

Part 1. Introduction

Over the past four years, 30 States have passed legislation¹ calling for new, rigorous evaluations for teachers and principals that incorporate measures of student growth as one of multiple measures of teacher performance.² An additional 14 States are poised to implement similar changes in upcoming years.³ These new laws are changing the way school systems support and evaluate educators and spurring innovation as States and districts devise practical solutions to the challenges of measuring student growth.

This publication aims to help State education agencies, school districts and their partners consider innovative and emerging approaches to measuring student learning in the context of educator evaluations. In particular, States and districts can use this publication as a resource as they work to improve their systems over time and contemplate new approaches to measuring growth.

For the Reform Support Network (RSN), a technical support group for Race to the Top States, there has been a long arc of work focused on non-tested grades and subjects (NTGS) and how to include teachers of NTGS into evaluation systems requiring measures of student growth. The RSN focused mainly on student learning objectives (SLOs) because it was the solution of choice for the majority of States implementing new evaluation systems. Meanwhile, for teachers of tested grades and subjects, many States and districts use quantitative models of various forms, such as student growth percentiles or prediction models.

¹ For example, the Maryland State legislature passed the Education Reform Act of 2010, which calls for an educator evaluation system that “includes data on student growth as a significant component.” Similarly, in 2011, Michigan passed legislation calling for an evaluation system that “uses multiple rating categories that take into account data on student growth as a significant factor.”

² National Center on Teacher Quality (2014). *State by State Evaluation Timeline Briefs*. http://www.nctq.org/dmsStage/Evaluation_Timeline_Brief_Overview

³ *Ibid.*

Connecting This Brief to Other RSN Resources

The RSN has developed a suite of resources to help States incorporate measures of student growth into educator evaluations. States that use SLOs can consult the **Student Learning Objectives Implementation Toolkit** (SLO Toolkit), which includes resources and strategies to implement high-quality SLOs.

The RSN has also developed an **SLO library** of annotated SLOs, which includes sample SLOs from nine States, covering a range of grades and subject areas.

Finally, the RSN **hosted a webinar** to share States’ efforts to streamline assessments in response to perceptions that students are over-tested. Leaders from New York and Connecticut described how they are awarding grants to school districts for inventorying their assessments and eliminating assessments that are redundant or do not contribute to student learning or teacher effectiveness.

In February 2015, the RSN convened a group of experts⁴ and practitioners to consider emerging measures of student learning, such as portfolios and prediction models to grapple with the following questions:

1. What innovative and emerging approaches to measuring student growth are available for States and districts? What promising approaches are being implemented on a small scale?
2. How can States and districts evaluate and improve the way they incorporate existing measures of student growth into educator evaluations?
3. How might States and districts measure student growth in five years?

This publication captures key points that emerged from discussion of these questions during the convening.

Part 2. Innovative and Emerging Approaches: Prediction Models and Portfolios

The RSN surveyed the field to identify innovative or emerging approaches to measuring student growth and invited practitioners from those States and districts to present at the convening. Among the innovative and emerging approaches discussed at the convening were prediction models and portfolios.

Prediction Models

Background on Prediction Models

Prediction models use regression techniques to predict how well students will perform on a particular assessment based on his or her prior test scores and other factors, such as the student's special education status, English-learner status, mobility or socioeconomic status. Many value-added models can be classified as prediction models. A student's value-added estimate is calculated by subtracting his or her predicted score from the actual or observed score. A negative value-added estimate means that the student's score fell short of the predicted score; a positive value-added estimate means that the student's score exceeded it.⁵ The concept of value-added is also shown in the illustration on the following page.⁶

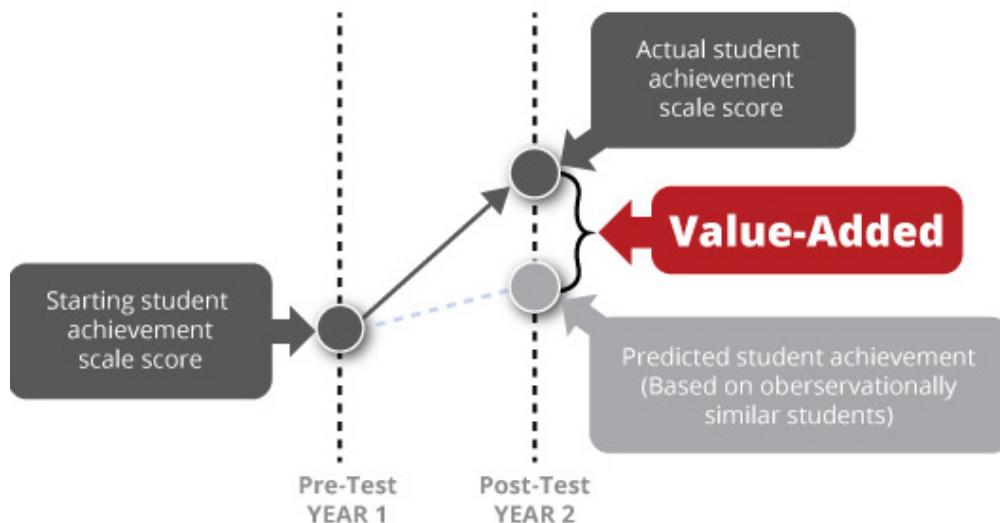
Several States (for example, [Pennsylvania](#), [New Mexico](#) and [Tennessee](#)) use value-added models or similar approaches to measure student growth. States and districts gravitate towards these models, in part, because they use data from large-scale, standardized assessments—the same tests used for school accountability—to determine value-added scores and to generate value-added estimates. Researchers have also found that value-added models are *generally accurate* measures of teacher effectiveness. A recent study found that when “a high value-added teacher” enters a school, his or her students' test scores tend to rise, and those students are more likely to go to college, earn higher incomes and less likely to become teenage parents.⁷

⁴The following experts and practitioners joined the RSN's student growth convening on February 22, 2015: Mark Conrad, Expeditionary Learning; Dru Davison, Shelby County Schools; Jennifer DeNeal, NC Department of Public Instruction; Elena Diaz-Bilello, National Center for the Improvement of Educational Assessment; Carole Gallagher, CRESST/ WestEd; Oliver Grenham, Adams County (CO) School District 50; Margaret Heritage, CRESST/WestEd; Betty Kennedy, Learning Sciences International; Scott Marion, National Center for the Improvement of Educational Assessment; Jesse Olsen, JumpRope Inc.; Michael Toth, Learning Sciences International; John White, SAS Institute; Nadja Young, SAS Institute

⁵For a simple, but detailed, explanation of how value-added models work, see the Oak Tree Analogy, a 10-minute video created by the Value-Added Research Center (VARC). You can view the video [here](#).

⁶Value-Added Research Center at the University of Wisconsin-Madison (2013). *Student Growth Measures*. <http://varc.wceruw.org/what-we-do/growth-measures.aspx>

⁷Chetty, R. et al. (2011). *The Long Term Impacts of Teachers*, National Bureau of Economic Research. http://obs.rc.fas.harvard.edu/chetty/value_added.pdf



Although value-added models are more accurate and reliable than many other measures of educator effectiveness, they have certain limitations. Convening participants acknowledged some of the common limitations associated with value-added models, including the following:

- Typically, teachers don't receive timely feedback from value-added models since student data from large-scale standardized tests are not available until the summer (or later).
- Value-added models are complex and hard for most educators (and the public) to understand.
- Some large-scale standardized assessments may be less instructionally relevant to teachers because they are not derived directly from the experience of students in the classroom.
- Standardized tests cover a narrow range of grades and subject areas, so most States and districts generate value-added estimates for only a small number of teachers (generally those mathematics and English language arts (ELA) teachers in grades three through eight). According to some estimates, teachers in non-tested grades and subjects comprise 69 percent of the teachers in States and districts.⁸

Summary of Experts' Discussion on Prediction Models

At the convening, experts discussed emerging practices from States and districts working to address some of these limitations. North Carolina, for example, uses a *wider range of assessments*—such as statewide end-of-course exams, final exams, career and technical education (CTE) certificate exams and so forth⁹—to generate value-added estimates for approximately 70 percent of teachers in the State.

Pinellas County Schools in Florida, piloted a unit *value-added model* that provides teachers with more timely feedback and covers a broader range of grades and subjects than traditional value-added models. This value-added prediction model used student data from common unit assessments created by teams of teachers in the district. Pinellas County administered the assessments twice during the year (once in the fall and again in the spring) to generate value-added estimates for teachers in all grades and subjects. Teachers received a report that

⁸Prince, C. et al. (2009). *The Other 69 Percent: Fairly Rewarding the Performance of Teachers of Nontested Subjects and Grades*. Center for Educator Compensation Reform. <http://www.maine.gov/education/effectiveness/other69Percent.pdf>

⁹North Carolina Department of Public Instruction, Educator Effectiveness Model: Student Growth. Retrieved June 29, 2015. <http://www.dpi.state.nc.us/effectiveness-model/student-growth/>

included their value-added estimates, along with feedback from classroom observations and student surveys. Teachers received feedback during the current school year, at a time when they were able to intervene with students and adjust their instruction.

UNIT VALUE ADDED MODEL PILOT, PINELLAS COUNTY SCHOOLS (FLORIDA)

In 2013–2014, Pinellas County Schools (“Pinellas County”)—with support from Learning Sciences International—launched a pilot to design a teacher evaluation that would incorporate student growth measures and help teachers improve their professional practice. Moreover, the district wanted to provide student growth scores to teachers during the current school year, instead of over the summer when teachers can no longer intervene with students or adjust their instruction.

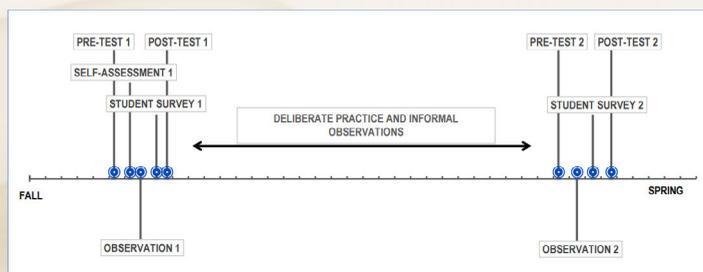
To meet these goals, Learning Sciences and Pinellas County designed a unit value-added model. Unlike traditional value-added models that use data from large-scale standardized assessments, the unit-value added model uses data from common assessments created by teams of teachers in the district. Teachers created two unit assessments in all subject areas—art, music, ELA, mathematics, science and social studies—and grade levels. Altogether, the district created 74 assessments, each of which covered a unit of instruction lasting approximately six weeks.

The district administered the first unit assessments in the fall (see timeline). At the beginning of the unit, students completed the pre-test. Over the next few weeks, teachers delivered instruction aligned to the unit assessment, while the district collected data on teacher performance using a self-assessment, a classroom observation and a student survey. At the end of the unit, students completed a post-test.

Researchers used pre- and post-test scores to generate value-added estimates for each student. The district created an “effectiveness profile” for each teacher showing student growth results along with data from the other measures. Teachers and their supervisors used these “effectiveness profiles” to identify strengths and areas for growth, and to adjust their instruction prior to administering the second unit pre- and post-test, which took place in the spring and informed individual teacher evaluations.

Researchers presented preliminary findings from the pilot at the convening. They found evidence that observation ratings correlated with student growth scores from both the end-of-unit and State’s value-added model and that the feedback teachers received in the fall led to improved instruction and student outcomes. Other convening participants noted that the results in Pinellas County may not be generalizable to the rest of the nation and the unit assessments covered a narrow range of the curriculum. Despite these qualifications, the Pinellas County pilot may represent a step forward in the field of measuring student growth due to its focus on instructional units and providing teachers with more timely feedback. Pinellas County has scaled up the unit value-added model to more schools during the 2014–2015 school year.

Pinellas County Public Schools - Pilot Timeline



- Fall unit and spring unit metrics were used to show growth within a school year.

© 2012 Learning Sciences International 1.877.411.7114

Portfolios of Student Growth

Background on Portfolios

A handful of States—including [North Carolina](#), [Missouri](#) and [Tennessee](#)—give teachers the option of using portfolios to measure their contributions to student growth. Generally, States provide this option to teachers in subjects where it’s difficult to measure growth using pencil-and-paper assessments, such as the fine arts, physical education and world languages. To assemble portfolios, teachers collect student work samples that

demonstrate growth across standards-based learning domains. States and districts often give teachers some flexibility to select the student work samples that comprise their portfolios, providing the sampling approach used allows for a representation of student performance levels and the teacher’s course load. Teachers submit their portfolios—with student work samples—to a reviewer who has expertise in the same content area. The reviewer then examines the student work samples and assigns the teacher a student growth rating. Generally, at least two reviewers evaluate each portfolio; if the first two reviewers—one of whom may be the teacher herself—assign different ratings, then a third reviewer may review the portfolio. To promote fairness and rigor throughout the review process, States provide extensive professional development to teachers and reviewers on how to use rubrics and analyze evidence of student growth.

Portfolios offer advantages for teachers and evaluators:

- Portfolios are an effective way to document information about student growth in performance-rich subject areas like fine arts, music and physical education.
- Portfolios are comprised of artifacts from classrooms and do not require districts and States to invest in new assessments.
- Portfolio development and review provides a rich source of professional learning; teachers carefully examine student work and reflect on their own practices when assembling their portfolios.
- Portfolios provide leadership opportunities for teachers who become reviewers.

Summary of Experts’ Discussion on Portfolios

Practitioners from Tennessee and North Carolina described their State’s portfolio models at the convening. Their presentations illustrated how States and districts can implement portfolio models in different ways. For example, in both North Carolina and Tennessee, multiple reviewers score each portfolio. However, in Tennessee, teachers self-assess their portfolios before sending them to a second reviewer; in North Carolina, teachers send their portfolios to two (or more) outside reviewers. The table below shows other key differences between the portfolios in North Carolina and Tennessee.

State	Scoring Process	Sampling of Student Work	Scale of Implementation
North Carolina	Two outside reviewers analyze work samples in each portfolio and assign a rating to the teacher. If they agree, the review process ends. If they disagree, the portfolio is sent to a third reviewer who analyzes the work samples and chooses one of the two previously-assigned scores.	An online platform randomly selects three students whose work samples are sent to reviewers. Teachers can “reshuffle” to have the platform select three different students.	The State has implemented portfolios for all teachers in AP/IB courses, fine arts, health and world languages.
Tennessee	Teachers self-score their portfolios before sending to outside reviewers. If the reviewer agrees with the teacher’s self-assessment, the review process ends. If the reviewer disagrees with the teacher’s self-assessment, the portfolio is sent to a second outside reviewer who breaks the tie.	Teachers assemble a “purposeful sampling” of student work that reflects the teachers’ classes, course loads and students. Sampling should also align with subject/grade-level standards.	Districts can select portfolios as measures of student growth in fine arts, world languages and physical education.

Experts cautioned that States and districts should carefully review all options for measuring student growth before adopting portfolios, unit value-added models, SLOs or a traditional value-added model. The framework introduced in the next section of this publication can help States and school districts select assessments and growth measures that are aligned to their theories of learning for students and theories of action for their evaluation systems.

Portfolio Models, Tennessee Department of Education (TDOE)

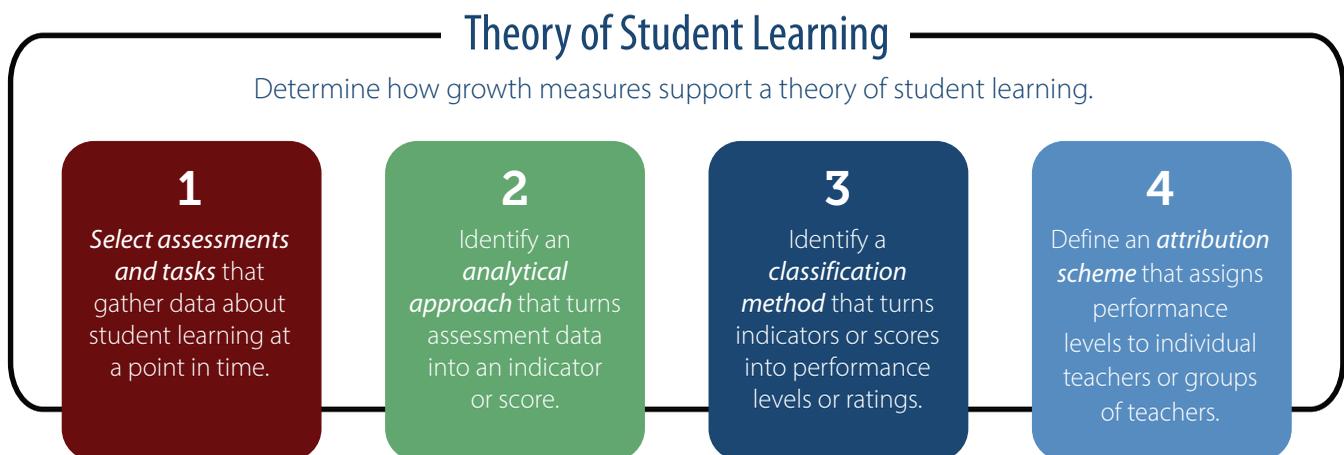
The TDOE developed a portfolio model of assessment under the leadership of Dr. Dru Davison. Its intent is to provide a holistic and meaningful picture of the value teachers add to student learning, using the work already happening in their classrooms. The key features of the Tennessee portfolio model include the following:

- **Evidence collection:** Teachers submit five evidence collections, comprised of a “purposeful sampling” of authentic student growth products. A “purposeful sampling” of student growth should be aligned to standards and reflect the teacher’s course load and students, including students at different points in mastering the standard(s). The TDOE has developed a cloud-based evidence collection tool that teachers can use to assemble their portfolios.
- **Alignment to learning objectives:** Teachers collect artifacts or evidence of student growth that aligns to their stated learning objectives.
- **Scoring:** Teachers score each collection of student work, then submit the collection to a content-specific peer reviewer for scoring. If the teacher and reviewer significantly disagree, then a second reviewer reviews the collection and resolves the discrepancy.

Reviewers score teachers on a five-point scale. Tennessee has created a scoring guide that reviewers can consult, but they have considerable discretion to exercise professional judgment when rating portfolios.

Part 3. Evaluating and Improving Growth Measures: A Framework for Incorporating Student Growth in Educator Evaluations

Scott Marion from the Center for Assessment presented a framework for incorporating measures of student growth into educator evaluations to his fellow participants. In brief, the framework can help States and districts organize the elements of different measures of student growth. Additionally, the framework can help States and districts understand the critical decision points and trade-offs associated with different approaches to measuring student growth. For instance, one district might be flexible about the types of measures it uses for evaluation if its primary goal is to improve instruction; another district that prioritizes the accurate classification of the highest and



lowest performing teachers might use assessments that are comparable across all classrooms. Finally, States and districts can use the framework to identify areas for improvement within existing efforts to measure and attribute student growth.

The framework consists of critical areas for States and districts to consider when incorporating measures of growth into educator evaluations.

- *Determine how growth measures support a theory of student learning* that explains what should be measured and how students acquire new knowledge and skills. A full discussion about learning theory is beyond the scope of this publication, but experts at the convening noted how important it is for districts and States to unify instruction, the assessments used to evaluate student learning and interpretations of student progress under a clear theory of student learning.
- *Select quality assessments and tasks* that gather data about student learning at a given point in time. Participants noted that more and more States use a *body of evidence*—comprised of common performance tasks, assessments created by departments or teams of teachers, AP/IB exams, diagnostic assessments, exams included in curriculum packages, commercially available interim assessments, etcetera—to determine a student’s baseline at the beginning of the year and the amount he or she has grown by the end of the year. Educators will need clear guidance on how to construct bodies of evidence and how to align the evidence with the learning goals for students.
- *Identify an analytic approach* that converts assessment data into some indicator or score. Common analytic approaches used by States and districts to measure growth include value-added analysis, growth percentiles, analysis of student work against a growth rubric and “gain scores” calculated by subtracting a pre-test score from a post-test score. Participants recommended that States and districts consider whether their chosen analytical approaches treat student growth as a computational exercise or yield meaningful information about student growth.
- *Identify a classification method* that turns the results of the analyses into a set of performance levels. In North Carolina, teachers receive one of three possible ratings for their contributions to student learning: Exceeds Expected Growth, Meets Expected Growth and Does Not Meet Expected Growth.¹⁰ Other States and districts rate teachers on four- or five-point scales. Participants noted that States and districts may encounter technical challenges if they have too many performance levels: If measurement errors are sufficiently large, the probability that a teacher lands in one performance category may be attributed more to chance than to his or her actual performance.
- *Clearly define an attribution scheme* that links student performance to educators. An attribution scheme should explain when student growth scores are assigned to individual teachers or shared by groups of teachers. “Shared attribution” of student growth scores may be appropriate when, for example, a student is taught the same subject by two different teachers or when the State or district wishes to foster teacher collaboration.

How do States and districts know if their approaches to measuring student growth are working? How can they improve these approaches over time? The framework includes a set of questions for State and district leaders to consider as they self-assess their efforts:

- How well does the approach approximate a teacher’s influence on student learning?
- To what extent do the assessments align with learning goals for the course/grade level or with the larger theory of action set for the evaluation system?

¹⁰ North Carolina Department of Public Instruction (2013). *Measuring Student Growth for Educator Effectiveness*. <http://www.dpi.state.nc.us/docs/effectiveness-model/student-growth/measuring-growth-guide.pdf>

- How does the approach inform a teacher's instruction? Does it identify effective instructional strategies or areas where students need additional reinforcement?
- What is the level of coherence between growth measures and other State or district priorities? For example, do educators understand how SLOs and the implementation of college and career-ready standards support improvements in teaching and learning?
- Does the approach minimize unintended consequences, such as creating incentives for educators to set less rigorous student growth targets or discouraging teachers from working with students who need the most support?
- When do teachers receive data and feedback on student growth? Is it at a time when they are still able to adjust instruction and intervene with students?
- Are growth measures comparable across schools and classrooms?

These questions may not be equally important to all States and districts. A hierarchy of criteria should reflect the State's or district's theory of action for evaluation. A State or district that prioritizes improving instructional practice might be more interested in making student growth data available earlier or helping teachers understand how to use growth data to inform instruction. Conversely, a State or district that emphasizes the need to accurately classify teachers might place more value on criteria that the growth measures should be comparable across classrooms or based on assessments that have been vetted by psychometricians.

Part 4. Concluding Thoughts

The RSN and experts concluded the convening by sharing their opinions about the future of student growth in educator evaluations. How will States and districts measure student growth in five years? Experts shared the following thoughts:

- States and districts might want to increase the amount of professional judgment that evaluators can use to make determinations about student growth. Professional judgment plays a large role in Tennessee's portfolio model, for instance.
- Growth measures might be supported by a theory of student learning that will answer questions such as, "How much growth is enough?" and "What will my students know if they meet this learning target?" Currently, some States and districts allow teachers to set SLO growth targets based on arbitrary expectations for growth on given assessments rather than considering the essential knowledge and skills that students must learn.
- States and districts might be responsible for ensuring that multiple sources of data are used to enhance what teachers know about their students, **without diluting accountability or evaluation policies**. In these jurisdictions, no single assessment will be considered sufficient to measure what students know. Instead, educators will use data from a variety of sources to inform the setting of learning targets for their students.

This publication features information from public and private organizations and links to additional information created by those organizations. Inclusion of this information does not constitute an endorsement by the U.S. Department of Education of any products or services offered or views expressed, nor does the Department of Education control its accuracy, relevance, timeliness or completeness.