Using Student Assessment Data in Teacher Evaluations

Mariann Lemke
October 2016
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.
Agenda

- State Assessment Data and Teacher Evaluations
  - Who, What, Why?
- What We Know: Value-Added Measures (VAMs)
- What We Know: Student Learning Objectives (SLOs)
- Other Growth Measures
- Q&A and Discussion
Policy scans show that states typically use state assessments in teacher evaluation systems to measure student growth through:

- Student learning objectives (SLOs) or similar goal-setting methods
- Statistical measures such as value-added models (VAMs) or growth models
Measuring Student Growth

- **Student growth** means the change in student achievement for an individual student between two or more points in time.

- **What you need:**
  - Data from two points or more points in time
  - Data that measure what students are supposed to have learned and teachers are supposed to have taught by that point in time
  - An approach to connecting the data
Why Student Growth?

Instruction → Student Learning
Why Student Growth?

Ms. Smith

Ms. Jones

Proficiency
Sample Value-Added Model
A student learning objective (SLO) is a measurable, long-term, academic goal informed by available data that a teacher or teacher team sets at the beginning of the year for all students or for subgroups of students.
Use of Value-Added Measures
The Every Student Succeeds Act (ESSA) specifies that the U.S. Department of Education (ED) cannot require anything of states with regard to educator evaluation (i.e., as condition of approval of waiver to required state plans).

Regardless of federal policy changes, state laws and regulations related to educator effectiveness remain in effect.
Current State Policy

- **43 states required objective measures of student achievement to be included in teacher evaluations.**
- **16 states included student achievement and growth as the “preponderant criterion” in teacher evaluations.** These states include AK, CO, CT, DC, DE, GA, HI, KY, LA, MS, NC, NM, NY, OK, PA and TN.
- **19 states included growth measures as a “significant criterion” in teacher evaluations.** Eleven of those states (AZ, FL, ID, IL, MI, MN, NJ, NV, OH, RI, VA) explicitly define what “significant” means for the purposes of including student achievement in teacher evaluations. Eight states (AR, IN, KS, MD, ME, MO, OR, SD) do not provide these explicit guidelines.
- **Eight states required objective evidence of student learning in teacher evaluations.** (MA, ND, SC, UT, WA, WI, WV, WY).
- **Seven states required that schoolwide achievement data be used in individual teacher performance ratings,** whereas 11 other states explicitly allowed the practice.
States Currently Using Value-Added Data in Teacher Evaluations

- **Florida:**
  - 50% of educator evaluation score

- **Louisiana:**
  - 50% of educator evaluation score

- **Minnesota:**
  - 35% of educator evaluation score

- **Ohio:**
  - Under 2016 statue, value-added score is optional

- **New Mexico:**
  - 50% of educator evaluation score

- **North Carolina:**
  - One of six standards

- **Tennessee:**
  - One of several approved measures used in portfolio
But…a Changing Landscape….

- Mass. reexamining role of student test scores in teacher evaluations
  (*Boston Globe*, September 27, 2016)

- N.J. Triples Weight of Tests in Teacher Evaluations
  (*Education Week*, September 13, 2016)
What We Know: Value-Added Models
Technical Characteristics

- **Bias/validity is a concern**: Potential exists, but studies suggest no concrete evidence of it.
- **Precision/reliability**: Documented variability in measures; can increase by pooling data over time.
- **Relationship to other measures**: Moderate relationships with other measures of teacher effectiveness.
- **Data needs**: Need high-quality assessment and linkage data.
Effects of the Use of Value-Added Measures

- Typically not used as an independent measure.
- Effects likely to vary depending on who uses data, their perceptions of data, and the specifics of policies—are measures to be used for accountability or improvement, and what are the specifics of supports or consequences?
  - Some evidence of teacher turnover, increased teacher performance, increased student performance with use of evaluation systems
  - May be useful for assignment or identifying coaches
What We Know: Student Learning Objectives
Implementation, Perceptions, and Relationship to Achievement

- Mixed perceptions of usefulness of SLOs; some evidence this may change over time and positive perceptions of use of data.

- Challenges related to assessments (selection, design) and accessing data, as well as with supports and communication.

- Current studies suggest some relationship between SLO quality and student achievement, but results are not consistent across content areas or studies.
Other Growth Measures
Pre-Test/Post-Test or Simple Growth

- Measuring growth = subtraction
- Scores must be on the same scale (could be a rubric); content must be aligned between time periods; need a reference point to interpret results
- Examples:
  - Math grade 4 and math grade 5 on a vertically scaled assessment
  - Spanish 3 pre- and post-test
  - Fitness pre- and post-test
Pre-Test/Post-Test With a Rubric

<table>
<thead>
<tr>
<th>Pre-Classification</th>
<th>Post-Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic</td>
<td>Basic</td>
</tr>
<tr>
<td>Proficient</td>
<td>Proficient</td>
</tr>
<tr>
<td>Advanced</td>
<td>Advanced</td>
</tr>
</tbody>
</table>

- Measuring growth = change in performance level
- Scores must be on the same scale; content must be aligned between time periods; need a reference point to interpret results
- Examples:
  - Writing, other performance-based content or skills
Another Approach: Portfolios

- Collection of student work showing growth related to relevant standards
  - May work especially well for courses with performance-based tasks or work that is scored via rubric
  - Requires a holistic rubric or repeated measures (e.g., writing assignments scored against the same rubric each time)
  - Must consider means to ensure consistency, quality
  - Could do within a single class or course
Q&A and Discussion
Reflection Activity
References and Resources


References and Resources


References and Resources


References and Resources


References and Resources


References and Resources


References and Resources


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