

# Alternative student growth measures for teacher evaluation: Profiles of early-adopting districts

Brian Gill  
Brittany English  
Joshua Furgeson  
Moira McCullough  
Mathematica Policy Research



REL 2014–016

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.

April 2014

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.

This REL report is in the public domain. While permission to reprint this publication is not necessary, it should be cited as:

Gill, B., English, B., Furgeson, J., & McCullough, M. (2014). *Alternative student growth measures for teacher evaluation: Profiles of early-adopting districts*. (REL 2014–016). 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

Many states and districts have begun to evaluate teacher performance and reward teachers based on their students' growth on state assessments, as measured by statistical techniques known as value-added models or student growth models. However, the state assessments necessary to evaluate most teachers are typically administered only in grades 3–8 and only in math and reading. To measure student achievement growth for teachers in all grades and subjects, a growing number of states and school districts are developing alternative student growth measures that do not depend on the state assessments that are typically used for school accountability under the federal Elementary and Secondary Education Act (ESEA). These alternative growth measures come in two forms: alternative assessment-based value-added models (VAMs) that use the results of end-of-course assessments or commercially available tests in statistical growth models, and student learning objectives (SLOs), which are determined by individual teachers, approved by principals, and used in evaluations that do not involve sophisticated statistical modeling.

Use of alternative growth measures that do not depend on state assessments is recent, and little is known about their validity and reliability or about how they are being used. This study describes how eight early adopting school districts are using these two types of alternative measures for assessing teacher effectiveness. The study aims to provide key pieces of information about the districts' experiences that can be used by other states and districts to decide whether and how to implement alternative assessment-based value-added models, or SLOs.

### Key findings

- *Selecting alternative assessments and implementing VAMs.* All four of the early-adopting sample districts that are applying VAMs to alternative outcomes are using locally developed or state-mandated end-of-course assessments. One of the districts uses commercial tests as well. Each district works with an outside provider to develop and implement its VAM, and all four districts apply the same VAM approach to alternative assessments that they are using for state assessments. Teachers in the four districts using alternative assessment-based VAMs receive written performance feedback about prior-year performance each fall, along with training to interpret the feedback.
- *Maximizing consistency in implementing SLOs.* In all four SLO early-adopting districts included in the sample, SLOs are required of all teachers across grades K–12, regardless of whether the teachers serve grades or subjects that include districtwide standardized tests. The SLO process is similar in each district. During the first months of school, each teacher identifies strengths and weaknesses of students. Teachers set learning goals aligned with standards; two sample districts require teachers to identify the instructional strategies to be used. Each teacher submits assessment goals to the principal, who has responsibility for reviewing and approving the SLOs. To promote consistency, three of the sample districts also require that SLOs be approved by the district office. Spring assessments—selected at the beginning of the year by the teacher—are used to measure growth. Teachers in the SLO sample districts receive feedback at the end of the school year or the following summer.

- *Selecting and designing alternative growth measures.* Districts using SLOs chose them as a teacher-guided method of assessing student growth, while those using alternative assessment-based VAMs were motivated to use VAMs partly to take advantage of existing assessments. Each of the eight sample districts received financial support from external funders to develop and implement the alternative measures, often for use in performance-based teacher compensation systems. SLOs require substantial teacher, principal, and district participation. Three of the four SLO districts formed design teams composed of teachers—sometimes represented by teachers union officials—and district staff to collaboratively develop the SLO approach. Districts implementing alternative assessment-based VAMs did not usually require a comprehensive design process because the assessments were already in place.
- *Identifying commonly reported uses of alternative growth measures.* All eight sample districts reported multiple uses for the measures, most commonly including performance-based compensation, teacher evaluation, and professional development. One district using alternative assessment-based VAMs and one district using SLOs are already implementing the measures for formal teacher evaluation, and three other sample districts are piloting alternative assessment-based VAMs to prepare for state mandates. SLOs are used not only for measuring teacher performance but also to help teachers plan instruction.

## Contents

Summary i

Why this study? 1

What the study examined 2

**How districts implement alternative assessment-based value-added models 4**

Locally developed or state-mandated end-of-course assessments are being used by all four sample districts using alternative assessment-based value-added models while one district also uses commercially available tests 4

The general approach to calculating value-added is similar across districts 6

**How districts implement student learning objectives 6**

Teachers of all grades and subjects create student learning objectives 7

Districts provide training, support, and guidelines for preparing student learning objectives 8

All districts created systems to promote consistency and rigor in student learning objectives 9

**How districts chose and designed the alternative growth measures 9**

All districts received external funding that supported the design process 9

Design and implementation required close collaboration between administrators and teachers for student learning objectives but not necessarily for alternative assessment-based value-added models 10

**How districts provide feedback and use the performance ratings 10**

Principals generally provided earlier performance feedback to teachers in districts using student learning objectives than in districts using value-added models 11

In addition to evaluation, districts use performance ratings for teacher compensation and professional development 11

Implications and next steps 13

Appendix A. District administrator interview protocol A-1

Notes Notes-1

References Ref-1

### Boxes

1 Definitions of alternative student outcome measures and approaches to measuring teacher performance 1

2 Sample selection, data collection, and analysis 3

3 What the literature says about alternative assessment-based value-added models 5

4 Background and relevant literature on student learning objectives (SLOs) 7

### Tables

1 Districts and alternative growth measures used 3

2	Coverage of end-of-course assessments in sample districts using alternative assessment-based value-added models, by subject, grade, and district	5
3	Coverage of high school end-of-course assessments in sample districts using alternative assessment-based value-added models, by subject and district	6
4	Intended uses of alternative measures, by measure and district	12

## Why this study?

Many states and districts are beginning to use student achievement growth as measured by state assessments (often using statistical techniques known as value-added models or student growth models) alongside classroom observations to evaluate and reward teachers. The U.S. Department of Education has used Race to the Top and the Elementary and Secondary Education Act (ESEA) waiver process to encourage states to incorporate student achievement growth into teacher evaluation, with the aim of raising professional standards and focusing evaluation on improvements in student achievement. But this approach has limited application in most states, because their assessments are typically administered only in grades 3–8 and only in math and reading, so student growth measures based on state assessments cannot be used for most teachers. Value-added models (VAMs) using state assessments can often be estimated for less than 20 percent of teachers (Whitehurst, 2013).

*Many states and districts are beginning to use student achievement growth alongside classroom observations to evaluate and reward teachers*

Many states and school districts are developing alternative student growth measures that can be applied to teachers in more grades and subjects (see additional information on these measures in box 1). One type of measure applies statistical methods in a VAM or student growth model using other standardized assessments, such as end-of-course assessments or commercially available tests. Another type uses student learning objectives (SLOs), which are customized goals set by each teacher at the beginning of the school year with the approval of the principal. SLOs are based on the aims of the course and the abilities of students. Thus, the selection of the specific targets can differ across classrooms, schools, and districts. SLOs might use standardized assessments, but they do not involve statistical modeling. SLOs are becoming more popular in part because they can be used with teachers of any grade and subject. (For a review of the literature on the reliability, validity, and discrimination of alternative growth measures, see Gill, Bruch, and Booker, 2013, and boxes 3 and 4 later in this report.)

---

### Box 1. Definitions of alternative student outcome measures and approaches to measuring teacher performance

**End-of-course assessments.** Assessments designed by states or districts to assess student achievement on the content covered in a course.

**Widely used commercial tests.** Standardized, nationally normed assessments such as the Stanford Achievement Test and the Iowa Tests of Basic Skills.

**Value-added measure (or model).** A statistical approach that estimates a teacher’s impact on student achievement by controlling for observed student characteristics and students’ prior achievement. The achievement measures in value-added models can be state standardized tests, end-of-course assessments, or widely used commercial tests.

**Student growth models.** Approaches that can be used for purposes similar to those of value-added measures, despite conceptual differences (such as the student growth percentile method). This report views student growth models and value-added models as equivalent.

**Student learning objectives.** Classroom-specific achievement growth targets set by individual teachers and approved by principals or other administrators.

---

Implementation of alternative growth measures is recent, and little is known about how they are used. To learn more, this study looked at their use to measure teacher effectiveness in eight early adopting school districts. Because all five states in the Regional Educational Laboratory Mid-Atlantic Region have committed to including student growth measures in teacher evaluation in return for ESEA flexibility, this study addresses a high-priority issue in the region.

### **What the study examined**

This study examined two types of alternative growth measures—alternative assessment-based VAMs, and SLOs—and how they are being used to measure teacher performance in eight early adopting school districts. More specifically, the study addressed the following research questions:

- For alternative assessment-based VAMs, what student outcomes have districts begun to use to include teachers in other subjects and grades than reading and math in grades 3–8?
- How are districts calculating teacher value-added based on alternative assessments?
- Which teachers are districts evaluating using SLOs?
- What procedures have districts established to enable teachers to create SLOs, to assess their completion, and to promote their rigor and consistency?
- How and why did districts select or develop the alternative growth measures?
- How and when do districts provide teachers with performance feedback?
- In what ways do districts use alternative student growth measures?

***This study looked at the use of alternative growth measures to measure teacher effectiveness in eight early adopting school districts***

This report is based on interviews with one key administrator in each of the eight districts, supplemented by document reviews (for further description of the research methods, see box 2). The report describes the different measures used by the sample of districts within the two categories of VAMs and SLOs and details the scope of the teacher populations the measures are used to evaluate. The report also chronicles the motivations behind the selection and development of the specific measures implemented and discusses their uses in teacher evaluation.

The alternative measures used by the eight districts are classified as VAMs or SLOs throughout this report. VAMs (or student growth models) that use end-of-course assessments and VAMs that use widely available commercial tests may differ in how well they align with a district’s curriculum, but they are similar in two important ways. First, they are standardized, meaning that they are used and scored consistently for students in a particular grade and subject districtwide. Second, the teacher’s contribution to student growth is measured using a sophisticated statistical model. Consequently, the findings for both types of VAMs are reported together as *alternative assessment-based VAMs*. By contrast, SLOs only sometimes involve assessments that are consistent districtwide, and they do not use statistical methods to identify a teacher’s contribution to student growth.

In this report, four districts use alternative assessment-based VAMs, and the other four use SLOs. Additional information about each district is provided in table 1. Although District A was identified as an early adopter of SLOs, the district stopped using SLOs during this study. The district has been included in the findings because it had substantial experience with SLOs and because its reasons for terminating the use of SLOs could be instructive.

---

## Box 2. Sample selection, data collection, and analysis

**Sample selection.** A recent literature review (Gill et al., 2013) and the National Council on Teacher Quality online database were used to identify an initial pool of 13 districts using alternative student growth measures to evaluate teacher effectiveness. The sample was not limited to districts in the Mid-Atlantic Region because so few of them have much experience with alternative growth measures and because the issues are not unique to the region. District websites were examined to identify the measures in use, the first year of implementation, and the purpose of the measure.

Potential sample districts were grouped by the type of alternative measure being used to ensure the inclusion of at least one district using each measure and then ranked within each type based on the year the measure was first implemented; districts with the longest period of implementation within each type of measure were prioritized. Rather than being selected as a representative sample, districts were selected for length of implementation because they were more likely to have progressed beyond a pilot stage and would have better insight into the medium- and long-term benefits and challenges of the measure. Four of the districts implementing value-added measures (VAMs) and four implementing student learning objectives (SLOs) were identified as having enough implementation experience and agreed to participate. One administrator from each district was interviewed.

While the sampling was limited, supplemental document reviews and cross-site analyses, together with use of an interview protocol (see appendix), allowed for some degree of triangulation to enhance the validity and reliability of the findings (Merriam, 2001; Wiersma, 2000).

**Data collection and analysis.** The interviews, conducted by phone, lasted about an hour and were recorded with the permission of the administrators. The interviewers used the recordings to create district-level summaries, which also incorporated findings from documents collected on district websites and from district administrators, such as evaluation reports, teacher handbooks, and lists of frequently asked questions. These summaries were analyzed to identify variations and commonalities across districts along the dimensions of implementation raised by the research questions.

*Districts with the longest period of implementation within each type of measure were prioritized because they would have better insight into the medium- and long-term benefits and challenges*

---

**Table 1. Districts and alternative growth measures used**

District	Measure	Years of implementation	District size
A	Student learning objectives	4	Large (> 50,000 students)
B	Student learning objectives	5	Large
C	Student learning objectives	4	Medium (10,000–50,000 students)
D	Student learning objectives	> 5	Large
E	Alternative assessment–based value-added model	3	Large
F	Alternative assessment–based value-added model	3	Small (< 10,000 students)
G	Alternative assessment–based value-added model	2	Medium
H	Alternative assessment–based value-added model	3	Medium

**Note:** Many of the districts requested anonymity, and so districts are not identified in this report.

**Source:** Interviews with district administrators supplemented by a review of documents on district websites.

---

The eight sample districts use or have used alternative assessment-based VAMs and SLOs for a variety of purposes. Five districts have made (or plan to make) alternative growth measures a formal part of teacher evaluations. The three districts that have no plans to make the alternative growth measures a formal part of teacher evaluations—all SLO districts—have nonetheless attached consequences in the form of annual bonuses or salary increases that are based on achievement of SLOs.

The following sections present the study findings, grouped by related topics.

### **How districts implement alternative assessment-based value-added models**

This section examines the implementation of alternative assessment-based VAMs in the four sample districts using them. (Box 3 presents background information on alternative assessment-based VAMs drawn from a review of the literature.)

**Locally developed or state-mandated end-of-course assessments are being used by all four sample districts using alternative assessment-based value-added models while one district also uses commercially available tests**

Districts using alternative assessment-based VAMs rely on student assessments that vary by developer (commercial testing companies, states, and local districts), alignment with specific courses, and coverage of grades and subjects. They have in common two key characteristics: they are standardized—the same assessments are given systematically to students in a particular grade, subject, and course across the district, with a consistent scale for scoring—and they measure a teacher’s contribution to student growth using a sophisticated statistical method rather than comparing simple before and after changes. (See box 3 for additional information.)

All four districts that use alternative assessment-based VAMs apply statistical methods to end-of-course assessments, and one of the four also uses commercially available tests.<sup>1</sup> Application of VAMs to these assessments is relatively recent in all four districts; the earliest implementation of alternative assessment-based VAMs in the sample districts was in the 2009/10 school year, and the most recent began in 2011/12.

Alternative assessments can be aligned to district, state, or national content standards. Two districts developed their own end-of-course assessments aligned to district curricula and standards. One district uses state-mandated end-of-course assessments that were developed to correspond to state standards. The fourth sample district implements both state-mandated end-of-course assessments aligned to state standards and two commercial tests. This district administers one commercial test—unaligned with district or state standards—and receives value-added estimates as part of a state-level program to introduce value-added analysis. The state also requires districts to choose a second commercial test from an approved list to identify high-value-added teachers for struggling students.

Districts vary in the subjects and grades covered by end-of-course assessments and commercial tests (table 2). All four sample districts using alternative assessment-based VAMs (Districts E, F, G, and H) employ end-of-course assessments in high school, and two districts use them in middle school as well. All four districts have end-of-course assessments

***Alternative assessment-based VAMs have in common two characteristics: they are standardized and they measure a teacher’s contribution to student growth using a sophisticated statistical method***

### Box 3. What the literature says about alternative assessment–based value-added models

Alternative assessment–based value-added models (VAMs) are student growth measures that apply statistical methods to standardized assessments. These include student growth percentile measures as well as measures explicitly intended to assess a teacher’s value-added. The models are typically used in districts that also apply VAMs to traditional state-mandated assessments, extending the models to additional assessments beyond reading and math in grades 3–8. In most cases the alternative assessment–based VAMs are applied to end-of-course assessments (that may be used statewide or locally only) or to widely used commercial assessments such as the ITBS or the PSAT.

The literature on alternative assessment–based VAMs is small and is reviewed in depth in Gill et al. (2013). Several studies have found that estimates of teacher value-added based on widely used commercial assessments are as stable over time as estimates based on state assessments (McCaffrey, Sass, Lockwood, & Mihaly, 2009; Corcoran, Jennings, and Beveridge, 2011) and correlate with other measures of teacher performance. Multiple studies have found moderate correlations (between 0.4 and 0.6) between teacher value-added estimated using commercial tests and using state assessments (Bill & Melinda Gates Foundation, 2010, 2012; Kane, McCaffrey, Miller, & Staiger, 2013; Mihaly, McCaffrey, Staiger, & Lockwood, 2013; Corcoran et al., 2011), though one study found smaller correlations (Papay, 2011). Studies have also found positive correlations between teacher value-added based on commercial tests and classroom observation estimates—often stronger than correlations between teacher value-added based on state assessments and classroom observation (Bill and Melinda Gates Foundation, 2012; Mihaly et al., 2013).

Findings on end-of-course assessments indicate that VAM estimates can differentiate among teachers but that reliability across years varies. VAMs based on district end-of-course assessments in math, English language arts, science, and history sometimes differentiate among teachers more reliably than VAMs based on state assessments (Johnson, Lipscomb, Gill, Booker, & Bruch, 2013). Another study of the same district using grade 8 science and language arts tests found only weak to moderate correlations—0.15 and 0.31—between estimates using end-of-course tests and estimates using state tests (Lipscomb, Gill, Booker, & Johnson, 2010). The data were from the first year’s administration of end-of-course tests, and the end-of-course tests and state assessments covered different content. Finally, one study found low covariance for estimates of high school teacher value-added across years using state-mandated end-of-course assessments (Jackson, 2012).

*Studies of end-of-course assessments indicate that VAM estimates can differentiate among teachers but that reliability across years varies*

**Table 2. Coverage of end-of-course assessments in sample districts using alternative assessment–based value-added models, by subject, grade, and district**

Subject	K–3	4–5	6–8	High school
English/language arts	E, F	E	E, H	E, F, G, H
Mathematics	E, F	E	E, H	E, F, G, H
Science	E, F	E, F	E, F, H	E, F, G, H
Social studies	F	F	E, F, H	E, G, H
Noncore subjects (for example, art, music, physical education)	E	E	E	E

**Note:** Many of the districts requested anonymity, so districts are not identified in this report.

**Source:** Interviews with district administrators supplemented by a review of documents on district websites.

**Table 3. Coverage of high school end-of-course assessments in sample districts using alternative assessment–based value-added models, by subject and district**

English language arts	Math	Science	Social studies
English 1 (E, F, G, H)	Algebra I (E, F, G, H)	Biology (E, F, G, H)	U.S. history (E, G, H)
English 2 (E, F, G, H)	Algebra II (E, F)	Chemistry (E, F, H)	
English 3 (E, F, G, H)	Geometry (E, F, G, H)		
	Precalculus (E, F, H)		

**Note:** Many of the districts requested anonymity, so districts are not identified in this report. District E has end-of-course exams for all high school courses, including those not listed in the table.

**Source:** Interviews with district administrators supplemented by a review of documents on district websites.

covering at least two courses in each of the four core academic subjects (English language arts, math, science, and social studies). One district administers end-of-course assessments in elementary grades for subjects not included annually in the state assessment regime, such as science. That district also uses end-of-course assessments across grades K–12 in noncore subjects, including music, art, and physical education.

*Each district uses the same VAM approach for state assessments and alternative assessments*

All four districts have end-of-course assessments for high school English 1–3, algebra I, geometry, and biology (table 3). Three districts administer a U.S. history end-of-course assessment, the only one administered in social studies.

In the one district that uses commercial assessments, all students in grades K–8 take one or both of two widely used commercial assessments. One assessment is administered to students in grades 2–8. Students take the full test battery, but the district uses the results only for areas not covered by the state assessment. The other assessment measures math and reading performance in grades K–4.

### The general approach to calculating value-added is similar across districts

VAMs can incorporate different covariates, such as demographic characteristics, and use different estimation approaches. A student growth percentile approach, which is used in some states, is not typically labeled a VAM but produces similar results and is used for teacher evaluation (Goldhaber & Theobald, 2013). Each of the four sample districts works with an outside provider, such as a university center or private company, to develop and estimate its VAM—three providers in all—but the specific VAM statistical model varies. Each district uses the same VAM approach for state assessments and alternative assessments.

## How districts implement student learning objectives

This section describes findings related to the scope of teacher coverage and the implementation process for student learning objectives. (Box 4 presents additional information on SLOs.) Unlike alternative assessment–based VAMs, which are standardized, SLOs are designed to vary not only by grade and subject but also across teachers within a grade and subject. While this flexibility may raise some concerns about comparability of results across teachers, it also means that SLOs can be used for any grade and subject—an attractive feature when a district is seeking a student growth measure for all its teachers. Because SLOs require teacher effort to identify specific outcome measures, set goals, and administer

---

#### **Box 4. Background and relevant literature on student learning objectives (SLOs)**

Student learning objectives (SLOs) evaluate teacher performance within a teaching framework: teachers analyze student needs, establish goals for student growth, plan how to achieve goals, choose instructional strategies, and measure progress. Teachers develop the SLOs near the beginning of each school year. The process begins with an initial assessment of student needs, typically relying on a baseline assessment that is often designed by the teacher and informed by students' assessment results in prior grades. Teachers then set targets for individual student growth or for the percentage of students expected to achieve a proficiency standard (the SLO). Finally, at the end of the school year teachers assess student growth using the chosen measure. SLOs often involve collaboration and consultation with other teachers and administrators and typically must be approved by the principal; in some districts, SLOs are also audited for quality and rigor by district staff (Gill et al., 2013).

The literature evaluating SLOs consistently finds that most teachers achieve at least some of their SLO targets, though the percentages are lower than those for teachers rated satisfactory on the traditional rating scale. Specifically, studies in Denver find that most teachers (70–85 percent) meet at least one target (Goldhaber & Walch, 2012; Proctor, Walters, Reichardt, Goldhaber, & Walch, 2011); studies of other districts find that roughly two-thirds of teachers meet all targets (Tennessee Department of Education, 2012; Terry, 2008; Community Training and Assistance Center, 2013). Two studies have correlated SLO ratings with estimates of teacher value-added and have found small positive relationships (Goldhaber & Walch, 2012; Schmitt & Ibanez, 2010). No studies have reported on the reliability of measures of SLO ratings on a year-to-year basis. For more information on the SLO literature, see Gill et al. (2013).

Implementing an SLO system in a district is an intensive process, and districts may rely on other districts or on external consultants for assistance with the process. Among the districts in the study sample, several have used the first documented district to implement SLOs, Denver Public Schools, as a model and resource. That district's website provides documents guiding the SLO process, including a checklist for teachers developing SLOs and a rubric for assessing them (<http://sgoinfo.dpsk12.org/>).

*SLOs can be used for any grade and subject—an attractive feature when a district is seeking a student growth measure for all its teachers*

---

the measures, along with the principal's approval of every measure and goal, SLOs place much greater demands on teachers and principals than do alternative assessment-based VAMs. But SLOs allow teachers to customize them to correspond closely to course goals and the needs of the students they serve. SLOs are also more directly connected to instructional improvement than are alternative assessment-based VAMs, because they are an integral part of the teacher's planning of the school year.

#### **Teachers of all grades and subjects create student learning objectives**

To have a comparable measure of performance for all teachers, each of the four sample districts using SLOs requires all teachers in grades K–12 to complete SLOs, even in grades and subjects in which state standardized assessments are administered. Three of the four districts initially implemented SLOs as part of a performance-based compensation system. SLOs provided a measure of performance that could be applied to all teachers regardless of grade or subject. One district also noted that using SLOs across grades and subjects in which state assessments were also administered allowed the district to compare SLOs and achievement levels. Teachers in all four districts must include English language

learner students and students in special education in their SLOs. Each district has created a districtwide SLO template, and teachers complete and administrators approve the SLOs online through a web application. All four districts have implemented SLOs for at least three years in addition to any piloting.

### **Districts provide training, support, and guidelines for preparing student learning objectives**

To help teachers develop and principals evaluate SLOs, three districts provide in-person professional development. Two districts send training teams to schools during the first two months of the school year to provide individualized and targeted SLO support. Another district reviews SLOs during new teacher orientation and provides school-based training as needed. The fourth district has long experience with SLOs and believes that teachers already understand how to develop and complete them; new teachers receive SLO handbooks and examples. All four districts also use other tools, such as webinars, workshops, and extensive documentation to provide guidance on SLOs.

Preparing SLOs requires teachers to analyze student needs, set growth goals, and implement an appropriate assessment. To identify initial student needs, teachers often review students' prior-year academic data and assess students' knowledge. Teachers can develop the assessment items, use or adapt items from previous district or state exams, or administer or adapt commercially available achievement tests. In areas where students are expected to have no baseline knowledge, such as the first year of foreign language instruction, teachers can use data from previous classes on areas where students typically struggle. The assessment can involve multiple-choice, short-answer, or essay items. Each district reports working to ensure that assessments are rigorous, valid, and reliable. For example, districts require that essays be graded using a well-specified rubric and that assessment items vary in difficulty.

Using the identified student needs, teachers in each district set learning goals aligned to standards (district, state, or Common Core standards). Three of the districts require two SLOs per teacher; teachers in the fourth district are required to submit at least one SLO but can submit more. Although at least one goal usually refers to all of a teacher's students, not all students are necessarily expected to meet the goal (for example, goals might refer to 75–85 percent of students). A teacher's second SLO might identify a specific student population to be assessed and evaluated, determined by the teacher or based on the larger goals of the school or district.

All the sample districts require that goals set high expectations for student achievement—beyond “normal” and “not easily attainable” expectations—such as at least a year of growth as measured by the assessment developer or moving up one proficiency level. Teacher goals vary by grade and subject. For example, an elementary teacher might have one reading and one math goal, while a secondary teacher will set subject-relevant goals. Some goals might be mandated; for example, a district could require that physical education teachers set a cardiovascular fitness goal. Two districts also require that teachers identify the research-based instructional strategies that will be used to achieve the goals.

In each sample district teachers develop and finalize their goals and administer initial (baseline) assessments during the first two or three months of school. Districts encourage or require teachers to work with their principal, a team of administrators, or colleagues. In

***Preparing SLOs requires teachers to analyze student needs, set growth goals, and implement an appropriate assessment; all sample districts require that goals set high expectations for students***

three of the sample districts the teacher submits baseline assessment results and student goals around the end of October. In the remaining district, SLOs might cover only a portion of the year—long enough to reasonably demonstrate growth—and teachers can have multiple SLOs that cover sequential units or periods of instruction.

### All districts created systems to promote consistency and rigor in student learning objectives

After teachers submit SLOs, an administrator, usually the principal, must review and approve them. To promote consistency across district schools, three of the sample districts also require that SLOs be approved by a district official after the principal's approval; the fourth district plans to incorporate district reviews in the future. Two districts review the SLOs to ensure a consistent level of difficulty, and one district checks initial assessments to confirm that the identified student population demonstrates a genuine need. The remaining district is currently adding a measure of difficulty to SLO evaluations.

**Sample districts that use SLOs favored them as a method of assessing student outcomes because SLOs foster teacher ownership of instructional goals**

In all four sample districts using SLOs, teachers administer the assessment again in the spring, during the last two months of school. To measure teacher value-added, two districts require that the fall and spring assessments be identical. The other two districts provide more flexibility, allowing teachers to use parallel forms of commercial tests or—in a few subject areas—demonstrate that students meet a proficiency level. (Because meeting a proficiency level does not measure growth, the district that allows a proficiency level standard also requires a growth-based goal using fall and spring assessments in all core academic subject areas.) Teachers then grade the assessments and submit them for administrator review. In two districts the principal decides whether each SLO has been met (yes or no); in the other two districts the principal assigns a score based on a rubric.

### How districts chose and designed the alternative growth measures

All four districts using alternative assessment-based VAMs were already using the alternative assessments for other purposes or were required to use them by the state or as part of an externally funded initiative. Sample districts that use SLOs favored them as a method of assessing student outcomes because SLOs foster teacher ownership of instructional goals. The districts that use alternative standardized assessments liked that the measure fit within the existing assessment framework and VAMs. In particular, two districts reported choosing SLOs to empower and engage teachers and build collaboration between teachers and principals over shared goals. SLOs also had the advantage that they could be implemented for teachers in all grades and subjects.

### All districts received external funding that supported the design process

All eight sample districts received financial support from external funders to measure student growth and used the funding to design and implement the alternative measures. In general, districts approached funders for support after deciding to implement the alternative measure. Five of the sample districts received funding from the U.S. Department of Education's Teacher Incentive Fund to develop and implement performance-based teacher and principal compensation systems in high-need schools. Some districts received funding through the federal Race to the Top initiative, and some received funding from local or national foundations (including the Measures of Effective Teaching initiative of the Bill & Melinda Gates Foundation).

## Design and implementation required close collaboration between administrators and teachers for student learning objectives but not necessarily for alternative assessment-based value-added models

SLO frameworks were generally developed as collaborations by administrators and teachers. Implementing SLOs required considerable teacher and principal participation. To generate support for the process, three of the four districts using SLOs formed design teams of teachers—sometimes represented by teachers union officials—and district staff that developed the SLO approach collaboratively. Development took one year, on average, and ranged from roughly six months to two years, depending largely on whether a district used pre-existing alternative measures and whether development included a pilot period. One district tasked an autonomous team with developing and supporting SLOs to ensure that the measure could survive changes in district leadership. Another district discontinued implementation of the SLOs when a change in superintendent resulted in a policy shift. Design teams sought out support. Two districts worked with an outside consultant to develop SLOs, and two studied the SLO system in Denver, which had pioneered SLOs a decade earlier.

*Districts typically piloted the SLOs to identify challenges and problems; districts using alternative assessment-based VAMs piloted the value-added statistical models but did not need to pilot administration of the alternative assessments*

Three of the four districts implementing alternative assessment-based VAMs did not require a comprehensive design process because the assessments were already in place. One district had been using locally developed end-of-course assessments as part of students' grades and worked with a vendor to set up a process for using them in alternative assessment-based VAMs. Two districts using alternative assessment-based VAMs were not involved in design because the state or funder mandated the details. The fourth district had previously developed local end-of-course assessments to align with the district curriculum; it subsequently engaged teachers and administrators to work with an external vendor to design a VAM.

Sample districts typically piloted the SLOs to identify challenges and problems, and one district also reported piloting SLOs to build the support of the teachers union. Districts using alternative assessment-based VAMs typically piloted the value-added statistical models but did not need to pilot administration of the alternative assessments, which had been developed prior to their use for teacher evaluation.

## How districts provide feedback and use the performance ratings

In addition to receiving a performance-based rating, teachers in SLO districts and VAM districts alike typically receive some additional results, including benchmark comparisons, student achievement data, and (in districts using VAMs) other value-added analyses and feedback. The ratings can be used for formal evaluations, performance-based compensation (either a one-time bonus or a change in base salary), professional development, tenure, and teaching assignments.

SLOs are also used to help teachers plan instruction, a use unrelated to measuring teacher performance. For example, the sample districts indicate that the SLO structure facilitates conversations between teachers and principals on school and district goals. Developing SLOs also often requires teachers to create their own assessments and pushes them to use data to drive instruction. However, the role of SLOs in teacher evaluation could conflict with their role in instructional planning. Using SLOs to plan instruction requires developing SLOs in areas with the strongest student need and setting high expectations for

students. In a high-stakes setting, where SLOs can determine teacher evaluation ratings or compensation, there could be an incentive for teachers to set more easily attainable goals.

### **Principals generally provided earlier performance feedback to teachers in districts using student learning objectives than in districts using value-added models**

In the districts using SLOs, principals usually provide feedback to teachers in person at the end of the school year or during the summer. Principals sometimes also provide written feedback. One district requires principals to provide written feedback for teachers who received the lowest rating on the scale, and another district provides written feedback to teachers after reviewing the principal's feedback.

All teachers in the four sample districts using alternative assessment-based VAMs receive feedback on their ratings, but there are differences in timing, in how it is provided, and in who provides it. In all four districts teachers receive feedback in the fall of the following school year—later than in the districts using SLOs—after the relevant student tests are administered. One district provides feedback twice: in October teachers receive their individual value-added scores, and in November they receive a report that compares their score to those of other district teachers.

In all four districts the value-added reports provided to teachers include information beyond their scores, such as confidence intervals, significance levels, and students' achievement levels based on their prior-year test scores. In two of the four districts teachers also meet with principals; the other two districts allow principals to decide whether to provide additional feedback in person.<sup>2</sup> All sample districts using alternative assessment-based VAMs indicated that teachers receive professional development or training to help them interpret the reports they receive.

### **In addition to evaluation, districts use performance ratings for teacher compensation and professional development**

All eight sample districts reported multiple intended uses for the alternative measure (table 4). The most commonly reported uses include formal teacher evaluations, performance-based compensation, and professional development. Some districts also use alternative growth measures to inform teacher assignments or tenure decisions.

*Formal teacher evaluations.* Two sample districts already use alternative assessment-based VAMs or SLOs for formal teacher evaluation. One of these districts evaluates teachers using a district-specific VAM instead of the state's VAM. The district-specific measure was intended to more precisely link students with teachers through additional class roster checks. This district administers district-developed end-of-course assessments in all grades and subjects that are not covered by the state assessment. Student growth, measured by the VAM, accounts for 40 percent of a teacher's evaluation rating regardless of whether an end-of-course assessment or state test is used as the outcome. The remaining components come from classroom observations by peers and administrators. The other district uses SLOs as a minor component of the teacher evaluation system (10–20 percent); teacher observation ratings and VAM estimates determine most of the score.

***The most commonly reported uses for the alternative measure include formal teacher evaluations, performance-based compensation, and professional development***

**Table 4. Intended uses of alternative measures, by measure and district**

Intended use	Student learning objectives				Alternative assessment–based value-added model			
	District A	District B	District C	District D	District E	District F	District G	District H
Component of formal evaluation (weight)	No	No	Yes (15 percent)	No	Yes (40 percent)	Yes <sup>a</sup> (≤ 26 percent)	Yes <sup>a</sup>	Yes <sup>a</sup> (30 percent)
Compensation	Annual bonus	Annual bonus	Annual bonus and salary increase	Annual bonus and salary increase	Annual bonus <sup>a</sup>	Annual bonus	No	Annual bonus
Tenure decisions	District does not offer tenure	No	District does not offer tenure	No	Yes	District does not offer tenure	No	Yes
Dismissal	No	No	Yes	No	No	No	No	No
Professional development	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Teaching assignments/promotions	No	Yes	Yes	No	No	Yes	Yes	Yes

a. District is planning to use the measures for this purpose.

**Source:** Interviews with district administrators supplemented by a review of documents on district websites.

Three of the other four sample districts are implementing alternative assessment–based VAMs to prepare for a state mandate to include student growth in most teacher evaluations. These districts chose to begin early to refine and improve the approach and to familiarize teachers and other stakeholders with the measures.

**Three sample districts use or have used alternative assessment–based VAMs as part of a performance-based compensation system, and three use SLO ratings for performance-based teacher compensation**

Two of these districts have established frameworks for evaluating teachers using alternative assessment–based VAMs. In one of these districts all teachers in core subjects in grades 4–12 will be evaluated using VAMs, and individual value-added scores will account for 30 percent of a teacher’s evaluation rating, regardless of whether the model is based on state assessments, locally developed end-of-course assessments, or both. Individual value-added scores will be combined with schoolwide value-added scores (based on state assessments and alternative assessments), student perception surveys, and classroom observations to create a teacher’s overall evaluation rating. The other district plans to use estimates of teacher value-added to evaluate teachers in grades 3–8 as well as teachers of algebra I, algebra II, biology, chemistry, English (grades 9–11), and precalculus. Value-added scores will count for less than 30 percent of a teacher’s evaluation rating and will be used in conjunction with teacher performance measured through formal observation and classroom walkthroughs, as well as SLOs.

**Performance-based compensation.** Three sample districts use or have used alternative assessment–based VAMs as part of a performance-based compensation system. One district offers annual bonuses to teachers who receive value-added scores exceeding a certain threshold. The other two use the measure as one component of an overall performance measure, and teachers receive annual bonuses based on their scores on the overall measure. In one of the two districts bonuses are based on student growth at the school level, while in the other district bonuses are based on individual teachers’ value-added scores derived from the state assessment or a widely used commercially available test. The fourth sample district plans to introduce performance-based pay using value-added results on state or alternative assessments.

Three districts use SLO ratings for performance-based teacher compensation. In one district SLOs affect a teacher’s evaluation ratings and that rating affects teachers’ eligibility

for bonuses and base salary increases. In the other two districts SLO ratings determine part of the bonus or salary increase; the rest of the increase is determined by other criteria, such as teaching in a low-achieving school. In one of these districts teachers have to achieve a certain evaluation rating to be eligible for the bonus. In the other district teachers receive a bonus if they achieve one objective and a salary increase if they achieve two. In both districts bonuses for achieving SLOs range from less than \$1,000 per objective to \$2,000.

*Professional development.* All eight sample districts use their alternative growth measures for professional development, formally or informally. The four districts using alternative assessment-based VAMs reported that the value-added results guide or will guide professional development in the district. The SLO sample districts do not formally use SLOs to identify teachers for professional development, though districts reported that some principals use SLOs for that purpose on an ad hoc basis. All four districts that use alternative assessment-based VAMs indicated that they use or intend to use the results to identify teachers for additional professional development or coaching. One district also plans to align professional development units with the teacher evaluation framework. All four districts that use alternative assessment-based VAMs noted that the value-added scores allow the district to promote ongoing conversations on data use by teachers to better understand student growth in relation to their own instructional practices.

**All eight sample districts use their alternative growth measures for professional development, formally or informally**

*Teaching assignment.* One district indicated that it has used the alternative assessment-based VAM results to develop teaching assignments. The district has created high school teaching assignments that allow teachers to specialize in their strongest subject. The other three districts using alternative assessment-based VAMs do not formally adjust assignments based on the results but suggested that principals might do so on their own. None of the districts using SLOs use them formally to adjust teaching assignments, though individual principals may do so informally.

*Tenure.* Two districts link tenure decisions to alternative assessment-based VAM ratings. In one district overall evaluation scores—based in part on estimates of value-added for some teachers—will affect tenure decisions. In another district teachers will have a choice beginning in 2013/14: give up their tenure (to opt into a system that provides bonuses based in part on value-added results) or remain in the current system (where tenure is unaffected by value-added and they are ineligible for the bonus). New teachers will be ineligible for tenure but eligible for the performance-based bonus. Three districts do not link alternative growth measures to tenure decisions, and three other districts do not have teacher tenure and were unsure about the measure's role in developing future teacher contracts.

### **Implications and next steps**

The findings presented here raise several questions about the implementation and use of the alternative growth measures, including the reliability and validity of the measures, the extent to which they differentiate teacher performance, and the specific costs and benefits of the measures for districts. To provide an in-depth description of the implementation of alternative growth measures, research on these measures should:

- Include data from a variety of stakeholders in each district.
- Assess how much differentiation scores on the alternative measures produce among teachers in each district.
- Look for evidence of the validity and reliability of the alternative measures.

- Describe the district-developed infrastructure for collecting and analyzing data on the alternative growth measures.
- Examine the perceived benefits and drawbacks of using each type of alternative growth measure for evaluating teachers.

This research would be useful to states and districts across the country as they work to incorporate student growth measures in their new teacher evaluation and compensation systems.

## **Appendix A. District administrator interview protocol**

---

### **A. Introduction and consent for audio recording**

The purpose of this section is to introduce the respondent to the study and request consent for audio recording.

1. Thank you for taking the time for the interview today. My name is [name] and I work for Mathematica Policy Research, which is conducting this study for the Mid-Atlantic Regional Educational Laboratory (REL). As I mentioned [*in my email/when we spoke on the phone*], we are conducting case studies of districts implementing measures of teacher performance that include student growth but that measure growth on outcomes other than state assessments. These alternative outcome measures might include nationally normed assessments, curriculum-based assessments developed by the district, or “student learning objectives” (SLOs) developed by individual teachers for their own students. Student growth might be measured through a simple subtraction of end-of-year achievement from beginning-of-year achievement, through a student growth percentile method, through a value-added statistical model, or implicitly in SLOs that are based on an understanding of baseline achievement levels in a class at the beginning of the school year. We plan to examine which alternative outcome measures are used, how the growth measures are implemented, the challenges and obstacles in implementation, how the measures are being used, and the effectiveness of these measures in differentiating teacher performance. We expect our findings to benefit states and districts considering the use of such alternative growth measures. The purpose of this interview is to gather information on the implementation of [*insert specific growth measures used in the district*] in the district from your perspective as a district administrator. We will be interviewing teachers, principals, and union representatives in the district to gather information on their perspectives as well. We will not identify you by name in the study reports.

Before beginning, I would like to get your permission to record the interview. This will ensure that we accurately capture information about district policies and will prevent me from having to ask questions and take notes at the same time. The audio recordings from all individuals interviewed within your district will be compiled into a single district summary that is for internal analysis purposes only and will not be published or shared with anyone outside the study team. You will not be identified in the district summary. This summary will then be aggregated with other district summaries and analyzed for trends and variations across multiple districts. Do I have your permission to record the interview?

### **B. Type of student outcome and growth measure used**

2. Please describe the student outcome measure(s)—other than state assessments—used in growth measures of teacher performance implemented by the district.
  - a. Prompt: Does the district use any student assessments developed by the district or schools in the district?

- b. Prompt: Does the district use SLOs for purposes of measuring student growth associated with individual teachers?
  - c. Prompt: Does the district use any end-of-year, nationally normed or summative assessments purchased from testing companies (not including interim or benchmark assessments)?
2. In what subjects and grades are these measures implemented? Do these measures apply to teachers of English language learner students and special education students or students with IEPs? For how many school years has each measure been in place?
  3. (If the district uses SLOs) Please briefly describe any formal guidelines the district provides to teachers or principals regarding the development of SLOs. Do teachers receive any training or professional development on developing SLOs? How about principals? Is there a district-wide template used by teachers? What is the process for approving SLOs submitted by teachers? Are teachers required or permitted to submit more than one SLO for a class? Does this vary by grade and subject area? Does the district have any way to determine whether the SLOs are of comparable difficulty for different teachers?
  4. I would like to learn a bit more about how the alternative measure(s) used in the district was designed. What was the process for deciding on the specific measure currently implemented in the district? Did the district use examples from other districts or states? Was the student outcome measure administered in the district prior to its use in growth or value-added teacher performance measures? If not, was there a piloting period for the alternative assessment? How was it selected or designed?
  5. Does the district use a value-added model (VAM) for assessing teacher performance (aside from school accountability)? If not, how does the district measure growth (e.g., percentile method, simple subtraction)? How long has the growth/VA model been in place? Has the district made any changes to this approach to measuring growth during that period?
  6. (If the district uses a widely recognized statistical model for measuring growth/value-added.) What is the name of the growth/VAM, and who designed it? Is the district using the same statistical model for alternative measures as it uses for state assessments?
  7. Prior to the implementation of this alternative growth model, how did the district measure teacher performance (e.g., measures of professional practice via classroom observations)? Is the district continuing to use these measures in addition to the alternative measure(s) you've described?

### C. Implementation of data collection and analysis

1. I would like to learn more about the administration of these alternative student assessments in the district. (For each assessment) How is the assessment administered (e.g., timing, frequency, who administers)? What is the process for collecting assessment results?

2. What is the procedure for analyzing assessment results and assigning teacher performance ratings? Who conducts the analysis? Is there a department in the district specifically responsible for the analysis? If not, does the district use an external vendor for this task?
3. Please describe any quality control measures you have in place for the collection and analysis of the teacher performance data (e.g., use of external vendors for test administration/data collection, evaluation of model or VAM estimate bias by external consultant, auditing of teacher SLO documentation).
4. What kinds of obstacles, if any, have you encountered in collecting and analyzing this data? For example, have you found you have sufficient staff and resources? (If applicable) Has the district encountered any difficulties collecting documentation of SLOs from teachers?
5. How has the district addressed these issues or challenges? From your perspective, has the district made any significant improvements to its approach?

#### **D. Differentiation produced by the alternative growth or value-added measure**

1. Please describe the scale for rating teacher performance on the alternative measures. Is it binary (reached target or not)? Categorical (e.g., beginning, developing, proficient, advanced)? Continuous (e.g., percentile, NCE, z-score)? (If continuous) What type of rating/estimate is used? (If categorical) What are the categories used?
2. Please describe the distribution of teacher performance across these ratings for each year the alternative measure has been in place. What proportion of teachers is in each category? (If respondent is unsure across all categories) Approximately what proportion of teachers is in the top and bottom categories? Are there any reports or documents you can provide that might offer some additional detail? If not, is there another staff member who might be able to provide more information?
3. Please describe the scale for rating teacher performance on any other measures currently or previously used in the district (e.g., measures of professional practice or state-test based VAM estimate—use response to B.5. to frame question appropriately). Has the district made any effort to compare teacher performance across the different measures currently or previously used in the district? If so, how does the distribution of teacher performance on these current measures compare to the distribution of teacher performance on the alternative measures that you previously described? How does the distribution of teacher performance under your district's old system compare to the distribution of teacher performance on the alternative measures?
4. Do you have any evidence on the proportion of teachers whose scores on the alternative measure are statistically distinguishable from average?
5. Do you have any information about the reliability or validity of the results? In other words, has the district examined teacher scores across multiple years or measures? Do individual teachers tend to earn similar scores from one year to the next? Do individual teachers tend to earn similar scores across different measures (within the same year)?

## E. Use of alternative growth or value-added measure

1. For what purposes is the district using each alternative student outcome measure used in assessing teacher performance?
  - Is it used for evaluating teachers? In what grades and subject areas?
    - Is it used as part of a performance-based compensation system? In what grades and subject areas?
    - Is it used as a means of targeting professional development to specific teachers?
  - Is the measure used to make decisions about teacher assignments?
  - Are there any other ways in which the district is using these measures?
  - If there are others in the district who can provide additional information on this issue, can you give us their contact information?
2. What type of feedback is provided to teachers regarding their performance on each measure? When is this feedback provided? Who provides the feedback? Is it provided in writing? In person?
3. If used in teacher evaluation, is the measure combined with any other measures of teacher performance (for example, value-added or other growth measures based on state assessments or classroom observation measures)? How much weight does each performance measure receive in a teacher's overall evaluation rating? Does the use of the measure in teacher evaluation vary by grade and/or subject area?
4. Does a teacher's performance rating on the alternative measure affect whether he or she is awarded tenure? Does it affect whether a teacher is targeted for additional monitoring and support, such as professional development, coaching, or additional interim evaluations?

## F. Benefits and drawbacks

1. Could you summarize some of the main benefits, from your perspective, of implementing this alternative VAM or growth measure of teacher performance in the district? For example, have you been able to measure the performance of a larger percentage of teachers in the district? How have teachers in the district responded to the use of alternative student outcomes to measure their performance?
2. What do you perceive as the primary costs of implementing the measure? Are there specific monetary costs to the district (e.g., administering additional tests and collecting data)? Has the district encountered any obstacles implementing the technical aspects of collecting and analyzing the teacher performance data? Has the measure allowed for a sufficient level of comparability across teachers? If SLOs are used in the district, has the district been able to ensure that all teachers are setting sufficiently high standards for student growth?
3. Could you describe the district's approach to communicating with the various stakeholders as the alternative measure was rolled out? How do you perceive the district's communication of the use of the alternative measure in evaluations to principals, teachers, and the teachers union/association?

## G. District context

1. Have there been any changes in the district administration or policies that may have affected the implementation of these alternative measures of teacher performance (for example, change in superintendent or implementation of a performance-based pay system)?
2. Please describe the teachers union/association's involvement in the development of the VAM or growth measure. To what extent did the district engage the union/association in development and implementation? Did the district have the support of the teachers union/association for the implementation of the VAM or growth measure of teacher performance using an alternative student outcome? Why or why not? (If applicable) How did the union/association respond to the incorporation of the VAM or growth measure into the teacher evaluation system?
3. Has the district received funding or resources from external sources (such as, federal government, foundations) to develop or implement this measure?
4. Is there anything else that you think might be important to know about the development and implementation of this alternative growth or value-added measure in the district? Are any other key stakeholders in the district with whom you think it would be important for us to speak?

## **Notes**

---

1. One additional district among the sample using student learning objectives also uses end-of-course assessment-based value-added models, but the administrator who was contacted had little knowledge about implementation of the alternative assessment.
2. One of these districts indicated that scores were not shared with principals during the pilot period to allow teachers to become more comfortable with VAMs.

## References

- Bill & Melinda Gates Foundation. (2010). *Learning about teaching: Initial findings from the Measures of Effective Teaching project* (MET Project Research Paper). Seattle: Author. Retrieved June 3, 2013, from <http://eric.ed.gov/?id=ED528382>
- Bill & Melinda Gates Foundation. (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains* (MET Project Research Paper). Seattle: Author. Retrieved June 3, 2013, from <http://eric.ed.gov/?id=ED540960>
- Community Training and Assistance Center. (2004). *Catalyst for change. Pay for performance in Denver: Final report*. Boston: Community Training and Assistance Center. Retrieved June 3, 2013, from <http://www.ctacusa.com/PDFs/Rpt-CatalystChangeFull-2004.pdf>
- Corcoran, S. P., Jennings, J. L., & Beveridge, A. A. (2011). *Teacher effectiveness and high- and low-stakes tests* (Working Paper). New York: New York University. Retrieved June 3, 2013, from [https://files.nyu.edu/sc129/public/papers/corcoran\\_jennings\\_beveridge\\_2011\\_wkg\\_teacher\\_effects.pdf](https://files.nyu.edu/sc129/public/papers/corcoran_jennings_beveridge_2011_wkg_teacher_effects.pdf)
- 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 January 7, 2014, from <http://eric.ed.gov/?id=ED544205>
- Goldhaber, D., & Theobald, R. (2013). *Do different value-added models tell us the same things?* (Carnegie Knowledge Network Brief No. 4). Stanford, CA: Carnegie Foundation for the Advancement of Teaching.
- Goldhaber, D., & Walch, J. (2012). Strategic pay reform: A student outcomes-based evaluation of Denver's ProComp teacher pay initiative. *Economics of Education Review*, 31(6), 1067–1083. <http://eric.ed.gov/?id=EJ989067>
- Jackson, C. K. (2012). *Teacher quality at the high-school level: The important of accounting for tracks* (Working Paper No. 17722). Cambridge, MA: National Bureau of Economic Research. Retrieved June 3, 2013, from <http://www.nber.org/papers/w17722>
- Johnson, M., Lipscomb, S., Gill, B., Booker, K., & Bruch, J. (2013). *Value-added models for the Pittsburgh Public Schools* (Report to Pittsburgh Public Schools). Washington, DC: Mathematica Policy Research. Retrieved June 3, 2013, from [http://www.mathematica-mpr.com/publications/pdfs/education/value-added\\_pittsburgh.pdf](http://www.mathematica-mpr.com/publications/pdfs/education/value-added_pittsburgh.pdf)
- Kane, T. J., McCaffrey, D. F., Miller, T., & Staiger, D. O. (2013). *Have we identified effective teachers?* (MET Project Research Paper). Seattle: Bill & Melinda Gates Foundation. Retrieved June 3, 2013, from <http://eric.ed.gov/?id=ED540959>

- Lipscomb, S., Gill, B., Booker, K., & Johnson, M. (2010). *Estimating teacher and school effectiveness in Pittsburgh: Value-added modeling and results* (Report to Pittsburgh Public Schools). Washington, DC: Mathematica Policy Research. Retrieved June 3, 2013, from [http://www.mathematica-mpr.com/publications/PDFs/education/pittsburgh\\_vam.pdf](http://www.mathematica-mpr.com/publications/PDFs/education/pittsburgh_vam.pdf)
- McCaffrey, D. F., Sass, T. R., Lockwood, J. R., & Mihaly, K. (2009). The intertemporal variability of teacher effect estimates. *Education Finance and Policy*, 4(4), 572–606. <http://eric.ed.gov/?id=EJ863346>
- Merriam, S. B. (2001). *Qualitative research and case study applications in education*. San Francisco: Jossey-Bass.
- Mihaly, K., McCaffrey, D., Staiger, D., & Lockwood, J. R. (2013). *A composite estimator of effective teaching*. Santa Monica, CA: RAND Corporation.
- Papay, J. P. (2011). Different tests, different answers: The stability of teacher value-added estimates across outcome measures. *American Educational Research Journal*, 48(1), 163–193. <http://eric.ed.gov/?id=EJ911163>
- Proctor, D., Walters, B. Reichardt, R. Goldhaber, D., & Walch, J. (2011). *Making a difference in education reform: ProComp external evaluation report, 2006–2010*. Prepared for Denver Public Schools. Denver: The Evaluation Center.
- Schmitt, L., & Ibanez, N. (2011). *AISD REACH program update: 2009–2010 Texas Assessment of Knowledge and Skills (TAKS) results and student learning objectives (SLOs)* (Austin Independent School District Department of Program Evaluation Publication No. 09.83RB). Austin, TX: Austin Independent School District.
- Tennessee Department of Education. (2012). *Teacher evaluation in Tennessee: A report on year 1 implementation*. Retrieved June 3, 2013, from <http://eric.ed.gov/?id=ED533726>
- Terry, B. D. (2008). *Paying for results: Examining incentive pay in Texas schools*. Austin, TX: Texas Public Policy Foundation. Retrieved June 3, 2013, from <http://www.broadeducation.org/asset/1128-paying%20for%20results.pdf>
- Whitehurst, G. J. (2013). *Teacher value-added: Do we want a ten percent solution?* The Brown Center Chalkboard, April 24. Washington, DC: Brookings Institution. Retrieved June 3, 2013, from [www.brookings.edu/blogs/brown-center-chalkboard/posts/2013/04/24-merit-pay-whitehurst](http://www.brookings.edu/blogs/brown-center-chalkboard/posts/2013/04/24-merit-pay-whitehurst)
- Wiersma, W. (2000). *Research methods in education: An introduction*. Boston: Allyn and Bacon.

