>
<td>• A relatively small number can be developed for most commonly taken courses.</td>
<td>• Current assessments are designed for purposes other than assessing student growth and/or as a measure of teacher effectiveness.</td>
</tr>
<tr>
<td>• May be readily adapted in some fields.</td>
<td>• Assessments may require time for teachers to work together to develop consistent scoring patterns.</td>
</tr>
<tr>
<td>• Can increase teacher buy-in and professional growth.</td>
<td>• Requires attention to ensuring security and comparability across classrooms.</td>
</tr>
</tbody>
</table>
SLOs are a set of goals that measure an educators’ progress in achieving student growth targets.

- Educators or educator teams establish learning growth targets for groups of students based on available data, monitor student growth toward the targets, and, at the end of an agreed-upon time period, determine the degree to which students meet the targets.
- Evaluators support this work by approving the SLO, engaging in mid-year reviews, reviewing and scoring the final results, and discussing lessons learned with educators.
## Student Learning Objectives

How and where are SLOs used in practice?

<table>
<thead>
<tr>
<th>District/State</th>
<th>Teachers Required to Use SLOs</th>
<th>How SLO Data Are Used?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Austin, Texas</td>
<td>All teachers</td>
<td>Compensation</td>
</tr>
<tr>
<td>Charlotte-Mecklenburg, North Carolina</td>
<td>All teachers</td>
<td>Compensation</td>
</tr>
<tr>
<td>Denver, Colorado</td>
<td>All teachers</td>
<td>Compensation</td>
</tr>
<tr>
<td>Georgia</td>
<td>Teachers in Nontested Grades and Subjects</td>
<td>Evaluation</td>
</tr>
<tr>
<td>Houston, Texas</td>
<td>All teachers</td>
<td>Evaluation</td>
</tr>
<tr>
<td>Indiana</td>
<td>All teachers in the default model</td>
<td>Evaluation</td>
</tr>
<tr>
<td>New Haven, Connecticut</td>
<td>Teachers in Nontested Grades and Subjects</td>
<td>Evaluation</td>
</tr>
<tr>
<td>New York</td>
<td>Teachers in Nontested Grades and Subjects</td>
<td>Evaluation</td>
</tr>
<tr>
<td>Rhode Island</td>
<td>All teachers</td>
<td>Evaluation</td>
</tr>
<tr>
<td>Washington, DC: DCPS</td>
<td>Teachers in Nontested Grades and Subjects</td>
<td>Evaluation</td>
</tr>
</tbody>
</table>
What does the research say?

- Elements of SLOs align with and promote best practices in planning for instruction, goal setting, reflection and assessment.
- High-quality objectives have been linked to their attainment by the teachers’ own measure(s) as well as by student achievement measured on independent, standardized state and national assessments (CTAC, 2008).
- Pre and post tests within SLO administration have challenges including corruptibility, lack of clear conceptualization about the nature of the pre-test, inflated fall-to-spring gains, weak analytic approaches, and treated non-equated tests are if they shared the same score (Marion, 2012).
Student Learning Objectives

Lessons learned:

• Identify quality assessments or build district capacity in assessment literacy
• Provide good SLO examples and guidance to support teachers in setting rigorous but realistic targets
• Provide training to calibrate SLO raters and link SLO results with PD opportunities
• Help administrators and teachers find time for the SLO cycle
• Clarify expectations with consistent communication
## Student Learning Objectives

<table>
<thead>
<tr>
<th>Benefits</th>
<th>Implementation Challenges</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Reinforce best teaching practices.</td>
<td>• Difficult to create comparability and rigor without common assessments or common requirements for assessment.</td>
</tr>
<tr>
<td>• Can be adapted to new assessment structures as they are developed.</td>
<td>• Predictive validation of SLOs in alignment with growth measured by a value-added or student growth measure has been completed on a limited scale only.</td>
</tr>
<tr>
<td>• Acknowledge the value of educator knowledge and skill.</td>
<td>• Requires significant time and attention from administrators and evaluators.</td>
</tr>
<tr>
<td>• Encourage collaboration.</td>
<td></td>
</tr>
<tr>
<td>• Can have face validity as teachers are often tasked with developing the objectives.</td>
<td></td>
</tr>
</tbody>
</table>
Key Considerations

• No measure or method is perfect; continuous improvement is the name of the game.
• On-going effective training around the use of any measure or process is necessary.
• Further research and evaluation in these early stages is critical to refining and comparing approaches.
• Ultimately, all methods should be used to improve performance and support professional growth.
  • Evaluators and teachers need supports in using these data and having meaningful conversations to reflect on lessons learned and establish next steps for professional growth.
References


Marion, S. el al (2012). *Considerations for Analyzing Educators’ Contributions to Student Learning in Non-tested Grades and Subjects with a Focus on Student Learning Objectives*. Center for Assessment.
Lisa Lachlan-Haché
P: 202-403-6214
E-mail: llachlan@air.org
1000 Thomas Jefferson Street NW
Washington, DC 20007-3835
Phone: 877-322-8700 or 202-223-6690
Website: www.tqsource.org
AISD
REACH
Supporting & Rewarding Success in the Classroom
Student Learning Objectives

✧ Content
  ✧ Two Objectives – Individual and Team
  ✧ Based on data & Campus Improvement Plan
  ✧ Targets of student growth
  ✧ Collaborative

✧ Rationale
  ✧ Empower Teachers
  ✧ Align Campus and District Expectations
SLO Design

✧ Phase 1
  ✧ 2 Individual SLOs
  ✧ Pre and Post

✧ Phase 2
  ✧ Tiered Targets

✧ Phase 3
  ✧ Individual
  ✧ Team
  ✧ Campus
Student Learning Objectives

✧ Process and Systems
  ✧ Dedicated Staff of 3
    ✧ Visit Schools and provide guidance
  ✧ Needs Assessment
    ✧ AIMS/Schoolnet
    ✧ Observation/Anecdotal Evidence
✧ Learning Content
  ✧ State Standards vs District CRMs
✧ Learning Objective
Student Learning Objectives

✧ Process and Systems
  ✧ Assessment
    ✧ Item Bank, Resources
  ✧ Growth Target
    ✧ Formula - (100-Pre-test)/2
✧ Approval
  ✧ Web-based Database
  ✧ Principal Approval
  ✧ SLO Team Approval
Student Learning Objectives

✧ Process and Systems
  ✧ Rating – 2 of SLO team must agree
  ✧ Sample versus Whole
✧ End of Year
  ✧ Verification Process
  ✧ Audit
Pilot Evaluation System

![Pie chart showing the distribution of evaluation components:]

- Professional Expectations: 10%
- Individual SLO: 20%
- Team SLO: 10%
- School-wide Measure: 10%
- Peer Observation (2): 15%
- 2 Administrator Walkthroughs: 10%
- 1 Administrator Observation: 15%

**Teachers**
- **Student Growth**
  - Individual SLO: 20%
  - Team SLO: 10%
  - School-wide Measure: 10%
  - *Observations of Instructional Practice and Classroom Climate*
  - 1 Administrator Observation: 15%
  - 2 Administrator Walkthroughs: 10%
  - Peer Observation (2): 15%

- **Student Input**
  - Student Survey: 10%

**Professional Expectations**: 10%
Successes

✧ Campus Based Training for Teachers and Principals
✧ Inclusion of All Teachers
✧ Teacher’s Professional Growth
✧ Reflective Practice
✧ Inclusion of performance measures in evaluation
Challenges

✧ Limited assessment tools
✧ Data and collection systems
✧ Embedding practice into campus culture
✧ Common assessment versus teacher developed assessment
✧ Concern over SLOs in the evaluation and fidelity
How much do teachers improve with experience?

This graph replicates one of the most memorable graphs from the original Human Capital Diagnostic. It shows the return to experience of math teachers in grades 4-8 in terms of impact on student achievement. Using novices as a baseline (those whose years of experience = 0 at the start of their first school year), the graph traces the improvement in value-added as a teacher gains experience. As much of the national research confirms, our teachers develop rapidly in their first 3-4 years. A teacher in her second year provides approximately 1-3/4 months additional instruction. Note that on average teachers do not develop as much after their third year of experience. The returns to experience in other tested subjects is similar.

Teachers who enter the year with three years of experience provide just under 3 months more instruction than they did as novice teachers. On average, this is the peak of a teacher’s performance.

In 2011, 26% of our first-year teachers left the classroom at the end of the year. 29% of second-year teachers left.
Many Views on Teaching

- Community
- Teacher
- Principal
- Master/Mentor Teacher
- Student
- Peer Teacher
<table>
<thead>
<tr>
<th>Novice Teacher</th>
<th>Struggling Teacher</th>
<th>Veteran/Expert Teacher</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regular visits to inform instructional practice</td>
<td>Targeted analysis of what works and what doesn’t</td>
<td>Administrative Load with Classroom Responsibilities</td>
</tr>
<tr>
<td>Classroom Management Strategies</td>
<td>Monitored progress toward improvement</td>
<td>Career Path Opportunities</td>
</tr>
<tr>
<td>Job-embedded, foundational PD</td>
<td>Diagnostic, skill-specific PD</td>
<td>Specialized, high-level PD</td>
</tr>
</tbody>
</table>
How Do We Define Measures of Effectiveness?

- Impact on Students
- Differentiates Teacher Performance
- Includes a Development Component
Who Will Measure?

- **Teacher**
  - Peer Consultation
  - PLC
  - SLO
  - Peer
  - Teacher Work Product

- **Principal**
  - Formal Observation
  - Teacher Conference

- **Master/Mentor Teacher**
  - Content Pedagogy
  - Classroom Environment
  - Teacher Work Product

- **Community**
  - Essential but Currently Lack Capacity to Collect Data

- **Student**
  - Value-Added
  - Student Surveys
  - Hard to Staff

- **Teacher**
Guiding Principles

To capture the great teaching that occurs in our classrooms, without...

- Added Work for Teachers
- Compliance Orientation
- Reliance on a Single Perspective
Test Development
Hillsborough County Public Schools

Michelle R. Watts, Ph.D.
Supervisor of Assessment
Demographics

• Hillsborough County Public School district is the 8th largest school district in the United States.
• We have over 190,000 students and 250 school locations.
• All of our courses, K-12, are measured with a semester examination, developed by the district.
Test Development History

• We have had semester examinations for high school courses for approximately 25 years.
• The examinations were used to maintain curriculum consistency across the large district.
• The examinations counted for 33% of the student’s semester grades.
• The examination results were used for teacher inservice training and curriculum evaluation.
Test Development History

• Our system is primarily a paper and pencil system (40,000 students in high school take 7 examinations in 4 days.)
• Student answer sheets are scanned into a web-based testing system.
• We are beginning to administer more tests online. For the 11-12 school year, we administered music, art, and pe exams to 15,000 students in each of grades 1-5.
Teacher Evaluation

• During the 2005-06 school year, the district received 10 million dollars from the state of Florida to pay the top 25% of teachers a bonus based on student achievement. This program continued for 5 years.

• We then created tests for all content areas including music, art and physical education for elementary students.

• We no longer award teachers bonuses. Student achievement is now included in our teacher evaluation process.
Test Development Process

• Test development is a shared responsibility between the assessment and curriculum staff.
• Curriculum is responsible for selecting teams of teachers to work on district examinations and for the content measured each semester.
• Assessment is responsible for training teachers on item writing and test development.
Test Development Process

• Develop a test blueprint – negotiated with current teachers and the content supervisor
• Develop test items
• Select items for the test
• Review the test
• Administer the test
• Revise the test
Test Development Process

• Current teachers of the course are selected to develop test items and the test blueprint.
• This process increases the content validity of the examination.
• We are building capacity in our teachers by training them to create test items.
• The number of the teachers selected varies from 2-6 depending on the number of students and teachers for each individual course.
Test Development Process

- The assessment staff review and revise all items.
- The teachers know the course content but have difficulty writing items that work well.
- After the test has been administered, we use the item statistics and overall test performance to select items for revision.
Challenges

• Writing items that are above the knowledge level – separating item difficulty and item cognitive levels

• Estimating the ability of students - the tests tend to be more difficult than desired

• Maintaining content valid tests as curriculum is changed or modified
Challenges

• Revising items to improve the test quality
• Maintaining test security and investigating improper test administration
• Distribution of current exams – administration of old archived examinations
• Scaling the examinations for grading within the 6 day deadlines
Challenges with Teacher Evaluation Process

- Understanding the value add calculations
- Maintaining correct teacher/student/course links
- Scaling the value add calculations across 700 examinations
- Course/tests with low student enrollment
Hella Bel Hadj Amor

P: 202-403-5176 > F: 202-403-5444
E-Mail: hbelhadjamor@air.org

REL Midwest
1120 East Diehl Road, Suite 200
Naperville, IL 60563-1486

General Information: 866-730-6735
Website: www.relmidwest.org