Summary

U.S. Department of Education
September 2013

Using alternative student growth measures for evaluating teacher performance: what the literature says

Brian Gill
Mathematica Policy Research

Julie Bruch
Mathematica Policy Research

Kevin Booker
Mathematica Policy Research

Key findings

States and school districts are exploring alternatives to state tests for measuring teachers' contributions to student learning. One approach applies statistical value-added methods to alternative student assessments such as commercially available tests and end-of-course tests. The evidence suggests that these methods can reliably distinguish among teachers. A second approach requires teachers to develop student learning objectives at the beginning of the school year; these can be used in instructional planning as well as evaluation. Ensuring consistency across teachers and schools is challenging, and implementation is demanding, but student learning objectives have the advantage that they can be implemented in any grade or subject.

ies NATIONAL CENTER FOR EDUCATION EVALUATION AND REGIONAL ASSISTANCE
Institute of Education Sciences

REL MID-ATLANTIC
Regional Educational Laboratory
At ICF International

REL 2013–002

The National Center for Education Evaluation and Regional Assistance (NCEE) conducts unbiased large-scale evaluations of education programs and practices supported by federal funds; provides research-based technical assistance to educators and policymakers; and supports the synthesis and the widespread dissemination of the results of research and evaluation throughout the United States.

September 2013

This report was prepared for the Institute of Education Sciences (IES) under Contract ED-IES-12-C-0006 by Regional Educational Laboratory Mid-Atlantic administered by ICF International. The content of the publication does not necessarily reflect the views or policies of IES or the U.S. Department of Education nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

Gill, B., Bruch, J, & Booker, K. (2013). *Using alternative student growth measures for evaluating teacher performance: what the literature says.* (REL 2013–002). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Mid-Atlantic. Retrieved from http://ies.ed.gov/ncee/edlabs.

This report is available on the Regional Educational Laboratory website at http://ies.ed.gov/ncee/edlabs.

Summary

States are increasingly interested in including measures of student achievement growth, or "value-added," in evaluating teachers. But annual state assessments, which are the typical measure of student growth, usually cover only reading and math teachers and only in grades 4–8. These state assessments thus cannot generally be used to measure contributions to student achievement growth for early elementary school teachers, most high school teachers, and teachers of other subjects.

As a consequence, a growing number of states and school districts are exploring alternatives for measuring teachers' contributions to student learning. These alternatives have the potential to be used for evaluating not only teachers who work in grades and subjects outside the annual state testing regime but also as complementary growth measures for teachers of tested grades and subjects.

This report reviews the literature on two categories of alternative measures for evaluating teachers:

- Alternative student outcome measures used in statistical growth (or value-added) models.
- Teacher-developed student learning objectives used for measuring growth.

Using alternative student outcome measures in statistical growth models

This literature review of studies of statistical growth models using alternative assessments (such as commercially available assessments like the Stanford Achievement Test and locally developed end-of-course exams) and other outcomes (such as student attendance) looked for evidence of the statistical properties of such measures. Despite differences in the student outcome measure, the statistical method used in the growth models to assess teacher value-added is similar to that used in state reading or math assessments. Key findings for growth/value-added models show that:

- Models based on widely used, commercially available assessments generally produce measures of teacher performance that correlate positively with other performance measures, such as teacher observations and student surveys. All the reviewed studies found positive relationships, with correlations up to 0.5.
- Models based on commercially available assessments yield results that are as stable over time as do models based on state assessments. Year-to-year correlations of teacher value-added based on commercially available assessments are positive but modest—consistent with year-to-year correlations for value-added measures based on state assessments. This finding suggests that growth models using these alternative measures—similar to those using state assessments—can be useful for teacher evaluation if applied judiciously. States and districts may want to use measures that average across several years of teaching or apply Bayesian "shrinkage" adjustments to reduce the likelihood that random error will mistakenly identify teachers as low performing.
- Little is known about growth/value-added models based on locally developed, curriculum-based assessments or nontest outcomes, but the available evidence suggests that they have the potential to reliably differentiate performance among teachers and schools. Just two studies were identified that examined the potential for using locally developed assessments to evaluate teacher performance, and both

examined the same district. The results suggest that such measures can reliably distinguish among teachers at the ends of the performance distribution. The same studies found that measuring schoolwide value-added using nontest outcomes (like attendance and course completion) can produce results that reliably distinguish school-level performance—but the studies did not analyze nontest outcomes at the teacher level.

More research is needed to inform the decisions of states and districts as they expand growth models to teachers and content not covered in state and commercially available assessments.

Measuring student growth using student learning objectives

Student learning objectives (SLOs)—classroom-specific growth targets chosen by individual teachers and approved by principals—are becoming popular as alternative measures of student growth because they can be used to evaluate teachers in any grade or subject. Although very little of the literature on SLOs addresses their statistical properties, key findings show that:

- SLOs have the potential to better distinguish teachers based on performance than traditional evaluation metrics do, but no studies have looked at SLO reliability. Most of the limited evidence on the statistical properties of SLOs is on the proportion of teachers achieving SLO objectives. Whether that differentiation represents true differences in teacher performance or random statistical noise is unknown.
- Little is known about whether SLOs can yield ratings that correlate with other measures of teacher performance. Only three studies have explored the relationship between SLO ratings and standardized assessment-based (value-added) growth measures. These studies found small but positive correlations. More research is needed as states and districts roll out SLOs as teacher evaluation measures and instructional planning tools.
- Until some of the research gaps are filled, districts that intend to use SLOs may want to roll them out for instructional planning before using them in high-stakes teacher evaluations. Several studies found teacher concerns about fairness in SLO implementation. This is no surprise, because SLOs are difficult to make valid and reliable. They are by definition customized to individual teachers and based on the professional judgments of teachers and principals. Making SLOs an important component of high-stakes evaluation could undermine their validity, because it means that teachers are in essence grading themselves.
- Studies of teacher experiences with SLOs indicate that SLOs can require substantial training and technology infrastructure and that they can be time-consuming for teachers and evaluators alike.