MAY 2013

An Ocean of Unknowns

Risks and Opportunities in Using Student Achievement Data to Evaluate PreK-3rd Grade Teachers

Laura A. Bornfreund

EDUCATION POLICY PROGRAM
Early Education Initiative
education.newamerica.net

New America Foundation

This report is funded through generous grants from the Foundation for Child Development and the W. Clement and Jessie V. Stone Foundation.

The New America Foundation would like to thank the following individuals for sharing their time, expertise, and insights for this paper:

Erika Bolig, Anna Brown, Connie Casha, Vincent J. Costanza, Diane Donohue, Patty Ewen, Rita Fry, Robin Gelinas, Laura Goe, Kristie Kauerz, Toby King, Luke Kohlmoos, Jim Lesko, Scott Marion, Jennifer Mosley, Sam Pearcy, Janice Poda, Tamika Pollins, Carolyn Schokley Ralph, Elliot Regenstein, Sharon Ritchie, Cortney Rowland, Chris Ruskowski, Thomas Schultz, Jonah Stuart, Amanda Szekely, Joann Taylor, Albert Wat, Rob Weil, K. Michelle Worrell, and Margery Yeager.

Contents

Executive Summary 1

Introduction 3

Background 6

The Untested PreK-3rd Grades 8

Experiments in Using Student Growth in the Early Grades to Evaluate Teachers 11

1. Student Learning Objectives 11

2. Creating or Identifying New, Shared Assessments 20

3. Shared Attribution 23

Recommendations and Considerations 25

Conclusion 28

Interviews Conducted 29

Notes 30

Executive Summary

What is the best way to use data to measure teacher impact on student learning? States and school districts are attempting to navigate these uncharted waters. As of 2012, 20 states and DC require evidence of student learning to play a role in evaluating teacher performance. As a result, better information on student learning is in high demand, and no grade level is immune. Historically, most states have required standardized testing only in grades three through eight. But now those 21 states, with likely more to follow, must figure out comparable ways to measure student learning in the "untested grades," as well, including pre-K, kindergarten and grades one and two. And even with testing in grade three, a lack of baseline data has implications for those teachers too.

Determining growth measures for these grades is among the most complex pieces of teacher evaluation reform. In this early stage of life, children's developmental growth—their acquisition of physical, cognitive, and social-emotional skills; their base of general knowledge; their strength of persistence and motivation; and their language and literacy ability—is directly linked to their academic growth. So measures of student learning should account for how young children actually learn and measure more than just reading and mathematics if we are to obtain an accurate picture of a teacher's impact on her young students' learning.

This paper provides a snapshot of how student achievement data are being used in teacher evaluation systems today and illuminates the issues causing states and school districts the most struggles. Most states are using one of or some combination of three approaches: student learning objectives, shared assessments, and shared attribution. The Early Education Initiative at the New America Foundation examines these approaches in five states (Colorado, Delaware, Florida, Rhode Island, and Tennessee) and three school districts (Austin, Texas; Hillsborough County, Florida; and Washington, DC). Each of the approaches carries its own risks and opportunities.

The first approach, student learning objectives (SLOs), centers on a teacher's students. The teacher—with his or her administrator—creates a measurable objective, identifies an assessment to measure that objective, and establishes a challenging but attainable target for students.

Opportunities with SLOs:

1. They foster school-level collaboration and shared priorities.
2. They can help improve instruction.
3. They can help teachers better meet individual student needs.
4. They can support a more well-rounded curriculum.
5. They attain teacher support.

Risks with SLOs:

1. They are resource-intensive to develop.
2. There is limited expertise at the district- and school-level.
3. They come with an inability to compare teachers.
4. They come with a high potential for manipulation.

The second approach is creating or identifying shared assessments at the district or state level.

Opportunities with Shared Assessments:

1. They facilitate comparisons across schools and districts.
2. They could build skills transferrable to the classroom.

Risks with Shared Assessments:

1. They require significant financial and time resources to develop.
2. There are too few appropriate assessments.
3. They could lead to curriculum narrowing and teaching to the test.
4. There are important concerns about test security.

The third approach, shared attribution, uses a school-wide, value-added score. Typically this is based on results from evaluations from third to fifth grade, such as third grade reading scores on a state's standardized test to determine the growth rating for a kindergarten, first, or second grade teacher.

Opportunities with Shared Attribution:

1. It promotes shared accountability.
2. It uses existing resources.

Risks with Shared Attribution:

1. It does not help to provide useful individualized information to teachers.
2. It does not help to differentiate teachers in a meaningful way.
3. It does not measure a teacher's impact on her own students' learning.

To maximize the opportunities and minimize the risks, the New America Foundation's Early Education Initiative makes three recommendations and poses nine considerations for state and district policymakers as they move forward with this work.

Recommendations

1. Account for specific attributes of PreK-3rd teachers.
2. Pilot and evaluate.
3. Do not use "shared attribution" measures from later grades as the sole measure of student growth to evaluate early grade teachers.

Considerations

1. Assessments are designed to be used in specific ways and do not always lend themselves well to other purposes.
2. States will need to decide whether there should be one statewide system or many different district-level systems, and be prepared to provide technical assistance to discover what measures are appropriate for young children, what skills should be measured, and how to measure them in accordance with developmentally appropriate guidelines.
3. While state and district officials may focus on improving numeracy and literacy in PreK-3rd, they should be concerned with whether students are developing crucial skills in the other domains of learning.
4. Engaging schools of education in conversations about teacher evaluation is important, so prospective teachers and principals can gain expertise in assessment models.
5. Creating a system in which teachers set goals and design measures to gauge student growth when their compensation or jobs depend on the results is rife with problems, as with SLOs.
6. Different delivery models of pre-K and kindergarten make it difficult to tie student growth to individual teachers in the earliest grades.
7. States should align their teacher evaluation systems with the Common Core State Standards before implementing the new assessments in the 2014-15 school year.
8. In evaluations of PreK-3rd grade teachers, states and districts should consider whether teachers administered student assessments appropriately and what they did with the data.
9. Since there is limited research on the approaches discussed in this paper, states and school districts should proceed cautiously in selecting assessments for measuring student learning in the early grades.

Regardless of the challenges states face in overhauling teacher evaluation systems, getting it right is crucial in the PreK-3rd grades. Research has confirmed, time and time again, that the quality of instruction and the quality of learning opportunities in children's formative years sets the foundation for their success as students, and, later, their success as adults.

Introduction

What are the best ways to assess what a kindergartner, or first grader, or second grader, has learned? Should a teacher ask her to name words presented on a flashcard on a specific day? Should a teacher track her reading level and writing ability over the course of the year? What about observing whether she is persistent, can stay focused on a task, or is able to solve challenging puzzles?

Policymakers have only started to grapple seriously with these questions in the early grades, and for years, child advocates have urged caution in assessing young children, worried that tests will be inappropriately administered or will not truly reflect what children know. Yet policies for evaluating teachers have already moved on to a question arguably more fraught: What is the best way to use data to measure a teacher's impact on her students' learning?

States and school districts are attempting to navigate these uncharted waters. Since 2009, 37 states and DC have amended their teacher effectiveness policies.[1] As of 2012, 20 states and DC require student learning to play a role in evaluating teachers' performance.[2] As a result, better information on student learning is in high demand, and no grade level is immune. Historically, most states have only required standardized testing in grades three through eight. But now those 21 states, with likely more to follow, have to figure out comparable ways to measure student learning in what are called the "untested grades" as well.

Educators are scrambling to meet the laws' requirements. "Crash and burn;"[3] "moving too quickly;" "not working;"[4] "just trying to a get system in place;" and "insanity;" are just some of the phrases used to describe plans for the untested grades and subjects. Another common lamentation: "Research is not part of the conversation." According to some experts, states in many cases are moving forward with new teacher evaluation systems without fully considering the child development or instructional ramifications of attempts to quantify student learning for high-stakes purposes in the PreK-3rd grades.[5]

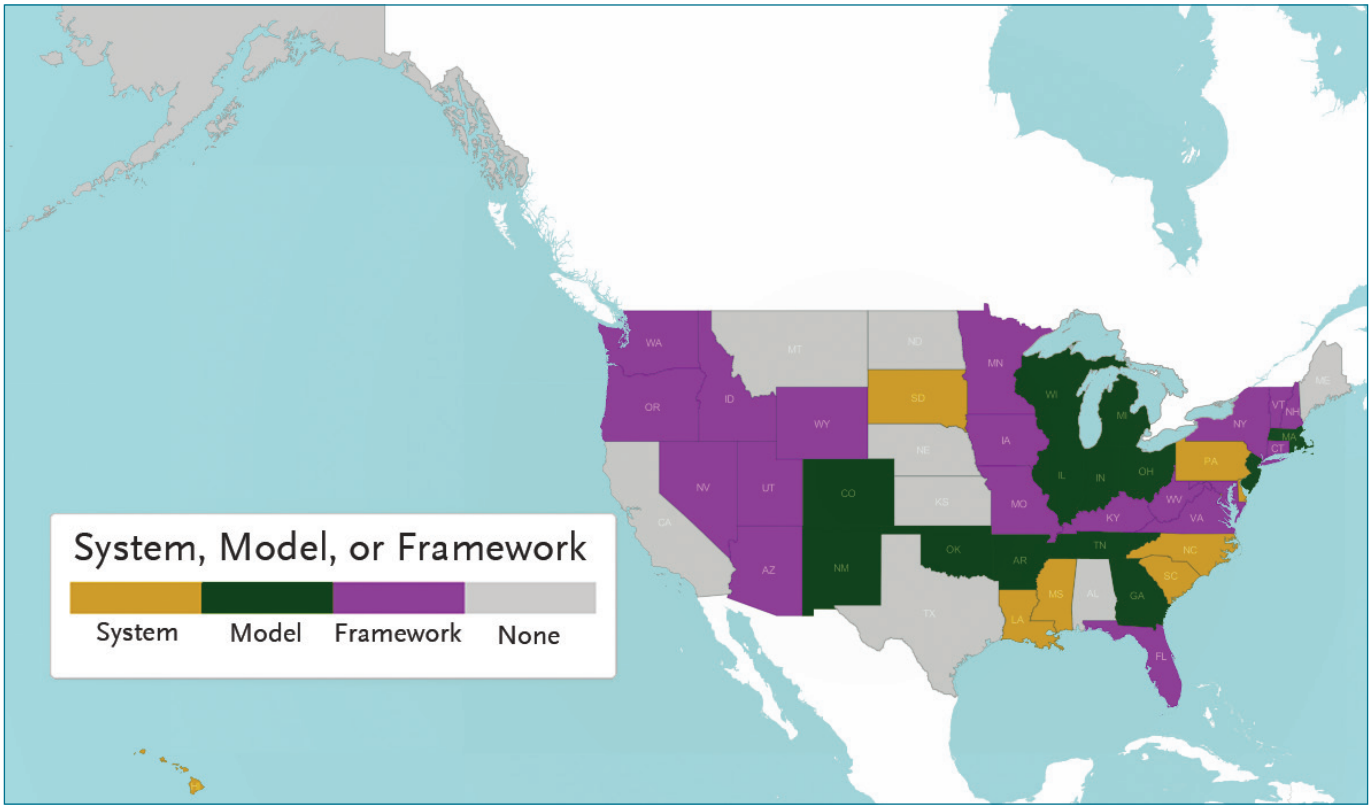
Still, policymakers in these states and elsewhere, guided by the education reform movement of the past decade, appear to agree that teachers' evaluations should include measures of student achievement, no matter what grade level. Few dispute that teacher evaluation systems do need an overhaul. Right now, they do not tell teachers, principals, or policymakers very much at all. Nearly all teachers are rated as effective, yet less than one-third of children are proficient readers by the end of third grade and only 75 percent of U.S. students graduate from high school on time. While out-of-school factors surely play a role leading up to these dire student outcomes, teachers are the most important in-school factor and having access to effective teachers, year-after-year, can have a profoundly positive impact on student success in school and in life.

States and school districts are already sailing on an ocean of unknowns. In the next couple of years, when teachers receive their annual performance evaluations, no matter what grade they teach, data on what their students have learned will be a significant factor in how they are judged. In some states, student performance could be weighted as high as 50 percent of a teacher's overall evaluation. (Other factors in revamped evaluations include observational ratings of teachers in the classroom, student surveys, and lesson plans.)

States and districts have to determine how to use student data to measure teacher effectiveness across the K-12 spectrum—and in many cases, the PreK-12 spectrum—without any waypoints toward proven practices. There are many issues to address, especially: how to expand teachers' and leaders' capacity to follow through on the new requirements, how to equip teachers and leaders with the expertise to choose assessments and analyze outcomes, and how to avoid further narrowing school curricula to the detriment of non-tested subject areas.

In some states, officials in departments of education are making the decisions, designing a model for districts to use if they choose, or allowing districts to design their own systems that meet certain parameters set by the state. Most states appear to be outlining key parameters, and leaving

Evaluation Design: Three Different Approaches States are Taking



***System**: The state designs the evaluation system, which districts must implement.*

***Model**: The state designs an evaluation system model, which districts can choose to use or create their own.*

***Framework**: The state sets certain evaluation parameters, which districts must adhere to in designing their evaluation systems.*

Source: State of the States 2012: Teacher Effectiveness Policies, National Council on Teacher Quality, http://www.nctq.org/p/publications/docs/Updated_NCTQ_State%20of%20the%20States%202012_Teacher%20Effectiveness%20Policies.pdf; Databases on State Teacher and Principal Evaluation Policies, Center on Great Teachers and Leaders at American Institutes for Research, http://resource.tqsource.org/stateevaldb/StateRoles.aspx.

the design decisions to the districts. As a result, experiments are underway (see the map above).

This paper is a snapshot of how student achievement data are being used in teacher evaluation systems today. The intent is to explain what is causing the most struggles and to describe the approaches and refinements already underway. It is framed around three methods that states and school districts are experimenting with to measure student learning for the purposes of teacher evaluation in pre-K and the early grades: developing Student Learning Objectives (SLOs), creating or identifying new standardized assessments for the untested grades, and using a whole school's or whole grade-level's test scores (known as shared attribution). This paper examines risks and opportunities in each method.

Teachers of the early grades have raised their own concerns about using student data in these new evaluation systems. Will these systems reflect the different ways in which young children learn? What types of achievement are appropriate to measure in the young? How should third-grade teachers be measured when there are no state test results for where their students stand at the beginning of the school year? And is it appropriate to include pre-K teachers in these new, more data-driven evaluation systems when they may teach in different settings, receive different levels of compensation, and have different training requirements than their K-12 counterparts in the same state?

Complicating this work further is the fact that states are not just implementing new teacher evaluation systems. At the same time, states are implementing the Common Core State Standards and preparing for the new Common Core assessments, which are set to come on line in the 2014-15 school year. The Common Core State Standards for English

Language Arts and Mathematics include kindergarten through second grade, but the formal assessments do not. The Partnership for Assessments of Readiness for College and Careers (PARCC)—one of two groups that received federal grant funds to create common assessments—does, however, plan to create formative assessments for teachers to use in kindergarten through second grade.[6]

Further, some states have recently updated or developed professional teaching standards as well as their early learning standards, prompted by the Common Core State Standards and the Race to the Top Early Learning Challenge. Many states are undertaking work to improve their longitudinal data systems, including how to link K-12 information to that from preschool. And they are pushing efforts to improve their lowest performing schools. As a result, some teachers are being assessed based on new evaluation tools while simultaneously implementing new curricula, lesson plans, assessments, and new intervention or improvement strategies.

Implementing all of these reforms well will take a significant amount of planning and resources at all levels, from classrooms to state departments of education. In the case of teacher-evaluation policies and the use of student data, policymakers should continue to ask themselves at what point the costs outweigh the gains. Substantial time and resources have already been invested to develop value-added models (see box on page 20), to create student achievement tests, to identify other tools as multiple measures of teacher performance, including observation tools, and to determine the appropriate weighting of each factor to accurately rate teacher performance and impact on student learning. Yet we do not know that this work will actually lead to improved teaching and learning. Groups of education researchers continue to question whether it even makes sense to use student growth data to evaluate teachers at any grade level.[7]

Still, states and school districts are sailing full speed ahead without much of a plan. That is why it is important to look now at the opportunities and risks of existing and emerging approaches for measuring student growth in the early grades, and for using this data in teacher evaluation systems. The PreK-3rd grades lay the foundation for a student's success throughout her years in school. The skill of teachers in these grades is critically important, especially for children who are receiving limited support at home for their cognitive and social development. Taking the right approach to evaluating the work of these teachers is essential to improving students' learning, not only in the early grades but also throughout their lives.

Notes on Methods

Research for this paper came from a review of studies on teacher evaluation and in-depth interviews with officials in five states and three school districts, selected based in part on conversations with experts in the field: Colorado, Delaware, Florida, Rhode Island, Tennessee, Austin (TX), Hillsborough (FL) and Washington, DC.

This paper does not delve into other important aspects of teacher evaluation, such as the promising use of observation tools to evaluate how teachers teach. For more on observation, see *Watching Teachers Work: Using Observation Tools to Promote Effective Teaching in the Early Years and Early Grades* (New America, 2011).[1]

This paper also does not explicitly address concerns with assessing young children, but others have[2] and New America is planning future papers to provide more direction to policymakers on this issue.

---

1 Lisa Guernsey and Susan Ochshorn, *Watching Teachers Work: Using Observation Tools to Promote Effective Teaching in the Early Years and Early Grades* (Washington, DC: New America Foundation, 2011), http://newamerica.net/sites/newamerica.net/files/policydocs/Watching_Teachers_Work.pdf.

2 National Research Council, *Early Childhood Assessment: Why, What, and How*, (Washington, DC: The National Academies Press, 2008).

Background

Leading up to the federal No Child Left Behind law, a growing body of research had begun to demonstrate that teachers are the most important school-based factors influencing student achievement—so much so that having three ineffective teachers in a row can have significant negative effects on a student's learning. One of the goals of NCLB was to solve the problem of inexperienced and out-of-field teachers being assigned to the most struggling schools with students, often poor and non-white, who need the most help.[8] The law requires that all teachers in Title I schools be "highly qualified."[9]

Following the institution of No Child Left Behind, education reformers began criticizing the highly-qualified teacher provision, saying it focused on the wrong issues. Among the critics was Michelle Rhee, then chancellor of the Washington, DC public schools. In a 2008 interview, she questioned whether it really mattered if you had a Ph.D. or a master's degree.[10]

The seminal 2009 paper *The Widget Effect*, released by The New Teacher Project, brought teacher effectiveness front and center, stating, "A teacher's effectiveness—the most important factor for schools in improving student achievement—is not measured, recorded, or used to inform decision-making in any meaningful way."[11] Discussions on how to improve teaching and learning shifted toward using student outcome data as a way to better identify and reward highly effective teachers, and to help less effective teachers improve or counsel them to leave the classroom.

The Obama Administration amplified the call for a focus on teacher effectiveness and student learning outcomes in developing its education agenda. In 2009, Arne Duncan announced the Race to the Top, a $4-billion competitive grant program established with funds from the American Recovery and Reinvestment Act. On teacher effectiveness, Race to the Top built upon pre-existing models—such as those in Tennessee—which tie students' growth to their teachers' evaluations.[12] (See box on value-added modeling, page 20.)

The Department of Education calls for states to "use multiple valid measures in determining performance levels, including as a significant factor data on student growth for all students."[13] The Department defines student growth as the change in achievement for an individual student between two or more points in time and specifies that for students in the untested grades and subjects, learning can be measured through "alternative measures of student learning and performance," as long as those measures are "rigorous and comparable across classrooms." Winning states have until the end of the 2013-2014 school year to use their federal funds to complete this work.

If it were not for the economic downturn and the fact that so many states were facing extremely tough budget decisions, the response to Race to the Top might not have been so enthusiastic and, in some cases, rushed.

If it were not for the economic downturn and the fact that so many states were facing extremely tough budget decisions, the response to Race to the Top might not have been so enthusiastic and, in some cases, rushed. Many state legislatures hurriedly passed laws that allowed teacher evaluations to be linked to students' achievement.[14] Initially, 11 states changed laws to be more competitive for Race to the Top. In all, since 2009, 36 states and DC have passed legislation or made policy changes to develop new or update existing teacher evaluation systems.[15]

The administration reinforced its interest in teacher evaluation by including similar guidelines in the latest round of the Teacher Incentive Fund (TIF) grant program. And again, in 2011, it did so when Secretary Duncan announced the administration's plan to grant states waivers from some of the more unpopular and unrealistic requirements of No Child Left Behind, such as the mandate that every student be "proficient" in math and reading by 2014. To qualify,

states had to make a firm commitment to improve their teacher evaluation systems, including multiple measures and incorporating student growth as a significant factor of teacher performance.

Some states have used these federal initiatives as an opportunity to think about what new teacher evaluation systems should look like. They are considering challenges such as how to conduct more rigorous observations; how to train those classroom observers; how best to use student growth and/or achievement to measure teacher effectiveness; how to assess the impact of an individual teacher on the collective growth of a school; and how to use professional development to help educators improve their practice. Other states have hastily adopted plans that, while possibly improving upon current systems, raise concerns of their own. One of those big concerns is what kind of student learning outcomes will be used to evaluate teachers of pre-kindergarten, kindergarten, and the early grades.

Teacher Evaluation: Significant Recent Events[1]

June 2009	Publication of *The Widget Effect*, a report exposing problems in current methods for evaluating teachers
August 2009	Michelle Rhee, chancellor of the District of Columbia Public Schools, launches IMPACT, a new evaluation system
January 2010	Deadline for first applications for Race to the Top, the U.S. Department of Education program that includes a focus on reforming teacher evaluation
August 14, 2010	*The Los Angeles Times* publishes teachers' "value-added" scores
September 2011	Obama administration announces plan to offer states flexibility under No Child Left Behind law
January 2012	Race to the Top problems in Florida, Hawaii, and New York
February 28, 2012	*New York Post* uses "value-added" data in revealing city's "worst teacher"
September 10, 2012	Chicago Teachers' Union strike
January 8, 2013	*Measures of Effective Teaching* report published by Gates Foundation
January 16, 2013	Florida Teachers' Union argues new education reform law is unconstitutional
January 17, 2013	NYC misses deadline for submitting evaluation system to state
January 19, 2013	Los Angeles Teachers' Union reaches agreement on the use of students' standardized test scores in their evaluations

1 Simone Pathe and Jaywon Choe, "A Brief Overview of Teacher Evaluation Controversies," *PBS Newshour*, February 4, 2013, accessed on February 19, 2013 at http://www.pbs.org/newshour/rundown/2013/02/teacher-evaluation-controversies.html; Politics K-12 blog at *Education Week*.

The Untested PreK-3rd Grades

The recent teacher evaluation reform movement has called for the inclusion of multiple measures, including measures of student growth for all teachers. Yet determining growth measures for these grades is among the most complex pieces of teacher evaluation reform.[16] The following four questions provide insight into why.

What Does Achievement Look Like in Young Children?

The rate of development and growth from birth through age eight is faster than at any other time, though not all children will develop or learn a new skill at the same time.[17] All children, but especially young children, pre-K through third grade, need more concrete instruction and opportunities to explore by engaging in hands-on activities.[18]

At this stage, children's growth—their acquisition of physical, cognitive, and social-emotional skills, their base of general knowledge, their strength of persistence and motivation, and their abilities in language and literacy—is directly linked to their academic growth. For example, a student's ability in later grades to focus, pay attention, and persist with challenging tasks is strengthened by his or her executive function (paying attention, completing tasks, controlling impulses, sharing with classmates, waiting their turn, etc.) and social-emotional development during early childhood.[19]

Standardized Assessments K-2

In pre-K and the early grades, there are different types of standardized assessments used:

- Direct assessments are administered directly to the child. For example, a teacher listens to each of her students read a passage to determine their reading level, examining the students' accuracy and the types of errors made.
- Observational measures are conducted during a specific activity. The teacher uses a rubric or checklist to determine whether a child demonstrates specific skills during an allotted timeframe.
- Authentic assessments are observations that are conducted during the regular flow of the day.

In *Developing Kindergarten Readiness and Other Large-Scale Assessment Systems: Necessary Considerations In The Assessment Of Young Children*, author Kyle Snow, senior scholar and director at the National Association for the Education of Young Children explains that an assessment should be administered in the same way, every time it is given.[1]

> "For example, an item on a direct mathematics assessment may have been developed to allow children to use some manipulable (e.g., counters, blocks) to help solve a basic addition problem. When conducting this assessment, all children should have access to the appropriate manipulable. Likewise, in a direct assessment, even if in the estimation of the assessor the child guesses the correct response, the response given is the one accepted, and the scoring protocol for the assessment would likely (if it is psychometrically sound) account for some proportion of correct guesses."

For more on early childhood assessment refer to the resources listed on page 10.

1 Kyle Snow, "Developing Kindergarten Readiness and Other Large-Scale Assessment Systems," National Association for the Education of Young Children, December 2011, accessed January 18, 2013 at http://www.naeyc.org/files/naeyc/file/research/Assessment_Systems.pdf.

An effective teacher in the early grades contributes to students' learning in all of these developmental domains. All teachers should regularly observe and assess students to help differentiate instruction to meet each child's needs in all these areas.

If the purpose of teacher evaluation systems is to measure the impact a teacher has on her students' learning, as well as to provide insight for teacher improvement, then assessment methods must account for how young children actually learn and measure more than just reading and mathematics. Such holistic assessments do exist for pre-kindergarteners and kindergarteners, but there are far fewer options for first and second graders. And those that do exist are not necessarily appropriate for teacher evaluation systems.

Are the Tests Really Telling Us What Children Have Learned?

Regardless of the kind of measure used, it is more difficult to obtain reliable and valid assessment data for young children—especially for high-stakes purposes—than it is for older children. Tests given at one point in time do not necessarily provide a complete picture of a student's knowledge and skills, and this is especially true in the early grades. Young children are not able to read and respond to test questions independently, and they have difficulty staying focused for an extended period of time. At these ages, the assessor—often the child's teacher—typically administers an assessment one-on-one by conducting observations of a student performing a specific task. Oftentimes these assessments are designed for purposes that do not require the same objectivity and standardization as grade 3-8 accountability tests.

Some early childhood experts suggest an altogether different kind of measure for early grade teachers. Elliot Regenstein, senior vice president for advocacy and policy at The Ounce of Prevention Fund, observes, "The nature of how you administer assessment in K-2 doesn't quite mesh with using assessment as a teacher accountability tool. The correct teacher accountability measure may not be 'how did your kids perform?' but did you administer the test appropriately and did you do something appropriate with the results?"[20]

Kristie Kauerz, program director for PreK-3rd education at the University of Washington, agrees. "In a perfect world," she says, "the variables included in a teacher evaluation system for the early grades would not just be on child performance, but would also include what kinds of data sources teachers are using and how teachers are using those data to inform their practice."[21] She notes that these practices are important not only for PreK-3rd teachers, but for teachers of every grade-level and subject.

Relying on Math and Reading Scores: Will Narrowness Distort Teaching?

States are primarily in search of measures for evaluating literacy and mathematics in the early grades—mirroring practices in grades 3-12—but this will not allow for a full picture of a young child's learning or his teacher's impact.

Among early childhood experts, there have been concerns about the use of literacy assessments that focus on a limited range of basic skills—such as naming letters of the alphabet—often absent of context. These kinds of assessments can lead teachers to narrow what they teach to focus on the specific skills assessed.[22]

Among early childhood experts, there have been concerns about the use of literacy assessments that focus on a limited range of basic skills—such as naming letters of the alphabet—often absent of context.

The same can be true for entire subject or topic areas. *The 2012 Common Core report, Learning Less: Public School Teachers Describe a Narrowing Curriculum*, found that 81 percent of elementary school teachers felt that "other subjects were getting crowded out by extra attention being paid to math or language arts."[23] While the survey only included teachers in grades 3-12, it is reasonable to assume that the same consequences could occur in the early grades for the same reasons. In early grade classrooms, the emphasis placed on teaching reading and math could pressure teachers to reduce time for learning centers, investigative activities, play, and socialization, along with subject areas beyond literacy and numeracy.

How Can We Account for the Different Delivery Models of Preschool?

The number of children enrolled in state-funded pre-kindergarten programs for three- and four-year-olds continues

to grow. Some state-funded pre-K programs are housed in elementary schools while others are part of a local child care center. Most pre-K teachers employed by school districts and working in public schools are included in teacher evaluations. Still, there are issues to consider. Depending on the state, pre-K teachers do not necessarily need the same degrees or certification as K-12 teachers. In some states, pre-K is a half-day program; in others, it is a full-day program. And in some states there are both half-day and full-day pre-K programs. (This is true for kindergarten as well.) Additionally, pre-K classrooms typically have low teacher-student ratios, and sometimes, students are taught in a co-teaching environment. These factors make it difficult to tie students' growth to an individual teacher. Of the states reviewed, there is variation in how pre-K teachers are subject to teacher evaluation requirements.

Guidance from the Early Childhood Field for Assessing Young Children

- "Those charged with selecting assessments need to weigh options carefully, considering the appropriateness of candidate assessments for the desired purpose and for use with all the subgroups of children to be included. Although the same measure may be used for more than one purpose, prior consideration of all potential purposes is essential, as is careful analysis of the actual content of the assessment instrument." —From the National Research Council's *Early Childhood Assessment: Why, What, and How?*

- "A one-time snapshot of a child entering a kindergarten classroom cannot capture all of the cumulative experiences in programs, in the home, and in the community of a young child from birth to that day in kindergarten." —From *Developing Kindergarten Readiness and Other Large-Scale Assessment Systems: Necessary Considerations for the Assessment of Young Children*

- "Children should be assessed using age-appropriate methods on all domains of early learning and development." —From a report from the National Association of Elementary School Principals Foundation Task Force on Early Learning

- "Assessment tools should be chosen that have been shown to have acceptable levels of validity and reliability evidence for the purposes for which they will be used and the populations that will be assessed." —From the National Research Council's *Early Childhood Assessment: Why, What, and How?*

- "The younger the child, the more difficult it is to obtain reliable and valid assessment data. It is particularly difficult to assess children's cognitive abilities accurately before age 6." —From *Principles and Recommendations for Early Childhood Assessments*, submitted to the National Education Goals Panel

---

Resources on assessment:

Taking Stock: Assessing and Improving Early Childhood Learning and Program Quality: http://ccf.tc.columbia.edu/pdf/task_force_report.pdf

Developing Kindergarten Readiness and Other Large-Scale Assessment Systems: Necessary Considerations for the Assessment of Young Children: http://www.naeyc.org/files/naeyc/file/research/Assessment_Systems.pdf

National Association of Elementary School Principals Foundation Task Force on Early Learning Report: http://www.naesp.org/transforming-early-childhood-education-pre-k-grade-3

Principles and Recommendations for Early Childhood Assessments: http://govinfo.library.unt.edu/negp/reports/prinrec.pdf

Early Childhood Assessment: Why, What, and How?: http://www.nap.edu/catalog.php?record_id=12446

National Institute for Early Education Research: http://nieer.org/research/assessment

Experiments in Using Student Growth in the Early Grades to Evaluate Teachers

In order to meet new regulations and laws, states and school districts have moved ahead with plans to measure student growth in the early grades, often with limited deliberation on the challenges discussed in the previous section. There has not been much time for small experiments or multiple pilot projects to see what works. Most states are using one, or some combination, of three approaches: student learning objectives (SLOs), identifying or creating new assessments, and shared attribution. Each one carries different risks and opportunities.

1. Student Learning Objectives

An SLO has three components: measurable student goals or objectives; a growth target that is set with the baseline performance of the students in mind; and a specific assessment or tool to measure student' progress toward or achievement of the target. While teachers can base SLOs on student performance targets as measured by externally developed standardized tests, SLOs can also be based on district-created or teacher-developed assessments. According to the Department of Education, the ideal assessment is one that is both "rigorous and comparable across classrooms."[24]

SLOs are taking hold in many school districts across the country. In fact, more than 2,000 districts in 20 states are using SLOs to measure students' learning in the untested grades and subjects.[25] States adopting the practice include Colorado, Connecticut, Delaware, Georgia, Indiana, New Hampshire, New York, Ohio, and Rhode Island.

Generally speaking, at the beginning of the school year, the teacher and evaluator meet to discuss goals in one or multiple subject areas, select an assessment or tool to measure student learning growth, and establish an appropriate target for the whole class or a student cohort. SLOs can also be set by teams of teachers (organized by grade or subject area), by the school district, or by the state education agency.

The SLO approach may offer the best avenue for providing teachers with tools to inform their instruction, identifying

Variations in States' and Districts' Approaches for Measuring Student Learning in PreK-3rd Grades for Teacher Evaluation

States	Primary Method	Multiple Measures of Achievement Used?	Development of Assessment or Item Bank?
Colorado	SLO	More than one SLO	Yes
Delaware	SLO	Growth goal	Yes
Florida	Shared attribution or assessments	Not required	In progress*
Rhode Island	SLO	More than one SLO	Not at this time
Tennessee	Shared attribution or assessments	Raw achievement measure	No
School Districts	**Primary Method**	**Multiple Measures of Achievement Used?**	**Development of Assessment or Item Bank?**
Austin, TX	SLO	Individual and team SLO	Some items in a bank
Hillsborough, FL	Assessments	In most grade-levels more than one subject addressed	No
Washington, DC	SLO	More than one SLO	In consideration

** Florida's item bank is intended to be used by school districts to create subject-area or grade-level assessments if they choose to do so.*

The SLO Process from the Teacher Perspective

Imagine a first grade teacher, Maria, in a Title I school working with high-need children. After reviewing data on her students' progress in kindergarten and her observations from the first two weeks of school, Maria meets with her principal. They discuss the goals or objectives she has identified for her students in reading and writing, which are the two subject areas identified by the school as this year's priority for first grade. She plans to use a district-created assessment as the reading measure and a teacher-created rubric as the writing measure. She administers pre-tests to establish baseline data and determine the growth targets that make the most sense for her students. She sets the target that 80 percent of her students will advance two levels on the district reading assessment and that 80 percent of her students will score at least 3 out of 5 on the writing assignment. Maria's principal approves her plan goals, assessment choices, rubric, and achievement targets. Maria begins the process of monitoring children's progress toward the SLOs she has set.

effective and ineffective educators, and improving student learning. But there is limited research on SLOs, especially on their impact in early grade classrooms. Implementation of the approach is not without risks, including states and school districts' capacity to provide professional development, resources, and monitoring of the SLO process.

Opportunities

Fosters School-Level Collaboration and Shared Priorities

The SLO process is collaborative and can bring teachers in across grade levels to discuss student data, learning needs, and school priorities. SLOs can be applied to and used by all teachers and all subjects, yet at the same time they can be tailored to meet the specific goals of schools and needs of individual students. In some districts—such as Washington, DC—even when value-added teacher-level data exist, SLOs are used for all teachers as an additional measure of student learning.

Under Rhode Island's model, early grade teachers are required to write two SLOs. Teachers are expected to align their objectives with the most important content and standards for their grade, as well as with their administrator's priorities for the school. Teachers of the same grade level in the same school are encouraged to use the same set of objectives and assessment, which could include a portfolio scored by rubric. Specific targets, though, may vary if student starting points differ substantially across classes.

In Austin, all teachers complete an individual *and* a team SLO. In the early grades, for example, second grade teachers would work together and with the principal to identify an area of need for their students and craft an appropriate objective, select a common assessment, and determine a rigorous, attainable target that all second graders should be able to meet.

The Rhode Island Department of Education (RIDE) created a teacher evaluation model that includes SLOs. For preK-2nd grade, no value-added measure of student growth is used. The SLO accounts for the full 50 percent of the growth component. Districts can choose to adopt the RIDE model or to submit one of their own for approval. Pre-K teachers who are part of public school programs are required to complete SLOs, but are not measured under any other component of the state's teacher evaluation system.

In Washington, DC, teachers set objectives in collaboration with their principals. At the beginning of the school year, principals meet with the DCPS chancellor to set school goals, one of which must be focused on student achievement. After these meetings, teachers and principals meet to discuss how the schools' goals will guide the objectives, assessments, and targets for the teachers' SLO, referred to as a Teacher-Assessed Student Achievement goal (TAS

How IMPACT Evaluation Works

Washington, DC Public Schools' IMPACT separates teachers into categories based on whether value-added data are available. Teachers of pre-K (three- and four-year-olds) through third grade fall into what DCPS designates as "Group 2a" because their students do not participate in the statewide standardized tests called DC-CAS. Teachers in non-tested grades and subjects are evaluated based on the following measures:

- 75% classroom observation
- 15% teacher-assessed student achievement data (TAS)
- 10% commitment to the school community

goal) in DC. According to Sam Pearcy, teacher effectiveness coordinator for DC's evaluation system known as IMPACT, principals could decide to focus on one area that year—say, improving students' writing—and use the evaluation system as a lever to drive that initiative. More often, principals give their teachers a schoolwide framework in which to build their goals, allowing flexibility for teachers to determine goals that best capture the different topics or objectives of a particular course. "When there is a push to a schoolwide initiative," Pearcy said, "it is usually one standard goal and teachers add on more individualized goals."[26]

Focuses on Improving Instruction

SLOs provide a formal process for instructionally- focused conversations between evaluators and teachers as well as with grade level teams and resource teachers. Austin Independent School District officials deliberately chose not to allow teachers to select state tests as a measure for SLOs.[27] Teachers can use state results to inform their goals, but they cannot use those results as their assessment. Joann Taylor, assistant director of strategic compensation in Austin, says that the district "really wanted teachers to think about what was happening in their classrooms and to have that power to say that these are the things that I know that my students need to accomplish by the end of the year."[28]

At the beginning of each school year in Austin, teachers conduct a needs assessment of their classes. Since PreK-2nd grade teachers do not have state test data, they use observation and anecdotal evidence they have collected during the first three weeks of school to inform the writing of their individual SLOs. After completing the needs assessment, teachers consider which state standards they should focus on. Teachers essentially take a state standard and turn it into a statement of what their students will be able to do by the end of the year: *My kindergarteners will be able to write two legible sentences on a specific topic by the end of the year.*

At the beginning of each school year in Austin, teachers conduct a needs assessment of their classes. Since PreK-2nd grade teachers do not have state test data, they use observation and anecdotal evidence they have collected during the first three weeks of school to inform the writing of their individual SLOs.

Can Help Teachers Better Meet Individual Student Needs

In Rhode Island, teachers are expected to set tiered targets according to students' baseline data. Baseline data could include prior year assessment scores or grades; beginning-of-year assessments; or other evidence such as prior work samples. For example, students who begin below grade level would be expected to make substantial progress toward objectives by the end of the year, though they might not be rated as proficient. Meanwhile, students who begin on grade level would be expected to meet or exceed proficiency by the end of the school year. Tamika Pollins, of Rhode Island's Office of Educator Quality and Certification, says SLOs help to "capture learning for all children. Allowing tiered targets lets us make sure no child is being left behind."[29]

In Austin, teachers construct targets based on student

Example of an SLO Target for a Second Grade Teacher

Baseline Data: I am a second grade teacher. When I received my course roster, I used my students' first grade end-of-year Fountas & Pinnell reading level scores to identify ability groupings within my class. I found that four students were reading below grade level, 15 were on grade level, and five were above.

Targets:

1. The four students who are reading below grade level will move up at least three reading levels. Students at Level H will move to Level K or better, the student at Level G will move to Level J or better, and the student at Level I will move to Level L or better (H ⇨ K, G ⇨ J, I ⇨ L).

2. The fifteen students who are reading on grade level move up at least three levels to reach proficiency with Level M (or higher)texts.

3. The five students who are reading above level will move up at least three reading levels to reach proficiency with Level P (or higher) texts.

Rationale for Targets: I know that most students can achieve three levels of growth on the Fountas & Pinnell scale because 90% of my students moved up at least three reading levels last year. I used baseline data to establish students' starting points and then set individualized targets for students who needed to reduce the gap between their knowledge and grade level proficiency. For the remaining students, I set a goal for them to improve at least three levels by the end of the year.

Source: Rhode Island Measures of Student Learning, Edition II: http://www.ride.ri.gov/Portals/0/Uploads/Documents/Teachers-and-Administrators-Excellent-Educators/Educator-Evaluation/Education-Eval-Main-Page/Measures-of-Student-Learning-GB-Edition-II.pdf.

data and performance on the pre-test they have either identified or created to measure the objective. If student scores are clustered in the same general range, teachers construct a single growth target. If student scores are spread over a wide range, teachers construct a multi-tiered growth target.[30]

Could Support a More Well-Rounded Curriculum

While most states and school districts specify that SLOs should be aligned with state standards, they are not attached to a specific assessment, such as the state's reading and math assessments. So it could be possible for teachers to measure students' progress in a wide range of subject and developmental areas. For example, in Washington, DC, kindergarten teachers are encouraged to create objectives for students' pre-reading, reading, writing, and math skills, but teachers and administrators could choose to include SLOs for social studies and science as well. For pre-K teachers, the district recommends using *Teaching Strategies GOLD* as the assessment to measure their SLOs. GOLD is an assessment that measures seven domains of children's development: social-emotional, gross motor, fine motor, language, cognition, literacy, and numeracy. GOLD is a teacher-administered, observation-based assessment system for children from birth through kindergarten that combines ongoing assessment and performance tasks.

It is important to note, though, that principals or teachers in DC are not required to include all of the domains, nor are they required to use GOLD at all. For example, a principal could decide he wants pre-K teachers to focus on the literacy and numeracy components only.

Attains Teacher Backing

Teachers may see more value in the SLO process because, depending on the assessments used, data can be immediately relevant to lesson planning and instruction. Because SLOs require using and understanding student data, creating or selecting assessments, and setting growth targets, they can help expand teachers' skill sets.

Risks

Resource-Intensive

The effective use of SLOs can be resource-intensive, and both districts and schools need to develop the necessary capacity.

Austin, for example, has developed extensive materials and has three staff members dedicated to supporting teachers in the development of SLOs. These professionals visit school campuses during the first two months of school, helping teachers think through the SLO process and providing training for teachers on key components, including designing and using formative assessments and interpreting data. The district team also leads monthly SLO meetings for principals and provides ongoing training to help them understand their role in approving and monitoring teachers' SLOs.

How to track and monitor SLOs is an important consideration. In many cases, this requires new technology.

Washington, DC Public Schools also has a team dedicated to IMPACT that reviews teachers' SLOs for feasibility and helps with revision. The team also serves as a resource throughout the year, answering questions and providing training sessions to guide teachers through the process.

How to track and monitor SLOs is an important consideration. In many cases, this requires new technology. Rhode Island developed a statewide, computer-based system to assist teachers and evaluators in collecting, managing, and approving SLOs. Austin and Washington, DC have similar systems. In Austin, teachers enter all of the information pertaining to their SLOs into the district's web-based SLO database.[31] Once SLOs have been entered, principals can approve them or return them to a teacher for revision. Then the objectives are submitted via the database to the district SLO team for final approval.

It is also important to embed good SLO practices into campus culture, making them part of a continuous improvement strategy. Taylor says this is a key job for the principals in Austin. If they do not see SLOs as a vehicle to improve teacher practice and student learning, SLOs will not work as intended. "Where principals are lax about the systems and don't really monitor," she says, "that's where we get more reports of things that aren't necessarily on the up and up."[32]

Limited Expertise

School districts will need to consider how to expand teachers' and principals' knowledge about high-quality assessments. The National Council on Teacher Quality (NCTQ) reviewed assessment coursework in 180 teacher preparation programs.[33] According to NCTQ's findings, only 21 percent of the programs adequately cover assessment literacy. Less than 1 percent were found to adequately explain to prospective teachers how to analyze data obtained from assessments and less than 2 percent of programs explained how to use those data to inform instruction.

This lack of expertise has been a concern for Austin and DC. In Austin, once a teacher has determined her objectives, she must find adequate assessments. She can create an assessment of her own, use an existing external assessment, or use one of the district-created common assessments.[34] "In the first year of doing this we learned that teachers aren't really sure how to create a good assessment," admits Taylor.[35] The district has worked to build teachers' assessment literacy, bringing in experts to provide professional development during the summer. In DC, for example, if teachers create assessments for their SLOs, principals must approve them. Principals must then know how to identify a high-quality, rigorous assessment.

In addition to providing professional development for both principals and teachers to build their facility with assessments, districts may also need to invest in instructing teachers and principals in how to use data appropriately, construct clear and measurable objectives, and set rigorous but achievable targets.

In addition to providing professional development for both principals and teachers to build their facility with assessments, districts may also need to invest in instructing teachers and principals in how to use data appropriately,

construct clear and measurable objectives, and set rigorous but achievable targets.

Another challenge can be found in the pre-test/post-test design used in many SLO systems. In a 2012 paper,[36] Scott Marion of the Center for Assessment, Inc., with his co-authors, examines problems with using pre- and post-tests to determine student growth and to compare teachers, such as "treating non-equated tests as if they shared the same score scale."[37] Marion says there are ways to use growth measures to evaluate SLOs, but he cautions against doing so without consulting assessment experts first.[38]

Inability to Compare Teachers

To ensure that teachers are evaluated fairly across a school and district, it is important to consider how uniform and comparable the SLOs may be. Given the individualized nature of SLOs, this is difficult. In Rhode Island, Pollins says it is the state's recommended "best practice" for districts to use common objectives and measures across classrooms at the same grade level or in the same subject area.[39] There is nothing in Rhode Island's regulation, however, that requires districts to do this.

To ensure that teachers are evaluated fairly across a school and district, it is important to consider how uniform and comparable the SLOs may be. Given the individualized nature of SLOs, this is difficult.

Taylor says that even with all the infrastructure Austin has built, comparability is still a concern for the district. It has tried to mitigate this problem by providing significant guidelines on SLOs, assessments that meet the district's criteria, and appropriate targets. The district SLO staff looks at every SLO to make sure it meets quality standards. Taylor says this is a critical piece of the program: "We really have thought about what is the difference between a sample versus the whole, and we have found that sampling doesn't really give us the consistency that we are looking for, so we have kept with that process of reviewing every SLO."[40]

Pearcy says there is limited availability of valid, reliable, and appropriate assessments at certain levels, especially in the PreK-3rd grades, in DC. "Where we have off-the-shelf assessments we are able to offer guidance around targets that are appropriate." He notes, though, that this is much more difficult to do for teacher-created assessments.[41]

High Potential for Manipulation

Teachers are highly involved participants in the SLO process. They work with their principals to set goals and targets for their students. Teachers identify or even create assessments to measure student progress on meeting the targets. And in the early grades, they are most likely administering those assessments as well. There are concerns about whether it is appropriate for teachers to play such a significant role in the evaluation process when they know the results will be tied to high-stakes consequences, which could include compensation and continued employment.

Bringing Uniformity to SLOs

Winston Churchill called democracy "the worst system except for all the others." Scott Marion, of the Center for Assessment, Inc., says the same is true for student learning objectives (SLOs) as a method of teacher evaluation. SLOs bring a real opportunity to improve educator practice and outcomes for students because they empower teachers to think deliberately about student learning and growth. If they are implemented without attention to the potential risks, though, the new system will amount to little more than the old system.

To try to address the challenges of comparability among teachers and assessment validity, states can establish centralized assessment banks. An "open" assessment bank allows teachers to share any assessments they are currently using with others across the state. A "vetted" assessment bank houses assessments that have been identified and reviewed by groups of teachers, state officials, and sometimes testing experts to determine whether or not the measures are valid and appropriate for the grade level or subject area for which they are recommended.

Delaware and Colorado are two states creating assessment banks to assist in the use of SLOs for teacher evaluation. DC and others plan to move in this direction too.

In the next several years, current and future adopters of the SLO approach will likely take similar steps to bring more consistency to the objectives and assessments used.

Delaware's Teacher Evaluation System: Measures for the Student Improvement Component

MEASURES: / ROLE:	**Measure A: DCAS**. Instructional scale scores for reading and/or mathematics in grades three (3) through ten (10).	**Measure B: Internal/External.** Measures developed or identified by educator groups across the state and approved by DDOE.	**Measure C: Growth Goals**. Common goals developed or identified by educator groups across the state and approved by DDOE.
Group 1: Includes any educator who instructs reading and/or mathematics in DCAS grades three (3) through ten (10).	50%	50%	
Group 2: Includes any educator who generally reports student grades in any subject or grade where DCAS reading and math is not administered and/or a Measure B asessment is available.		50%	50%
Group 3: Includes any educator who generally does not report student grades and any educator who cannot otherwise be cagegorized into Groups 1 or 2.			100%

Delaware's K-2 teachers fall into Group 2 and pre-K teachers fall into Group 3.

Source: Delaware Performance Appraisal System II Guide Revised for Teachers, http://www.doe.k12.de.us/csa/dpasii/specialist/DPASIISpecFullGuide.pdf.

Delaware

After receiving one of the first federal Race to the Top grants in 2010, the Delaware Department of Education began to evolve its existing teacher evaluation system to incorporate multiple measures of student growth. Delaware was already using student learning objectives as part of teacher evaluation, but to add some standardization to the process, it decided to create an assessment bank for teachers to use when selecting assessments for their SLOs, according to Diane Donohue, special assistant for educator effectiveness.[42]

For the "student improvement" component of Delaware's teacher evaluation system, teachers are grouped in one of three ways, depending on whether or not their students participate in the Delaware Comprehensive Assessment System.[43] Early childhood teachers, birth through pre-K,[44] who are hired by public school districts and funded by IDEA, Title I, or Delaware's state-funded pre-K program, fall into "Group 3."[45] Kindergarten through second grade teachers fall into "Group 2" as explained in the chart above.

For the "Growth Goals" measure, the Delaware Department of Education brought a group of early childhood educators together to identify specific goals for early education. These goals incorporate the eight developmental domains included in Delaware's Early Learning Guidelines. Under the guidance of the Department of Education, educator groups have also identified or created about 200 different assessments that can be used for measure B and C.

Colorado

To help districts, schools, and teachers to identify fair, valid, and reliable assessments of student learning outcomes, the Colorado Department of Education has created "content collaboratives" for all subject areas. The members of the content collaboratives include researchers and educators from diverse districts and grade levels, including teachers from the early grades of elementary school. These individuals are charged with identifying high-quality assessments for all grades and content areas, using principles and an "assessment review tool" established by a technical steering committee. An assessment is rated based on whether it: aligns with the Colorado Academic Standards; has rigorous and clear scoring criteria; is fair and unbiased for all; and engages students in authentic situations that can be applied to other content areas and contexts. In the future, the collaboratives may also create assessments.

An Example of Delaware's SLO Process from the Teacher Perspective

Imagine a first-grade teacher named Bradley in a Delaware classroom with 20 students. Because his students do not yet participate in the Delaware Comprehensive Assessment System (DCAS), he falls into "Group 2," which means for the student improvement component of his evaluation, 50 percent is based on his students' growth on an alternate assessment (measure B) that he selects and the other 50 percent is based on a specific growth goal (measure C).

At the beginning of the school year, Bradley meets with his principal to discuss assessments and targets for both sets of measures. He has to select four measures in total; at least one must fall under measure B. Basing his selections on what he already knows about his students, Bradley accesses Delaware's "online shopping mall" of assessments. Bradley also has the option of selecting an externally approved assessment. If Bradley does not find an assessment he thinks would be appropriate, he can request to use a different one, but doing so would require approval from both his evaluator and the state.

For measure B, Bradley selects the Developmental Reading Assessment (DRA) and a teacher-created math assessment. The pre-test is given to all students, but Bradley can choose to narrow the focus of his target to a cohort of students, such as his 10 lowest-performers as measured by a pre-test. He could also set different targets for various cohorts of students all based on the same assessment. At the end of the year, he will administer the post-tests.

For measure C, Bradley must select at least one growth goal in either English Language Arts or Math, or he can select one from each subject. The first grade growth goals are one set of 46 in total. Bradley and his principal could agree to a goal that Bradley writes on his own if it is more appropriate for his students' needs. Bradley decides to select a reading fluency goal and a math goal:

- 80 percent of the targeted group of students will show improvement in fluency in the last six weeks of school as compared to the first six weeks of the school year.
- The number of targeted students who demonstrate an understanding of place value will increase by 30 percent based on the baseline data in the spring.

Bradley discusses his selected growth goals and the tools used to measure student growth with his principal, and together they determine appropriate targets based upon his students' needs.

Assessments that have been evaluated—along with their ratings and reviews—are placed in the online resource bank as optional tools for districts to use as they develop plans to meet the state's student learning outcomes measure requirement. Districts have the option to use other assessments and district officials can vet them using the same assessment review tool that is being used by the collaboratives.

In Colorado, pre-K teachers are included in the evaluation system if they have a state teaching license. The state does not require teachers who are part of the Colorado Preschool Program to have a state teaching license. Head Start teachers are not required to be licensed either. Early Childhood Special Education teachers, however, are required to have a state teaching license.[46]

School districts can, however, establish their own requirements for pre-K teachers and some do require they obtain a state teaching license. In these cases, pre-K teachers are required to be evaluated.

For kindergarten, the Department of Education is trying to use assessments that teachers are already administering to children. Tara Boertzel-Schuenemann, student growth consultant in the Educator Effectiveness Unit, says that various state laws[47] already require kindergarten teachers to provide a variety of different information on their students under

A Note on Terminology

In Colorado, the use of the term "license" has caused some confusion. Pre-K programs are licensed by the Department of Human Services. But individual teachers in those programs are not required by the State of Colorado to obtain an individual teaching license. (Districts may require pre-K teachers to have a teaching license.) At a recent state meeting of educators, some believed that all pre-K teachers were required to be evaluated because they work in a "licensed" program. This is not the case, and Colorado Department of Education officials say they are working to make sure the distinction is clear.[1]

1 Interview with Tara Boertzel-Schuenemann, student growth consultant for the early childhood education work group, Colorado Department of Education (November 20, 2012).

Colorado's Achievement Plan for Kids (CAP4K) and the Colorado Read Act. "If we can allow assessments to serve multiple needs, it means less testing for children and it makes it easier on the teachers too," says Boertzel-Schuenemann.[48] The department recommends that for pre-K and kindergarten, multiple domains of learning be included in the measures of student outcomes. For example, among suggested assessments are ones currently being used to assess school readiness. This raises concerns about whether these assessments are appropriate for use in teacher evaluation. Researchers caution that these tools should only be used for the purposes for which they are developed and validated.

Learning from Colorado and Delaware

By creating assessment banks, Colorado and Delaware have made it easier for districts to identify high-quality evaluative tools and for teachers to select assessments that align with specific objectives they plan to measure. Additionally, Delaware has taken steps to provide more standardization through its suggested growth goals and assessment bank, while still allowing teachers flexibility to meet the specific needs of their students.

Delaware and Colorado have tried to standardize by encouraging the use of assessments that have already been vetted. Identifying and/or creating assessments to fill the banks can be a labor-intensive process. Delaware teacher groups have identified or created nearly 230 tests so far, from pre-K to 12th grade. Too many options, though, can be confusing, and teachers will surely need support in identifying the most appropriate measures of their students' growth. State officials recognize the need for ongoing review to remove and replace assessments that are not working.

Delaware officials say they are aware of issues that can arise from testing for accountability purposes, such as teaching to the test and narrowing the curriculum. Ruszkowski says, "If administrators see teaching to the test or cheating, we expect them to step up and reflect it in the appropriate areas of the evaluation such as quality of instruction or professionalism."[49]

By creating assessment banks, Colorado and Delaware have made it easier for districts to identify high-quality evaluative tools and for teachers to select assessments that align with specific objectives they plan to measure. Additionally, Delaware has taken steps to provide more standardization through its suggested growth goals and assessment bank, while still allowing teachers flexibility to meet the specific needs of their students.

In Colorado, using the assessments in the bank is not required either. To help districts select good assessments, the state's content collaboratives created a detailed assessment review tool and made it available on its website to help districts evaluate other assessments.

Bringing uniformity to SLOs may be the course others will follow. For example, eventually DC plans to move away from teacher-created assessments altogether.[50] DCPS eventually plans to follow the wake of states like Colorado and Delaware by providing a bank of assessments that have been vetted by experts for validity, reliability, and appropriateness.[51]

2. Creating or Identifying New, Shared Assessments

Other states and districts are identifying or creating new assessments to measure student growth or achievement, often using value-added models to determine students' growth between two points in time. (For more on value-added modeling, see box below.) Hillsborough County, Florida and the state of Tennessee are two examples of places expanding the standardized-assessment coverage to every grade and subject area.

In the Hillsborough County Public Schools (HCPS), all teachers are evaluated using value-added measures. In 2007, Hillsborough developed a series of end-of-semester and course exams as part of Florida's pay-for-performance initiative. In 2010, as one of seven districts selected to participate in the Gates Foundation's Measures of Effective Teachers (MET) project, the county identified or created assessments for then-untested grades and subjects, from kindergarten through 12th grade. This school year, HCPS added an assessment to measure the effectiveness of teachers in public pre-K programs.[52] Forty percent of a teacher's evaluation is based on her students' growth on these assessments, as measured by Hillsborough's value-added model. The district worked with the University of Wisconsin Value Added Research Center to develop a customized value-added model to measure all grades and subject areas.[53]

Tennessee is another state that has opted to identify an assessment for every grade and subject area. An early Race to the Top winner, it was one of the first states to make substantial changes to its teacher evaluation system. In 2010, the Tennessee legislature passed the First to the Top Act, which required 50 percent of the evaluation to be comprised of student achievement data for all teachers, 35 percent of which must be based on student growth as measured by the

On Value-Added Modeling (VAM)

Value-added models measure a teacher's contribution in a given year by comparing the end-of-grade test scores of his students to the scores of those same students in the previous school year, as well as to the scores of other students in the same grade. Students' past test scores are used to predict future test scores, assuming that students usually score approximately as well each year as they have in past years. Actual scores are then compared to the predicted scores and the difference between those scores is considered the teachers' "value-add." Some VAMs control for factors such as a student's past performance, socioeconomic status, race/ethnicity, teacher years of experience, and school size. But not all do so; which variables are included depends on the model. Controlling for out-of-classroom variables can increase the fairness of the model. Additionally, the more years of data used for a particular teacher, the more accurate the data will be. As Matthew DiCarlo, senior research fellow at the Albert Shanker Institute, writes, "Well-designed value-added models can, on the whole, go a long way toward controlling for the many test-influencing factors outside teachers' control, to no small extent because prior achievement helps pick up on these factors. But it is inevitable that even the best models will penalize some teachers and reward others unfairly (this is true of almost any measure)."[1]

Tennessee pioneered the use of a "value-added model" (VAM) in the mid-1990s to track the educational value provided at the classroom, school, and district levels by teachers and schools. William Sanders, then a statistician at the University of Tennessee-Knoxville, developed the Tennessee VAM known as TVAAS. It is still used today as a component of Tennessee's updated teacher evaluation system. There is much debate on whether, and to what extent, value-added data should be included as a measure in teacher evaluation.[2]

---

1 Matthew Di Carlo, "On Teacher Evaluations, Between Myth And Fact Lies Truth," *Shanker Blog: The Voice of the Albert Shanker Institute*, accessed April 29, 2013, http://shankerblog.org/?p=8093.

2 Eva L. Baker, Paul E. Barton, Linda Darling-Hammond, Edward Haertel, Helen F. Ladd , Robert L. Linn, Diane Ravitch, Richard Rothstein, Richard J. Shavelson, and Lorrie A. Shepard, "Problems with the Use of Student Test Scores to Evaluate Teachers" Economic Policy Institute, 2010, accessed April 25, 2013 at http://www.epi.org/page/-/pdf/bp278.pdf.

Creating or Identifying New, Shared Assessments: Examples from Florida

In Florida's Duval County Public Schools K-2 teachers will be rated on their students' performance on the teacher-administered Florida Assessments for Instruction in Reading (FAIR).[1] Osceola County Public Schools is also using FAIR for first and second grade. Already some educators have concerns. A second grade teacher in Florida shares her frustrations with FAIR: "The FAIR is given three times throughout the year. Students are measured on their probability of reading success based on their reading of a passage. If a student scores at the highest level available for second grade the first time the assessment is given there is no way to measure growth. FAIR does not allow students to read third-grade passages. The high-scoring student reads the same passage at each assessment point."[2]

Other Florida districts are creating or identifying new assessments as well. As part of Florida's Race to the Top grant, the state is creating assessments for hard-to-measure content areas such as music and art.[3] Additionally, the Florida Interim Assessment Item Bank and Test Platform project—also part of Race to the Top—will create K-2 assessment items. Michelle Worrell, project manager for the item bank and test platform, asserts that, "We are not, however, creating assessments. Districts will be able to access the items within the IBTP to create district assessments for K-2 should they choose to do so."[4]

1 Duval County Public Schools, "Collaborative Assessment System for Teachers (CAST)," accessed on December 12, 2012 at http://www.duvalschools.org/static/wearedcps/joindcps/cast.asp.

2 Interview with anonymous second grade teacher (March 5, 2013).

3 Florida Department of Education, "Race to the Top Assessments," accessed on December 12, 2012 at http://www.fldoe.org/arra/racetothetop/assessments/.

4 Email interview with K. Michelle Worrell, project manager, Item Bank & Test Platform, Office of Race to the Top Assessments, Florida Department of Education (February 7, 2013).

Tennessee Value-Added Assessment System (TVAAS) or a comparable measure. The other 15 percent must be based on a raw measure of student achievement adopted by the State Board of Education. The department has identified a number of suitable "off-the-shelf" assessments.[54]

Opportunities

Facilitate Comparisons

Districts that require the use of specific assessments have an easier time comparing teachers across the district and even the state. Additionally, requiring specific assessments allows states or districts to identify measures that are aligned with state standards. After meeting to discuss appropriate assessments with early grade educators from across the state, education officials in Tennessee decided to allow districts to use the Stanford 10 (reading and language arts, math and science) to create growth scores for teachers in first through third grades.[55] The department of education also intends to develop a first and second grade state assessment to align with the Common Core assessments for use beginning in 2015.[56]

Could Build Teachers' Skills

When new tests are created by the state or district, teachers are often invited to participate, potentially increasing their support for the effort and honing their abilities to create and identify good assessments. Many of the assessments used by Hillsborough Schools are teacher-created. District officials saw developing tests internally as a way to build teacher capacity.[57] Hillsborough Schools provides extensive training for teachers serving as item writers for the district tests, helping them to identify characteristics of a good test and to write high-quality test items. HCPS officials recognize that teachers take these skills back to their own schools, using them in the classroom as well as sharing them with colleagues.

Risks

Requires Financial and Time Resources

The process of identifying appropriate assessments, or creat-

ing new ones that are valid and reliable, can be costly and time-consuming. So is reviewing and updating the assessments. Hillsborough County, according to Michelle Watts, supervisor of data analysis, has created or identified more than 900 examinations used in the district every year, all developed in-house by teachers. When the district made curriculum changes, her team updated the tests to maintain content validity. With nearly a thousand tests, this updating process is continuous. "We never finish, so you think you have got a test put to bed and it is done, and then we find that we do need to go back and look at it and make some changes," Watts says.[58]

If a district opts to create new assessments, it needs to consider whether it will pilot the assessments before officially rolling them out, as well as how often it will change assessments or assessment items. Hillsborough does not pilot items before they are used on tests, but when the district finds tests that are problematic changes are made.[59]

Having teachers create tests has posed some challenges for HCPS. District staff review test items after the teacher teams submit them and have often found questions that are not challenging enough. Watts explains in a webinar: "Elementary teachers tend to make tests that are too easy for the students, students tend to get most of the items right, which of course means then that those few items that aren't highly easy for the student are the only items that discriminate, so trying to get items that are a little bit more challenging for elementary teachers has been a challenge for us."[60]

Too Few Appropriate Assessments

Some assessments are being used for multiple, but not necessarily compatible, purposes. In the early grades, for example, some formative assessments are being used to measure teacher effectiveness, instead of for their intended purpose of providing a snapshot of student achievement at the beginning of a course of study. Repurposing assessments in this fashion may feel like an easy option for school districts, but it is not always the best option. Researchers caution against using a test for purposes for which it has not been designed, because doing so could reduce its validity.[61] It is important for districts to discuss the intended purpose of an assessment with test makers, to determine if there could be validity issues. Washington, DC, for instance, does not allow the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) to be used as part of teacher evaluation for this reason, recognizing that "its intent is purely formative, and calculations of 'a year's growth' using DIBELS are very complex." Kristie Kauerz, program director for PreK-3rd education at the University of Washington, notes that using DIBELS alone for a high-stakes decision is a misuse of the assessment, as DIBELS is diagnostic in nature and measures a very limited skill set. But "as a provocative data point, however, to spur conversations about how individual children and subsets of children are doing and how that should effect teaching, DIBELS is fine," Kauerz explains.[62]

Researchers caution against using a test for purposes for which it has not been designed, because doing so could reduce its validity.

The Tennessee Department of Education brought together groups of educators to discuss appropriate measures for pre-K through third grade. For pre-K and kindergarten teachers, there is still no alternate assessment for the growth measure. So these teachers will continue to have the school-wide or district-wide rating based on state tests or SAT-10 scores from first and second grades.[63] According to a report on the first year implementation of Tennessee's evaluation system, "The early childhood educator group expressed strong interest in considering how ongoing assessment in early childhood, including screening tests, portfolios and kindergarten readiness, could be harnessed to develop an alternate growth model."[64] Luke Kohlmoos, director of evaluation for Tennessee's Department of Education, suggests that using a portfolio method could be a possibility in the future.[65]

Could Lead to Curriculum Narrowing and Teaching to the Test

States and districts that plan to create new assessments may also find it difficult to muster teacher buy-in. Many teachers have expressed concerns about too much testing. Policymakers and parents also worry that requiring new assessments can lead to more curriculum narrowing, more test preparation, and potentially less time for other areas of development.

In kindergarten in Hillsborough, assessment is limited to a reading measure. Teachers administer a district-created

reading assessment (pre- and post-test) that focuses on skills such as phonemic awareness, according to Anna Brown, director for assessment and performance management for HCPS.

Policymakers and parents also worry that requiring new assessments can lead to more curriculum narrowing, more test preparation, and potentially less time for other areas of development.

A Hillsborough County kindergarten teacher, who spoke on the condition of anonymity, said the district also uses the DRA, a diagnostic that establishes children's reading levels. She said that the more levels that students progress through, the more points teachers gain on their evaluations. This can produce an incentive to focus on drilling students on passages that are not connected to any content. She is concerned that some teachers ask students to read the DRA passages multiple times throughout the year, to see how they are doing. This sounds like a good practice, as it is important to see the progress new readers are making. However, when the time comes to give the end of the year post-test assessments, the passages are no longer cold reads. It is possible that students have seen them before and are reading from memory rather than using decoding skills they have gained during the year. Additionally, this practice raises concerns about teachers using "drill and kill," focusing only on what will be tested. This can be especially damaging in the early grades, when teachers should focus on instilling a love of reading and of learning.[66]

Concerns About Test Security

Hillsborough has also faced test security issues. Hillsborough requires teachers to sign a test security agreement. The district has spent time and resources investigating multiple incidents of improper test administration. As evidenced by the recent indictments of 35 teachers in an Atlanta, Georgia cheating scandal, the concerns are very real.[67]

3. Shared Attribution

The shared attribution method refers to the use of a district-wide, school-wide, value-added measure—or the scores from a specific grade level or subject area—to assess teacher effectiveness. In other words, the individual growth rating for a teacher could be based on the school's performance as a whole in reading and/or math. For example, a kindergarten teacher could receive his student growth rating based on his school's third-grade reading scores on a state test.

This method is a simple way to assign all teachers a value-added score. States and school districts that put all their emphasis on shared attribution seem to be using this method as a placeholder while officials figure out a better way. In Florida, for example, the shared-attribution method is a temporary measure because by 2014-2015, the state expects all districts to create or identify appropriate assessments for currently untested grades and subjects. Yet shared attribution already is being used for high-stakes decisions about Florida teachers. As of the 2011-12 school year, those who receive two consecutive unsatisfactory ratings can be dismissed.

Opportunities

Promotes Shared Accountability

Some experts suggest that this method can encourage school-wide collaboration, especially, as in the Tennessee system, across grade levels.[68] If a kindergarten teacher's rating is based on first graders' scores on the Stanford-10, she has a stake in students' success on that test. Kindergarten teachers might be more apt to meet with first grade teachers to discuss what their students need to know to be more prepared.

Uses Existing Resources

Another advantage is that no additional resources are required, because teachers are measured by student performance on existing assessments. In many Florida districts, for example, kindergarten through third grade teachers are evaluated by school-wide English language arts and math test scores on the FCAT, the state's test for grades 3-8.

Risks

Does Not Help to Provide Useful Individualized Information to Teachers

For teachers in the untested grades and subjects, the attributed rating gives them no information about how their current students are doing. More importantly, the shared attribution method does not give teachers any insight on what they should do differently to help improve learn-

ing outcomes. That is partly why Austin's school district decided to include shared-attribution scores, for all teachers, as only a small part of its system, relying primarily on SLOs instead, which can provide information useful to teachers for improving instruction in specific classrooms.

Does Not Help to Differentiate Teachers in a Meaningful Way

Other experts are concerned that this method waters down the validity of data, making it difficult to determine which teachers are most effective, due to the indirect relationship between instruction and scores on an assessment that might be given years later. There is no way to tell, for instance, which kindergarten teachers' students make more than one year's growth and which teachers' students do not. These ratings are also based in some cases on the performance of students who individual teachers never taught. For example, a second grade teacher who is new to a school is rated on the performance of third through fifth graders, even though she never taught those students.

Does Not Measure a Teacher's Impact on Her Own Students' Learning

In fact, many early grade teachers do not believe that shared attribution provides any accurate measures of their teaching. They say this method devalues what they teach and the contributions they actually make to their own students' learning during the year they teach them. Carolyn Schokley Ralph, a kindergarten teacher in Orange County, FL, said the value-added score tells her nothing about how her own students performed. "After the test results came in, we looked at the list of fifth graders and identified the ones who had been there since kindergarten," she said. "There were only 12 out of about 60."[69] She was frustrated that the student growth portion of her evaluation did not even include many students she taught when they were in kindergarten.

In Tennessee's first-year of implementation, all teachers of untested grades and subjects received a school-wide value-added score based on the state's standardized test for grades 3-8. Many educators expressed frustration that the school-wide data were not reflective of their own individual performances.[70] After holding meetings with educators from around the state, the education department decided to make changes. State officials worked with groups of educators, including groups for the PreK-3rd grades, to identify and develop other, more appropriate measures that could be used to determine growth. Still, officials are concerned that the school-wide growth scores carry too much weight. As a result, the department is working to get legislation introduced that reduces the growth percentage (35 percent) for teachers using school-wide growth scores.[71]

School-wide ratings as a sole growth measure also go against one of the key purposes of teacher evaluation: to provide feedback that enables teachers to improve how they teach. A second grade teacher in Orange County says that her grade level "was given the test results of 3rd-5th graders the previous year to let us know the areas they needed to improve."[72] It was not evident, though, how those scores were relevant to improving her students' learning and performance on FCAT.

Recommendations and Considerations

Based on our analysis of research and states' teacher evaluation plans, the Early Education Initiative at the New America Foundation puts forward three recommendations and nine considerations for federal, state, and district policymakers as they map out how to measure student learning in pre-kindergarten, kindergarten, and the early grades for the purpose of determining teacher effectiveness.

Recommendations

1. Account for Specific Attributes of PreK-3rd Teachers

Effective PreK-3rd teachers are essential for building a solid foundation for children's future academic success and for developing non-academic skills that will be critical throughout their lives. The PreK-3rd grades also have their own distinct challenges. These grades should not be lumped in with the tested grades and subjects. In the early grades, teachers are expected to be subject-matter experts in math, reading, science, and social studies. They also must help children develop executive-functioning skills such as focusing their attention, completing tasks, controlling impulses, sharing with classmates, waiting their turn, and more. They contribute to student's learning in multiple developmental domains including physical, social-emotional, cognition and general knowledge, persistence and motivation, and language and literacy.

At least three steps will help ensure that these teachers' abilities are fairly evaluated if student achievement data are to be used: 1) When setting policies related to the untested grades and subjects, treat PreK-3rd teachers as a distinct group. Policymakers should not assume that whatever works for the seventh grade history teacher would also work for early grade teachers. 2) The measures used to determine student growth and teacher effectiveness should measure not just reading and mathematics knowledge and skills, but also other skills that have bearing on how young children learn. Measures of teacher effectiveness should include tools for observing and rating teaching practices validated in PreK-3rd settings. 3) States and districts should include all teachers, PreK-12th grade, in conversations about suitable measures of children's learning, and align those measures with current research on appropriate assessment of children.

2. Pilot and Evaluate

Before full implementation, states and school districts should pilot student learning measures, coordinating a staged implementation of the teacher evaluation system to address issues that arise. High-stakes consequences should not be attached immediately, but applied after significant kinks have been worked out. States and districts should conduct ongoing evaluations of their systems.

3. Do Not Use "Shared Attribution" Measures From Later Grades as the Sole Measure of Student Growth to Evaluate Early Grade Teachers

It makes little sense to use "shared attribution" data—data from the 3rd-5th grades school-wide or from another grade-level's value-added scores—as the sole measure of student growth to evaluate teachers in the early grades. Data on students that a teacher either has never taught or taught years before are obviously not reflective of that teacher's talents or deficiencies. Neither is the information derived from these data of much help when it comes to informing the teacher's own instruction, since assessments may be given years after he or she has a certain group of students. What's more, these measures do not help administrators make decisions about the kind of professional development teachers might need to improve.

Considerations

As states and school districts continue to develop, implement, and refine the student growth measures of their teacher evaluation systems, there are various issues to address:

1. Assessments and Their Purpose

Assessments are designed to be used in specific ways and don't always lend themselves well to other purposes. States and school districts should be mindful of this and consult psychometric experts and the body of research on appropriate assessment.

Here is useful guidance provided by the Indiana State Department of Education concerning formative measures that other states and districts should heed:

> Formative assessments are used for the purpose of measuring student progress on a particular skill or content area. The results from such assessments are used to change or enhance instruction in order to ensure mastery of skill or content. The formative nature of the assessment is altered when data are used for evaluation purposes, and this can influence the way teachers prepare for and administer these tests.[73]

2. State, District, and School Capacity

There is a lot of work ahead for states, districts, and schools when it comes to implementing new teacher evaluation systems. To navigate such unchartered waters, ships will need a strong crew. To start, states will need to decide whether there should be one statewide system or many district-level systems. They will need to be prepared to provide technical assistance to help determine what measures are appropriate for young children, what skills should be measured in the early grades, and how those skills can be measured in accordance with developmentally appropriate guidelines. Training is needed to ensure data are not misused. States can also play a role in bringing educators together to share information and potentially promising practices.

States and districts will need to plan for additional staff members to support the assessment and evaluation methods selected. If states choose to create or identify new assessments for each subject area to measure student growth, they will need to invest resources in developing or selecting an appropriate growth model as well as in choosing assessments that are valid and reliable for teacher evaluation.

States will also need to think about how to capture and house evaluation data. This is a particular concern with the SLO and assessment bank approaches. An assessment bank requires an online platform for teachers to review and select assessments. The SLO process requires a lot of information to be approved and tracked. Teachers should be able to submit their SLOs to their principals electronically through a district system, so district staff can have a way to monitor the SLOs if they choose to do so.

3. Curriculum Narrowing

Following the enactment of No Child Left Behind, discussions of curriculum narrowing became more urgent. Stories from teachers abound about the lack of time for science, social studies, hands-on learning centers, recess, and anything else not included on state standardized tests. Previous research on this issue focused on the upper elementary grades. With the institution of more assessments with high stakes attached, PreK-3rd teachers may feel pressured to focus instruction on topics included on assessments or on the objectives they are using to measure student growth.

While states may focus on improving students' numeracy and literacy skills in PreK-3rd, they should be concerned with whether students are honing skills in the other domains of learning, such as social-emotional development and the domain referred to as "approaches to learning." Skills under the social-emotional domain include self-control and the ability to develop positive relationships with other children and adults. Skills in the "approaches to learning" domain include curiosity, persistence, self-organization, reasoning, and the ability to apply prior knowledge to new situations. All of these skills are important to student success in later grades. States and districts should design policies that reward teachers for helping children develop knowledge and skills across a wide range of domains.

4. Teacher and Principal Pre- and In-Service Development

Engaging schools of education in conversations about teacher evaluation is important. States should also update regulations for teacher and principal preparation programs to reflect what is required in evaluation systems. Preparation programs should include deeper instruction on creating good assessment measures and selecting high-quality assessments, setting rigorous learning goals and targets, and using data to inform practice and instruction.

If one of the primary purposes of the student learning objective (SLO) method is to provide teachers with data that can help inform—often immediately—their instruction, then they need professional development in how to use that information. Still missing from teacher preparation and often from professional development provided by districts is training in understanding and using data to help students improve.[74] Also important to the SLO model

are trainings on how to write meaningful goals and high-reaching, but attainable, targets for students. Training in how to differentiate those targets for different groups of students is also important. Principals need this training and support as well, as they are the primary approvers of teachers' goals, targets, and assessments.

Deep learning about how to select and create assessments is often absent from teacher preparation programs. As a result, teachers enter the classroom without the knowledge and skills necessary to construct rigorous, appropriate assessments to measure the learning goals they set. States and local school districts must build teachers' skills in this area, especially if they allow teachers to create their own measures, or encourage them to be a part of a district or state effort to create new assessments.

5. Teachers' Role in the Process

Creating a system where teachers set goals and design measures to gauge student growth on which their compensation or jobs depend is rife with problems. This is the case with SLOs. Research says that young children do better on assessments—often one-on-one—when adults they know and trust administer them.[75] In the PreK-3rd grades, it is often the teacher who does the assessing. This does not necessarily make sense, though, when the results are used for accountability purposes. States should conduct pilots to examine whether this approach leads to unintentional—or intentional—manipulation of the system.

6. Pre-K and Kindergarten Teachers

States and school districts with state-funded pre-K programs will need to determine whether or not the teachers employed as part of these programs will be included in teacher evaluation systems. In some places, like Oklahoma and Georgia, pre-K teachers are required to have a bachelor's degree and certification and they are paid on the same scale as K-12 teachers. That is not the case everywhere. Some public pre-K programs are located in public elementary schools, but not all of them. States and districts need to think about which teachers should be included: only those hired by public schools, or also those in community-based programs? States and districts will also need to decide who should evaluate pre-K teachers and how these teachers should be measured. Kindergarten teachers are already explicitly included in most of the new teacher evaluation models. Yet few, if any, of these models acknowledge that some kindergarten teachers are teaching half-day programs or are teaching two full classes of children, one in the morning and one in the evening. Judgments of teacher effectiveness based on kindergarteners' performance on assessments should, at the very least, take into account these variables.

7. New Common Assessments

Before states implement the new Common Core assessments in the 2014-15 school year, they should develop plans to align their teacher evaluation systems with the Common Core State Standards. According to a report from *Education Week*, 30 states have already done this and more have plans underway.[76] The Partnership for Assessment of the Readiness for College and Careers (PARCC) is designing K-2 formative assessment resources.[77] Will states adopting the new PARCC common assessments require, allow, or recommend these assessments be used in evaluating K-2 teachers even though they are not intended for that purpose?

8. Rating Teachers on How They Use Assessment and Data

In pre-K and early grade classrooms, teachers often administer assessments to small groups of children, or one-on-one. In evaluations of PreK-3rd grade teachers, states and districts should consider including information about whether teachers administered those assessments appropriately and what they did with the data. Did teachers use the information they gleaned to differentiate instruction for students at varying levels?

9. Research

In the PreK-3rd grade world, measuring student learning for the purpose of teacher evaluation is truly uncharted water, and there is limited research available on the approaches discussed in this paper. On shared attribution, studies have explored the Teacher Advancement Program (TAP),[78] and results show TAP schools have higher overall growth in the tested grades when compared to a control group of schools. But studies did not look specifically at the school-wide value-added component. Creating new assessments for teacher evaluation has not been adequately mapped out.[79] There have also been studies on the use of SLOs that show positive results, but again only in the tested subjects.[80] States and school districts should keep this fact in mind when making decisions about which approaches to use.

Conclusion

Building evaluation systems that can improve teaching and that can help distinguish effective from ineffective teachers is something policymakers have grappled with for years. There is little dispute that the impact a teacher has on student learning should be part of how she is evaluated. Figuring out how to do this well is a daunting task and some states are sailing without a rudder.

Undoubtedly, the toughest waters to navigate are the untested grades and subject areas. In most states, standardized testing programs only cover English language arts and mathematics in grades three through eight, and once in high school, leaving nearly three quarters of teachers lacking a ready method to determine their impact on student growth.

Identifying approaches to measure student growth in the PreK-3rd grades is complex. There are distinct challenges for this set of teachers. The developmental growth of children in the early grades is directly linked to their academic growth. The paper-and-pencil tests used with older kids will not work with children ages three through eight. And measures of literacy and numeracy alone do not allow for a full picture of a young child's learning or his teacher's impact—in fact, many experts would argue this is also the case for grades 3 -12.

Yet we cannot forget the potential of new teacher evaluation systems: to improve teaching and learning by ensuring that every child has access to an effective teacher, to identify those teachers who are already helping children to achieve, and to provide constructive feedback and new courses of action for teachers who are not.

To realize this potential, policymakers need to be ready to mitigate risks such as sparse expertise, lack of capacity, and curriculum narrowing. They need to improve upon other evaluative tools, such as instruments for observing and measuring what teachers do in the classroom.[81] That means they need to commit resources to improving in-service and pre-service development for teachers and principals, identifying valid and reliable measures that provide a well-rounded picture of student learning, and conducting ongoing evaluations of the system to make sure it is measuring what it is supposed to fairly and accurately.

The old way of doing things did not work for policymakers, principals, teachers, and most importantly students. All students—especially at-risk students—deserve a well-trained, effective teacher who can challenge them, instill a love of learning, and help them develop the knowledge and skills they need for success in school and life.

Regardless of these risks, though, overhauling teacher evaluation systems must continue. The old way of doing things did not work for policymakers, principals, teachers, and most importantly students. All students—especially at-risk students—deserve a well-trained, effective teacher who can challenge them, instill a love of learning, and help them develop the knowledge and skills they need for success in school and life. Getting this right is crucially important in the PreK-3rd grades. Research has confirmed, time and time again, how the quality of instruction and the quality of learning opportunities in children's formative years sets the foundation for their success as students and successful adults.

Interviews Conducted

Erika Bolig, elementary formative and interim assessment specialist, Michigan Department of Education
Anna Brown, director for assessment and performance management, Hillsborough County Public Schools
Connie Casha, director of early childhood education programs, Tennessee Department of Education
Vincent J. Costanza, early childhood program specialist, New Jersey Department of Education
Diane Donohue, special assistant for educator effectiveness, Teacher Leader Effectiveness Unit, Delaware Department of Education
Patty Ewen, early childhood consultant, New Hampshire Department of Education
Robin Gelinas, senior policy advisor, Education Counsel
Laura Goe, principal investigator for research & dissemination, National Comprehensive Center for Teacher Quality, Education Testing Services
Kristie Kauerz, research scientist and director of PreK-3rd education, University of Washington
Luke Kohlmoos, director of evaluation, Tennessee Department of Education
Jim Lesko, director, early development and learning, Delaware Department of Education
Scott Marion, associate director, Center for Assessment
Jennifer Mosley, research director, Teaching Strategies, LLC
Sam Pearcy, coordinator impact teacher effectiveness, District of Columbia Public Schools
Janice Poda, strategic initiative director, Council of Chief State School Officers
Tamika Pollins, teacher quality specialist, Rhode Island Department of Education
Carolyn Schokley Ralph, former kindergarten teacher, Orange County, FL
Elliot Regenstein, senior vice president, Advocacy and Policy, Ounce of Prevention
Cortney Rowland, senior policy analyst for teacher and leader effectiveness, National Governors Association
Chris Ruskowski, chief officer , Teacher and Leader Effectiveness Unit, Delaware State Department of Education
Jonah Stuart, director, Public Policy and Government Relations, Teaching Strategies, LLC
Amanda Szekely, senior policy analyst for early childhood education, National Governors Association
Joann Taylor, assistant director of strategic competition, Austin Independent School District
Albert Wat, senior policy analyst for early childhood education, National Governors Association
Rob Weil, director of innovation project, America Federation of Teachers
K. Michelle Worrell, project manager, Item & Test Bank Platform, Office of Race to the Top Assessments, Florida Department of Education
Margery Yeager, policy advisor, Education Counsel
Anonymous, kindergarten teacher
Anonymous, first grade teacher

Notes

1 National Council on Teacher Quality, "State of the States 2012: Teacher Effectiveness Policies," 2012, accessed January 7, 2013 at http://www.nctq.org/p/publications/docs/Updated_NCTQ_State%20of%20the%20States%202012_Teacher%20Effectiveness%20Policies.pdf.

2 Sarah Mead, Andrew Rotherham, and Rachael Brown, "The Hangover: Thinking about the Unintended Consequences of the Nation's Teacher Evaluation Binge," Bellwether Education Partners, 2012, accessed October 12, 2012 at http://bellwethereducation.org/wp-content/uploads/2012/09/Teacher-Quality-Mead-Rotherham-Brown.pdf.

3 Jay Mathews, "Rating D.C. Teachers No Easy Task," *Washington Post*, November 9, 2009, accessed January 30, 2013 at http://articles.washingtonpost.com/2009-11-02/news/36771705_1_teacher-evaluation-system-teacher-assessments-national-teacher.

4 Matt Dixon, "Don Gaetz: Florida Teacher Evaluation Not Working," *Florida Times-Union*, January 29, 2013, http://jacksonville.com/opinion/blog/403455/matt-dixon/2013-01-29/don-gaetz-florida-teacher-evaluation-not-working.

5 While third graders do take state assessments, in most cases, second graders do not. Therefore, there is a lack of baseline data to generate value-added scores for third grade teachers.

6 Partnership for Assessment of Readiness for Colleges and Careers, "K-2 Assessments," accessed on February 26, 2013, http://www.parcconline.org/K2-assessments.

7 Eva L. Baker, Paul E. Barton, Linda Darling-Hammond, Edward Haertel, Helen F. Ladd , Robert L. Linn, Diane Ravitch, Richard Rothstein, Richard J. Shavelson, and Lorrie A. Shepard, "Problems with the Use of Student Test Scores to Evaluate Teachers" Economic Policy Institute, 2010, accessed April 25, 2013 at http://www.epi.org/page/-/pdf/bp278.pdf.

8 "Highly Qualified Teachers for Every Child," U.S. Department of Education, August 2006, 2, accessed March 12, 2013 at http://www2.ed.gov/nclb/methods/teachers/stateplanfacts.pdf.

9 Under NCLB, states were required to measure the extent to which students in the state were being taught by highly qualified teachers (HQT), based upon these three criteria: they had (1) earned a bachelor's degree, (2) obtained full state certification or licensure, and (3) proved they knew each subject area they taught. Upon assessing teacher quality, states were required through the HQT provision to adopt goals and plans to ensure all teachers became highly qualified, then publicly report those plans and progress made toward teacher quality goals. U.S. Department of Education, "Fact Sheet: New *No Child Left Behind* Flexibility: Highly Qualified Teachers," March 2004, 2, http://www2.ed.gov/nclb/methods/teachers/hqtflexibility.pdf.

10 Michelle Rhee, Superintendent of D.C. Public Schools, *Charlie Rose*, July 14, 2008.

11 Daniel Weisberg, Susan Sexton, Jennifer Mulhern, and David Keeling, *The Widget Effect: Our National Failure to Acknowledge and Act on Differences in Teacher Effectiveness*, The New Teacher Project, 2009, accessed January 30, 2013 at http://widgeteffect.org/downloads/TheWidgetEffect.pdf.

12 "TVASS: An Introduction to Value-Added in Tennessee," *Taking Note* (State Collaborative on Reforming Education, June 2012), accessed on January 10, 2013 at http://www.tnscore.org/wp-content/uploads/2010/12/SCORE_TVAAS4.pdf.

13 "ESEA Flexibility," U.S. Department of Education, March 2013, accessed December 8, 2013 at http://www.ed.gov/esea/flexibility.

14 Asenith Dixon, "Focus on Teacher Reform Legislation in SREB States: Evaluation Policies," Southern Regional Education Board, June 2011, accessed September 26, 2013 http://publications.sreb.org/2011/11S07_Focus_Teach_Eval.pdf; "National Teacher Evaluation Systems," State of Arkansas Bureau of Legislative Research, August 14, 2012, Accessed September 26, 2013 at http://www.arkleg.state.ar.us/education/K12/TeacherIssueDocuments/12-097_Public_NationalTeacherEvaluationSystem_EdComm_BB_08-14-2012.pdf; National Council on Teacher Quality, "State of the States 2012: Teacher Effectiveness Policies," 2012, accessed January 7, 2013, http://www.nctq.org/p/publications/docs/Updated_NCTQ_State%20of%20the%20States%202012_Teacher%20Effectiveness%20Policies.pdf; "Comparison of States that Use Student

Achievement in Teacher Evaluation Systems," New York Governor's Office, June 24, 2011, http://www.governor.ny.gov/assets/documents/RTTTStateComparison1.pdf.

15 National Council on Teacher Quality, "State of the States 2012: Teacher Effectiveness Policies," 2012, accessed January 11, 2013 at http://www.nctq.org/p/publications/docs/Updated_NCTQ_State%20of%20the%20States%202012_Teacher%20Effectiveness%20Policies.pdf.

16 Cynthia D. Prince, Patrick J. Schuermann, James W. Guthrie, Peter J. Witham, Anthony T. Milanowski, Christopher A. Thorn, "The Other 69 Percent:Fairly Rewarding the Performance of Teachers of Nontested Subjects and Grades," Center for Education and Compensation, 2009, accessed October 12, 2012 http://www.cecr.ed.gov/guides/other69Percent.pdf.

17 "Principles and Recommendations for Early Childhood Assessments," National Education Goals Panel, 1998, accessed November 9, 2013 at http://govinfo.library.unt.edu/negp/reports/prinrec.pdf.

18 Laura Bornfreund, "Getting in Sync: Revamping the Preparation of Teachers in Pre-K, Kindergarten and the Early Grades" New America Foundation, 2011, accessed April 30, 2013, http://www.newamerica.net/publications/policy/getting_in_sync.

19 *Taking Stock: Assessing and Improving Early Childhood Learning and Program Quality*, The Report of the National Early Childhood Accountability Task Force convened by the Foundation for Child Development, the Pew Charitable Trusts and the Joyce Foundation, 2007, accessed November 9, 2012 at http://ccf.tc.columbia.edu/pdf/Task_Force_Report.pdf.

20 Interview with Elliot Regenstein, senior vice president, advocacy and policy, Ounce of Prevention (November 13, 2012).

21 Interview with Kristie Kauerz, director of preK-3rd education and research scientist at the University of Washington (January 16, 2013).

22 Dorothy S. Strickland and Shannon Riley-Ayers, "Early Literacy: Policy and Practice in the Preschool Years," National Institute for Early Education Research, 2006, accessed February 19, 2013 at http://nieer.org/publications/policy-matters-policy-briefs/policy-brief-early-literacy-policy-and-practice-preschool.

23 Common Core, *Learning Less: Public School Teachers Describe a Narrowing Curriculum*, The Farkas Duffett Research Group, March 2012, accessed February 19, 2013 http://commoncore.org/_docs/cc-learning-less-mar12.pdf.

24 Supporting State Efforts to Design and Implement Teacher Evaluation Systems," National Comprehensive Center for Teacher Quality, December 13, 2010, accessed October 12, 2012 at http://www.tqsource.org/webcasts/201012Workshop/Teacher_Effectiveness_Workshop_Glossary.pdf.

25 From presentation by William J. Slotnick, Communications Training Analysis Corporation (CTACUSA), The National Press Club, February 27, 2013.

26 Email Interview with Sam Pearcy, DC Impact November 28, 2012.

27 Interview with Joann Taylor, assistant director of strategic compensation, Austin Independent School District (September 21, 2012).

28 Ibid

29 Interview with Tamika Pollins, educator quality specialist, Rhode Island Department of Education (October 25, 2012).

30 "2011-2012 Student Learning Objectives Examples," Reach: Supporting and Rewarding Success in the Classroom, Austin Independent School District, accessed November 12, 2012 at

31 "REACH SLO Manual," Austin Independent School District, accessed November 12, 2012 at http://archive.austinisd.org/inside/initiatives/compensation/docs/SCI_SLO_Manual.pdf?v=2012.

32 Interview with Joann Taylor, assistant director of strategic compensation, Austin Independent School District (September 21, 2012).

33 Julie Greenburg and Kate Walsh, "What Teacher Preparation Programs Teach About K-12 Assessment: A Review", National Council on Teacher Quality, 2012, accessed on November 12, 2013 at http://www.nctq.org/edschoolreports/assessment/report.jsp.

34 Approved district assessments include: TPRI/Tejas LEE, DRA/EDL, DIBELS/IDEL, Flynt Cooter, Fitnessgram and NOVANET BASI.

35 Interview with Joann Taylor, assistant director of strategic compensation, Austin Independent School District (September 21, 2012).

36 Scott Marion, Charlie DePascale, Chris Domaleski, Brian Gong, and Elena Diaz-Bilello, "Considerations for Analyzing Educators' Contributions to Student Learning in Non-tested Subjects and Grades with a Focus on Student Learning Objectives," Center for Assessment, 2012, accessed November 15, 2012 http://www.nciea.org/publication_PDFs/Measurement%20Considerations%20for%20NTSG_052212.pdf.

37 Ibid.

38 Interview with Scott Marion, associate director, Center for Assessment (February 12, 2013).

39 Interview with Tamika Pollins, educator quality specialist, Rhode Island Department of Education (October 25, 2012).

40 Interview with Joann Taylor, assistant director of strategic compensation, Austin Independent School District (September 21, 2012).

41 Interview with Sam Pearcy, coordinator, IMPACT Teacher Effectiveness, Washington DC Public Schools, October 11, 2012.

42 Interview with Diane Donohue, special assistant for educator effectiveness, Delaware Departmetn of Education (January 29, 2013).

43 "Delaware Performance Appraisal System: DPAS II Guide Revised for Teachers," Delaware Department of Education, August 2012, accessed November 12, 2013 at http://www.doe.k12.de.us/csa/dpasii/files/DPASTeachFullGuide.pdf.

44 Teachers who fall under this category are working with infants and toddlers who are receiving special education services.

45 Interview with Jim Lesko, director, Early Development and Learning, Delaware Department of Education (February 12, 2013).

46 "Supporting Early Childhood Educator Evaluations," Educator Effectiveness at the Colorado Department Education, accessed on March 4, 2013 at http://www.cde.state.co.us/EducatorEffectiveness/downloads/FactSheets/ECE.pdf.

47 Colorado Reading To Ensure Academic Development Act (2012); Preschool to Postsecondary Education Alignment Act, Colorado Achievement Plan for Kids CAP 4 (2008).

48 Interview with Tara Boertzel-Schuenemann, student growth consultant for the early childhood education work group, Colorado Department of Education (November 20, 2012).

49 Interview with Chris Ruszkowski, chief officer , Teacher and Leader Effectiveness Unit, Delaware Department of Education (November 8, 2012).

50 Interview with Sam Pearcy, coordinator, IMPACT Teacher Effectiveness, Washington DC Public Schools, October 11, 2012.

51 Interview with Sam Pearcy, coordinator, IMPACT Teacher Effectiveness, Washington DC Public Schools, October 11, 2012.

52 "Evaluating Teachers of Untested Grades and Subjects Webinar," Midwest Regional Educational Laboratory, American Institutes for Research, August 29, 2012, accessed January 26, 2013 at http://www.relmidwest.org/WebTranscriptAug29.pdf.

53 "Multiple Measures of Teacher Effectiveness in Hillsborough County Public Schools: Implementing Value-Added Measures," Hillsborough County Public Schools, accessed October 15, 2012 at http://communication.sdhc.k12.fl.us/eethome/casestudies/Implementing%20Value-added%20Measures%20in%20Hillsborough%20County.pdf; Interview with Anna Brown, director for assessment

and performance management, Hillsborough County Public Schools (September 21, 2012).

54 "15% Approved Achievement Measures: Revised Options for 2012-2013," Team Tennessee, Tennessee Department of Education, August 16, 2012, accessed October 16, 2012 at http://team-tn.org/assets/misc/15%25%2Achievement%20Measure%20Options_12_13(5).pdf.

55 Interview with Luke Kohlmoos, director of Evaluation, Tennessee Department of Education (October 16, 2012).

56 "Teacher Evaluation in Tennessee: A Report on Year 1 Implementation," Tennessee Department of Education, 2012, accessed on February 8, 2013 at http://www.tn.gov/education/doc/yr_1_tchr_eval_rpt.pdf.

57 "Evaluating Teachers of Untested Grades and Subjects Webinar," Midwest Regional Educational Laboratory, American Institutes for Research, August 29, 2012, accessed January 12, 2013 at http://www.relmidwest.org/WebTranscriptAug29.pdf.

58 Ibid.

59 Ibid.

60 Ibid.

61 Kyle Snow, "Developing Kindergarten Readiness and Other Large-Scale Assessment Systems," National Association for the Education of Young Children, December 2011, 18, accessed January 18, 2013 at http://www.naeyc.org/files/naeyc/file/research/Assessment_Systems.pdf; Interview with Scott Marion, associate director, Center for Assessment, February 12, 2013; Interview with Kristie Kauerz, director of preK-3rd education and research scientist at the University of Washington (January 16, 2013).

62 Interview with Kristie Kauerz, director of preK-3rd education and research scientist at the University of Washington (January 16, 2013).

63 Interview with Luke Kohlmoos, director of Evaluation, Tennessee Department of Education (October 16, 2012).

64 "Teacher Evaluation in Tennessee: A Report on Year 1 Implementation," Tennessee Department of Education, 2012, accessed on February 8, 2013 at http://www.tn.gov/education/doc/yr_1_tchr_eval_rpt.pdf.

65 Interview with Luke Kohlmoos, director of Evaluation, Tennessee Department of Education (October 16, 2012).

66 Interview with anonymous kindergarten teacher, March 19, 2013.

67 David Beasley, "Atlanta Educators Surrendering In Cheating Scandal", Reuters, April 2, 2013, accessed on April 20, 2013 at http://www.reuters.com/article/2013/04/02/us-usa-schools-atlanta-idUSBRE9310YP20130402.

68 Scott Marion and Katie Buckley, "Approaches and Consideration for Incorporating Student Performance Results from "Non-Tested" Grades and Subjects into Educator Effectiveness Determinations," 2011, accessed September 20, 2012 at http://www.nciea.org/publications/Considerations%20for%20non-tested%20grades_SMKB2011.pdf.

69 Interview with Carolyn Schokley Ralph, a former kindergarten teacher, Orange County, FL (March 6, 2013).

70 State Collaborative on Reforming Education, "Supporting Effective Teaching In Tennessee: Listening And Gathering Feedback On Tennessee's Teacher Evaluations," 2012, accessed December 12, 2012 at http://www.tnscore.org/wp-content/uploads/2012/06/SCORE-Evaluation-Report.pdf.

71 Interview with Luke Kohlmoos, director of Evaluation, Tennessee Department of Education (October 16, 2012).

72 Interview with anonymous second grade teacher, March 5, 2013.

73 "Legislative Guidance: Assessments," Indiana Department of Education, February 2012, accessed March 23, 2013 at http://www.in.gov/legislative/interim/committee/reports/SEDUF8F.pdf.

74 Jennifer C. Kabaker and Clare McCann, forthcoming paper on data-driven instruction from the Federal Education Budget Project at the New America Foundation

75 Kyle Snow, "Developing Kindergarten Readiness

and Other Large-Scale Assessment Systems," National Association for the Education of Young Children, December 2011, 18, accessed January 18, 2013 at http://www.naeyc.org/files/naeyc/file/research/Assessment_Systems.pdf.

76 William Porter, Regina Riley, Lisa Towne, Sean M. Chalk, Amy M. Hightower, Sterling C. Lloyd, Carrie A. Matthews, and Christopher B. Swanson, "Moving Forward: A National Perspective on States' Progress In Common Core State Standards Implementation Planning," Education First and EPE Research Center, 2013, March 30, 2013 at http://www.edweek.org/media/movingforward_ef_epe_020413.pdf.

77 K-2 Assessments, Partnership for Assessment of Readiness for College and Careers, 2013, accessed March 12, 2013 at http://www.parcconline.org/K2-assessments.

78 TAP: The System for Teacher and Study Advancement, is a model developed by Milken Family Foundation Chairman Lowell Milken—in partnership with others involved in the Milken Family Foundation—targeted at attracting, retaining, developing, and motivating new talent in education. The TAP model is based on the following four elements: allowing teachers to pursue a variety of career paths, providing ongoing opportunities for applied professional growth, comprehensively evaluating teachers based on instruction, and awarding performance-based compensation. TAP is being implemented in diverse urban, rural, and suburban districts across the country; the National Institute for Excellence in Teaching (NIET) has also helped facilitate the use of the TAP system through state partnerships with Texas, Indiana, Louisiana, and South Carolina. http://www.tapsystem.org/what/what.taf?page=response.

79 According to Lisa Lachlan-Haché of the American Institutes for Research, in a webinar for REL-Midwest.

80 Community Training and Assistance Center, "It's More Than Money: Teacher Incentive Fund—Leadership for Educators' Advanced Performance," 2013, accessed March 1, 2013 at http://www.ctacusa.com/PDFs/MoreThanMoney-report.pdf.

81 Lisa Guernsey and Susan Ochshorn, "Watching Teachers Work: Using Observation Tools to Promote Effective Teaching in the Early Years and Early Grades," New America Foundation, 2011, http://newamerica.net/sites/newamerica.net/files/policydocs/Watching_Teachers_Work.pdf.

1899 L Street, NW
Suite 400
Washington, DC 20036
Phone 202 986 2700
Fax 202 986 3696

NEW AMERICA FOUNDATION

WWW.NEWAMERICA.NET