NEW AMERICA

EDUCATION POLICY

EARLY & ELEMENTARY EDUCATION POLICY

ABBIE LIEBERMAN, REED DESROSIERS, AND DAVID LOEWENBERG

WHY PRE-K AND THE EARLY GRADES ARE DIFFERENT

A Blog Series on Teacher Evaluation

SEPTEMBER 2016

About the Authors

Abbie Lieberman is a policy analyst with the Education Policy program at New America. She is a member of the Early & Elementary Education Policy team, where she provides research and analysis on policies that impact children from birth through third grade. Prior to joining New America, she interned on the legislative affairs team at Education Trust and at the U.S. Department of Education working on the 2013 Race to the Top District competition.

Reed DesRosiers was a summer 2016 intern for the Education Policy program. Before coming to New America, he taught at Frederick Douglass High School in Baltimore as a Teach For America corps member. He holds a master of science in education from Johns Hopkins University and a bachelor of science in policy analysis and management from Cornell University.

David Loewenberg is a fall 2016 intern with the Education Policy program at New America. Previously, he taught pre-K in St. Louis and interned in the U.S. Department of Education's Office of Early Learning. Loewenberg holds a bachelor's degree in government from the University of Texas at Austin.

Acknowledgments

This series of briefs benefited from the expert and editorial insight of Laura Bornfreund as well as the editing skills of Sabrina Detlef. We would also like thank the Foundation for Child Development for their generous support of this work. The views expressed in this series do not necessarily represent the views of the foundation, its officers, or employees.

About New America

New America is committed to renewing American politics, prosperity, and purpose in the Digital Age. We generate big ideas, bridge the gap between technology and policy, and curate broad public conversation. We combine the best of a policy research institute, technology laboratory, public forum, media platform, and a venture capital fund for ideas. We are a distinctive community of thinkers, writers, researchers, technologists, and community activists who believe deeply in the possibility of American renewal.

Find out more at **newamerica.org/our-story**.

About the Education Policy Program

New America's Education Policy program uses original research and policy analysis to solve the nation's critical education problems, serving as a trusted source of objective analysis and innovative ideas for policymakers, educators, and the public at large. We combine a steadfast concern for low-income and historically disadvantaged people with a belief that better information about education can vastly improve both the policies that govern educational institutions and the quality of learning itself. Our work encompasses the full range of educational opportunities, from early learning to primary and secondary education, college, and the workforce.

Our work is made possible through generous grants from the Alliance for Early Success; the Buffett Early Childhood Fund; the Foundation for Child Development; the Bill and Melinda Gates Foundation; the Evelyn and Walter Haas, Jr. Fund; the HeisingSimons Foundation; the William and Flora Hewlett Foundation; the Joyce Foundation; the George Kaiser Family Foundation; the W.K. Kellogg Foundation; the Kresge Foundation; Lumina Foundation; the McKnight Foundation; the Charles Stewart Mott Foundation; the David and Lucile Packard Foundation; the J.B. & M.K. Pritzker Family Foundation; the Smith Richardson Foundation; the W. Clement and Jessie V. Stone Foundation; and the Berkshire Taconic Community Foundation.

Find out more at **newamerica.org/education-policy**.

Contents

INTRODUCTION

Research confirms what many of us intuitively know: better teachers lead to better student outcomes on measures of academic achievement, college attendance, and career earnings.[1] In the last seven years, an increased focus on teacher evaluation has raised many questions about what constitutes an "effective teacher" and what systems best capture a teacher's effectiveness. As states and districts grapple with identifying and implementing the most appropriate evaluation tools, one of the many outstanding questions is how such tools can be most appropriately used in the early elementary grades, where learning and instruction often look different than in older grades.

States and districts have had systems in place to evaluate teacher performance for decades. Evaluation tools vary, but have historically relied upon classroom observations and principals' intuition. Unfortunately, many of these older systems do very little to differentiate between high- and low-quality educators. In 2009, TNTP (formerly The New Teacher Project) published an influential and oft-cited report, *The Widget Effect*, based upon the study of evaluation systems in 12 diverse school districts.[2] The findings were startling. Nearly all teachers were evaluated positively, with fewer than one percent of teachers receiving unsatisfactory ratings. Additionally, classroom observations did not inform professional development or staffing decisions. Nearly three out of four teachers reported that they never received specific feedback from their observations, and half of the districts had not dismissed a tenured teacher for poor performance in five years. Evaluation systems seemed to be more of a bureaucratic exercise than a system meant to improve the quality of the teaching workforce.[3]

Not long after *The Widget Effect* was published, the Obama administration announced the landmark Race to the Top grant competition.[4] The competition's cornerstone was a call for states to innovate their teacher evaluation models, and it offered more than one-quarter of the competition's

Research confirms what many of us intuitively know: better teachers lead to better student outcomes on measures of academic achievement, college attendance, and career earnings.

points for systems that promoted "great teachers and leaders." Many states applied for these grants and have since overhauled their teacher evaluation systems in an attempt to better measure teacher quality.

The administration also made teacher evaluations a priority with its No Child Left Behind (NCLB) waivers.[5] To escape the unrealistic expectations of the federal education law, states were required to develop evaluation systems that used multiple measures to evaluate teachers, including at least one measure of student growth. Desperate for flexibility under NCLB, almost all states applied for waivers despite controversies around teacher evaluation. [6] By the end of 2015, 43 states required student growth to be taken into account in teacher evaluation.[7]

Needless to say, teacher evaluation looks markedly different today than it did seven years ago in most parts of the country. These new systems differ by state, but often include some combination of classroom observations and measures of student growth, as well as, more recently, student feedback. These changes have not come without growing pains and most states are still adjusting their systems to find what works best for teachers, administrators, and most importantly, students.

One challenge has been determining the best way to evaluate early education teachers, particularly in the "untested grades," pre-K through second grade. During the same time that teacher evaluation systems have been changing, there has also been an increasing recognition of the importance of early education at the federal, state, and local level. We have a better understanding now than ever before about how important the early years of school are to children's development and future success. And much learning in these early years depends on the quality of children's interactions with adults and the relationships formed with teachers and peers.

Based on what we know about child development, teaching and learning should look different in pre-K, kindergarten, and the early grades of elementary school than they do for older students. Yet many states have developed "one-size-fits-all" evaluation systems that were not created with the needs of young children in mind. Early grade teachers and evaluators have questioned whether the way older grade teachers are evaluated is appropriate for teachers of younger children. Should teachers be observed and held accountable to the same rubrics, regardless of their students' age? Is it appropriate or even possible to accurately measure children's gains in pre-K and kindergarten? Based on what tests? Can young children accurately give feedback about their teachers? States and districts have grappled with these questions and others as they have implemented new teacher evaluation systems.

In 2013, New America released *An Ocean of Unknowns: Risks and Opportunities in Using Student Achievement Data to Evaluate PreK–3rd Grade Teachers*, a policy paper examining the best way to measure teacher impact on student learning.[8] In the summer of 2016, New America's Early & Elementary Education Policy team revisited this topic. We examined how several states and districts are differentiating their systems by grade level as they implement new teacher evaluation processes. We looked at three components of teacher evaluations and how they are used in the early grades: classroom observations, student-growth measures (commonly known as student learning objectives, or SLOs), and student surveys. We looked at how states are incorporating these components into their evaluation systems in this three-part series on teacher evaluation in the early grades.

State and district teacher evaluation systems are likely to continue to change as states implement the new Every Student Succeeds Act. As evaluation practices are revised and improved, it is imperative that the unique demands and circumstances of the early grades are not lost in the shuffle.

A LOOK AT CLASSROOM OBSERVATIONS

Picture yourself as an observer conducting a teacher evaluation, tasked with deciding where teachers fall on a scale of "ineffective" to "highly-effective," potentially affecting their pay or job security. You walk into a science lesson on the conservation of mass in a fifth grade classroom. The desks are in rows and the students are listening to their teacher at the front of the room. She asks them to predict whether the ice cube on her desk will maintain its mass when it melts. Using what they learned in the previous night's reading, she asks the students to explain their predictions in their journals. After 10 minutes, she asks them to share with the person next to them and then selects one student to read his answer aloud. By this time, the ice has melted and the students can see that the mass has remained the same.

Next you observe a science lesson in a kindergarten classroom, where the students are learning about the properties of different materials. With the children seated in a circle on the floor, the teacher reads *Captain Kidd's Crew Experiments with Sinking and Floating*.[9] Afterwards, she asks the students why they think some objects sink and others float. She writes their ideas down on the whiteboard. Then she pulls out a few everyday items from a bag and asks the class to predict whether they will sink or float. After guessing together, the students return to their tables, which have been equipped with similar items and tubs of water. Working in small groups they test their ideas out for themselves. They draw pictures of the items that sink under the "Sink" heading and the items that float under the "Float" heading using a graphic organizer. The teacher walks around checking in with each small group, asking probing questions.

Instruction in these two classrooms looks very different. As the observer, do you know whether it was good practice to have fifth graders write in journals rather than share with the whole class? Or whether there was the right balance between whole group instruction and student-centered learning in the kindergarten classroom?

Classroom observation allows principals or external observers to see teachers in action and offer feedback that can help them improve their practice.[10] But high-quality teaching should look different from one grade to the next, especially in the early years. Notice how the lesson plans, classroom environments, and role of the teacher differ in these grades. To effectively promote high-quality teaching across all grade levels, evaluators need a keen understanding of these differences.

As teacher evaluation systems have been changing in recent years, many states and districts have

updated their frameworks for observing teachers. However, many states continue to use one general framework across all grades. But observation tools are often created with a certain age group in mind and using them to evaluate teachers instructing different grades can be confusing or even unfair. For instance, some rubrics used for observing teachers in K–12 might be inconsistent with best practices in the early grades or fail to clarify how to identify certain measures in classrooms where instruction looks different.

Lisa Guernsey and Susan Ochshorn's 2011 paper, *Watching Teachers Work: Using Observation Tools to Promote Effective Teaching in the Early Years and Early Grades*, examines the importance of classroom observation as a tool to identify, promote and reward good teaching.[11] While observations are increasingly likely to inform personnel decisions, they should also play a prominent role in helping teachers understand the parts of their practice that are most beneficial to children, and the parts that they can change to be more effective. As Guernsey and Ochshorn explain, "professional development and formal evaluations will need to go hand-in-hand, with data from observations bridging the two."

In the early grades, a high-quality observation tool should emphasize the importance of certain types of interactions and teaching strategies that help students gain academic skills in areas like language, literacy, and math, and to develop social-emotional skills. Teaching in these years should be hands-on, young children should be engaged, teachers should be responsive and encourage children to build on their interests, and adults in the classroom should demonstrate an understanding of child development and learning.

Observation tools designed specifically for pre-K classrooms usually acknowledge this. For instance, Head Start and many other pre-K programs use tools like the Classroom Assessment Scoring System (CLASS) to observe teachers. CLASS measures interactions related to emotional climate, classroom organization, and instructional support.

Most state Quality Rating and Improvement Systems require the use of an observation tool like CLASS.[12] However, in pre-K and child care centers, especially those outside of the public school system, these tools are usually used to measure overall program quality, as opposed to formally evaluate teachers. These types of tools are appropriate for measuring quality teaching, but they rarely meet state requirements for teacher evaluation. As pre-K is more and more commonly folded into the public school system, states and school districts need to ensure that their observation tools can accurately evaluate quality instruction in pre-K and early grade classrooms.

Several states and districts, such as Illinois and the District of Columbia, recognize that using one classroom observation model for all grades and subject areas may be an ineffective or unfair way to evaluate teachers of younger children, specifically those in kindergarten through third grade, and pre-K when included. As such, they have developed separate rubrics, guidelines, or methods to better evaluate early educators. The lessons they have learned may help states that have not yet acknowledged the differences between evaluating teachers of young children and older students.

Illinois

Illinois has taken significant steps to ensure that early education teachers are evaluated on the practices that are best for young children. The state encourages districts to select one evaluation rubric for all staff, but acknowledges that teaching and learning in the early grades may require a different kind of tool.

Illinois is one of many states that has approved the use of the Charlotte Danielson Framework for Teaching for teacher observations across grade levels, including pre-K.[13] This framework, like many others, was created for use beginning in the upper grades of elementary school, raising concerns with how well the tool adapts to the early grades. To figure this out, researchers at the Center for the Study of Education Policy at Illinois State

University (CSEP) conducted a validation study of the Danielson Framework to determine if it is valid and reliable in the early grades.[14]

CSEP spent the first year taking an in-depth look at the content of the Danielson Framework to determine if it was aligned with what research says is important for children in pre-K through third grade. When comparing it to NAEYC's Standards for Professional Preparation Programs, CLASS, and the Head Start standards, they found that overall, it aligns with developmentally appropriate practice.[15] Unsurprisingly, Danielson is more academic than the early childhood-specific frameworks and has less emphasis on family engagement. According to Lisa Hood, director of the study at CSEP, "this doesn't mean Danielson can't be used to evaluate social-emotional interactions and family engagement, it just needs to be more intentional."[16]

Twenty-six teachers (14 pre-K teachers and 12 K–3rd grade teachers) in seven districts with a total of 620 students (50 percent in pre-K) participated in CSEP's validation study.[17] To test the framework's inter-rater reliability, the researchers paired internal observers (principals/center directors) with trained external observers and compared their classroom observation ratings on 17 components. A comparison of the ratings showed an inter-rater reliability average of 67 percent, with agreement between the internal and external observers ranging from as low as 42 percent in one component to as high as 92 percent in another. Internal observers tended to rate teachers higher than external observers on several components.

Based on these findings, CSEP is developing resources for the areas of the framework where inter-rater reliability was weakest, such as on using assessment, setting instructional outcomes, and more abstract concepts, like "developing respect and rapport" or creating a "culture of learning." In June, the CSEP team embarked on a three-year project to develop videos showing best practices for how pre-K and kindergarten teachers and their principals can navigate the observation tool and evaluation process.

Early grade teachers and evaluators have access to extensive documents created by a group of early childhood stakeholders that outline multiple examples of what each component of the Danielson Framework might look like in the early years.[18] Hood says CSEP has received positive anecdotal feedback about the examples, but it has not collected systematic feedback on whether principals actually go back to their schools and use the tool. Principals also have access to trainings provided by the Illinois Principals Association and guidance created by the Illinois State Board of Education's Performance Evaluation Advisory Council (PEAC) around pre-K through third grade evaluation.[19]

While CSEP found that most teachers earn a "proficient" rating (the performance levels are *unsatisfactory, basic*, *proficient*, and *distinguished*) in Danielson, it is possible for the tool to differentiate early childhood educator performance. According to Hood, most of the challenges with Danielson in the early grades are "user-oriented issues, instead of with the framework itself. When used well and when people have strong understanding of early childhood practice, Danielson works well. When they don't have this background, that's when there's an issue," she said.

District of Columbia Public Schools

District of Columbia Public Schools (DCPS), which has been an oftentimes controversial pioneer when it comes to teacher evaluation reform, uses IMPACT, a self-created teacher evaluation system.[20] IMPACT has been around since 2009, but a separate rubric to evaluate pre-K and kindergarten teachers that more accurately reflects developmentally-informed practice was created in 2011. The preK–K rubric was updated in 2016 through a collaborative process that involved content area experts weighing in to ensure that it is appropriate for the youngest learners. The rubric includes the same broad practices as those of the older grades, but differs in the way that it describes their implementation.[21]

The rubric for grades 1–12 focuses on what observers should see students doing, whereas the preK–K

rubric is more focused on whether the teacher is creating the conditions to make learning possible.[22] Accordingly, the early childhood rubric looks more at teacher actions instead of independent student actions. As depicted in the examples below, the preK–K rubric evaluates teachers based on how well they *encourage* students to take certain actions or behave in certain ways, whereas the grades 1–12 rubric rates teachers for how well students take certain actions *independently*. The early childhood rubric mentions the importance of learning environments and how teachers can encourage meaningful work and play, rather than just work. The grades 1-12 rubric makes no mention of learning environments or play. Furthermore, it gives specific guidance that observers should "consider students' developmental age when assessing" certain practices.

Figure 1 | Excerpts from DCPS Rubrics

DCPS PreK-K Rubric

ESSENTIAL PRACTICE 1	CULTIVATE A RESPONSIVE LEARNING COMMUNITY	
	1.A Supportive Community	**1.B Student Engagement**
LEVEL 4	All students are valued members of a welcoming and responsive learning community.* **The teacher proactively cultivates community with and among students.**	All students are engaged throughout the learning experience OR almost all students are engaged throughout the learning experience and the teacher responds to disengagement by inviting students back in to the learning experience. **The teacher promotes engagement by establishing purpose for what students are learning and doing.**
	For example, the pre-kindergarten or kindergarten teacher: • Consistently encourages students to praise and show appreciation for one another • Provides time, space, and structures for positive interactions between students such as morning meeting "shout-outs" • Has meaningful systems in place to encourage acts of kindness and compassion with peers and adults *See also examples from Level 3*	For example, the pre-kindergarten or kindergarten teacher: • Shares with students what they are working on and why, as appropriate • Has clear, student-friendly academic and developmental objectives for centers or workstations • Communicates daily developmental objectives in ways such as oral explanations, sight words, pictures, and voice recordings *See also examples from Level 3*

DCPS Grades 1-12 Rubric

ESSENTIAL PRACTICE 1	CULTIVATE A RESPONSIVE LEARNING COMMUNITY	
	1.A Supportive Community	**1.B Student Engagement**
LEVEL 4	All students are valued members of a welcoming and responsive learning community.* **Students are authentically welcoming and responsive to one another.**	All students are engaged throughout the learning experience OR almost all students are engaged throughout the learning experience and the teacher responds to disengagement by inviting students back in to the learning experience. **Students demonstrate deep investment in the learning experience.**
	For example, the students: • Demonstrate interest in the thoughts, opinions, and well-being of each other • Provide peers with meaningful and specific feedback/praise • Productively collaborate across difference (e.g., cultural, racial, linguistic, dis/ability, and/or gender) *See also examples from Level 3*	For example, the students: • Persevere when they struggle with challenging content or activities • Demonstrate interest in, commitment to, or excitement about what they are learning and doing *See also examples from Level 3*

According to Stephanie Shultz, who works on IMPACT, the preK–K rubric aligns with the "context and structures you are most likely to see with young learners, such as station-driven learning, play, morning meeting, etc." It also emphasizes language development, which is a crucial component of learning at this age.

The stakes are high for DCPS teachers: the observations make up a majority of evaluation scores for all preK–12 teachers.[23] This year, for the first time, school principals will be the only ones using the rubrics to evaluate teachers; in the past, external observers ("master educators") have played a prominent role.[24] Hiring principals who are instructional leaders is a priority for the district and all principals receive extensive training and support from the IMPACT team to become familiar with the tool.[25] It's important that this training enables principals to distinguish high-quality instruction in a kindergarten class versus a fourth grade class.[26]

Shultz says DCPS has "received a lot of appreciation from early childhood teachers, who say they 'see themselves in the rubric' and appreciate the distinction between this rubric and the one used with other students." The district should consider extending the preK–K rubric into the first through third grades to reflect the full continuum of early childhood.

Teacher evaluation systems should reward good teaching and promote improvements in practice. Teaching young children requires different skills and strategies than those used for older children, and the best observation tools acknowledge those differences. It can be difficult for a single tool to meet the needs of a teacher who is reading stories about floating boats and another who is teaching the law of conservation of mass, but specific guidance for teachers and observers on how standards and rubrics can be tailored for the early grades is one way to help ensure that evaluations accurately capture the quality of teaching.

A LOOK AT STUDENT LEARNING OBJECTIVES

With recent work highlighting the critical role that teachers play in student achievement, and with a nudge—or push—from education reformers and the Obama Administration, student growth measures have become a key part of teacher evaluation systems throughout the United States. The most common way to measure student growth achievement is with standardized test scores. Since state standardized testing usually does not begin until third grade, states and districts have developed other methods to assess how much the "untested" younger students learn. According to the National Council on Teacher Quality's 2015 *State of the States* report, 39 of the 43 states that require a

student growth component in teacher evaluations weigh student growth measures the same in untested grades/subjects as in tested grades.[27]

But what is the best way to use the gains made during the school year by pre-K through second grade students to evaluate a teacher, where growth is not as straightforward as the difference between scores on tests? Student Learning Objectives (SLO) have emerged as a tool to evaluate an individual teacher's impact on his or her students.[28]

New America's *An Ocean of Unknowns: Risks and Opportunities in Using Student Achievement Data to Evaluate PreK–3rd Grade Teachers*, looks closely at how five states (Colorado, Delaware, Florida, Rhode Island, and Tennessee) and three school districts (Austin, Texas; Hillsborough County, Fla.; and Washington, D.C.) were using SLOs and other measures of student growth in the early grades.[29] We revisit Delaware and Washington, D.C. to see how they have adapted their system of SLOs to better serve early grades. We also take a look at the changing use of SLOs in Georgia.

The SLO process varies, but it generally involves teachers collecting baseline student data, setting measurable goals for students evaluated by an assessment, usually with the approval of their school leader, and working towards these goals throughout a school year. As of 2014, SLOs were part of teacher evaluation systems in at least 25 states.[30]

This method of evaluating teachers has advantages. In addition to allowing schools to measure teacher impact in untested grades and subjects, educators using SLOs have reported that they promote data-driven instruction, increase collaboration between teachers and administrators, encourage closer tracking of student progress, and make use of higher-quality assessments.[31] In multiple studies, teachers have reported spending more time analyzing student data and reflecting on their practice.[32]

SLOs are not without limitations, however. The limited research on the link between SLO attainment and other measures of student achievement has yielded mixed results.[33] Some studies have found some positive relationship between SLOs and student achievement, but others have found inconsistent correlations across subjects and grades.[34] This inconsistency may partially be due to the fact that the use of SLOs varies widely between districts, schools, and even classrooms. Additional, rigorous research is needed to evaluate the consistency and accuracy of SLOs across subjects and grade levels, especially for PreK–2nd.

> **Questions remain about how effective the use of SLOs is, relative to other teacher evaluation measures, in differentiating teacher performance.**

Additionally, questions remain about how effective the use of SLOs is, relative to other teacher evaluation measures, in differentiating teacher performance. Traditional evaluation, most commonly classroom observations, has been shown to fail in capturing a nuanced picture of professional performance, since the vast majority of teachers are rated highly.[35] An Institute of Education Sciences study looking at eight early-adopting school districts found that the inclusion of alternate student growth measures, such as SLOs, yielded greater differentiation in teacher performance than ratings based on classroom observation.[36] An evaluation in New Jersey also found that including student growth measures increased differentiation in teacher ratings, even though the vast majority of teachers remained in the higher ratings categories.[37]

Setting goals for students requires walking a fine line between the ambitious (high goals) and the attainable (realistic ones). There are concerns that evaluating teachers on goals that they set for themselves creates an incentive for them to lean towards the attainable rather than the ambitious.[38] Principals can counteract this by creating systems

and rules that hold staff to high expectations. It takes significant time, resources, and expertise, however, for school leaders to implement these systems and create a culture that encourages teachers to aim for a balance of ambitious and attainable.[39] To do this well, principals need sufficient professional development and support from the school district. Another challenge is that teachers might be essentially "grading themselves," because they are highly involved in the SLO process—often writing, administering, and grading the assessments.[40] There is a high potential for manipulation, especially when results are tied to high-stakes consequences such as merit pay, dismissal, or tenure.

Of the less-discussed challenges are those surrounding the implementation of SLOs in the early grades.[41] Because of the significant variation in children's development from pre-K through second grade, it can be difficult to create or select developmentally-appropriate assessments that reliably measure a teacher's impact on students.[42] As New America's Laura Bornfreund and Clare McCann have explained, "the developmental growth of children in the early grades is directly linked to their academic growth…measures of literacy and numeracy alone do not account for a full picture of a young child's learning or his teacher's impact in laying the foundation for…long-term success in school."[43] Despite these challenges, many states and districts have gone ahead with SLOs in the early grades.

Delaware

In Delaware, the use of SLOs in the early grades remains a factor in the state's teacher evaluation system, but the state's use of SLOs has adapted over time to address concerns raised by teachers and administrators. In collaboration with their principal, teachers in all grades, including pre-K, set growth goals for students at the beginning of the year and can choose from a list of more than 200 approved assessments (both internal and external) to evaluate student progress.[44] Teachers can also work with principals to select measures not included in the assessment bank. The state's online database allows teachers to reference student growth data from previous years to help them set reasonable and appropriate goals.

The assessment bank was developed quickly and, at first, some assessments faced quality issues, prompting pushback from some teachers. But, according to Laura Schneider, director of educator effectiveness at the Delaware Department of Education, the state has listened to concerns and improved assessments over time thanks to an ongoing review process that includes regular reliability and validity tests of the assessments.[45] In response to teacher and administrator feedback, Delaware has recently reduced the number of SLOs required from four to two because of time and resources involved with their administration. Delaware's requirements surrounding SLOs do not differ by grade level.

Buy-in from teachers and administrators around the use of student growth measures in general has been challenging, but open communication and professional development have helped those at the school-level see the benefits of SLOs. While results are tied to formal teacher evaluations, "it's about measuring student growth and not adult behavior," Schneider said. "Implementation has been improving over time as people realized that."

District of Columbia Public Schools

In Washington, D.C., a district that has aggressively pursued reforms in teacher evaluation, SLOs (referred to as Teacher Assessed Student Achievement Data, or TAS) are used to measure student growth in all grades and account for 15 percent of a teacher's overall evaluation.[46] DCPS provides guidance around which assessments are appropriate for use in each grade. DCPS suggests that pre-K teachers use the observation-based assessment, Teaching Strategies GOLD, that evaluates students across seven domains: social-emotional, language, physical-gross motor, physical-fine motor, cognition, literacy, and math. While TS GOLD is recommended for all

pre-K teachers, the decision of what assessment to use and how heavily to weight GOLD as a TAS assessment is ultimately left to principals.

DCPS may serve as an example of how schools and districts can overcome one of the key challenges of SLOs: the inability to compare teachers. While the implementation and assessment of SLOs often varies between classrooms and schools, the common use of TS GOLD across pre-K classes creates a standardized system in which teachers can be compared and trends at the classroom, school, and district level can be identified.

There is more variation in selected assessments in kindergarten through second grade. DCPS recommends a few appropriate assessments for these grades, but teachers have the flexibility to select other assessments or create their own with approval from their principal.[47] According to the TAS guidance, "school leaders decide what assessments, weights, goals, tracking systems, and data collection methods are appropriate for their school." While principals approve their teachers' SLOs, DCPS reviews every single one, all 10,000 of them. According to Michelle Hudacsko, Deputy Chief of IMPACT, principals review their teachers' TAS goals for appropriate rigor and then the district does a final review of all goals each year for workability, to ensure the goal can be scored appropriately at the end of the year.

Georgia

While Delaware and DCPS have adapted and continued to use SLOs over the years, Georgia has recently reversed course on the use of SLOs.[48] Legislative action in 2016 determined that districts are no longer required to use SLOs as part of their teacher evaluation systems.[49] Student growth now accounts for only 30 percent of a teacher's evaluation score, reduced from 50 percent. Cindy Saxon, associate superintendent of teacher and leader effectiveness for the Georgia Department of Education said, "this decision was an effort to reduce the testing burden on students and teachers." Districts now have considerable discretion in selecting which type of student growth measures to use, be it SLOs or some other measurement such as an off-the-shelf assessment.

While the state is moving away from SLOs as a required student growth measure, Georgia's innovative "resource library" may be an example for other states looking to support the use of SLOs. Containing more than 2,000 assessment items across courses and content, the resource library allows districts from across the state to access and share assessments for SLOs. The library also includes about 220 state exemplars for districts to use in their entirety, which were designed by Georgia educators to provide support for districts.[50] It's too soon to tell yet if districts will continue to use SLOs now that they are not required by the state.

Pulling back on SLOs may become a more widespread trend as states implement the new Every Student Succeeds Act (ESSA). The law allows for much greater flexibility around teacher evaluations. In an EdWeek article shortly after the law was passed, Molly Spearman, the state superintendent for South Carolina, was quick to say that she'll be delaying the state's requirement for all districts to use SLOs by two years and "would like to make those measures an artifact examined by evaluators, not a specific, weighted component of each teacher's review."[51]

States and districts that continue to use SLOs must ensure that teachers and principals have the time, resources, and support to learn how to use them effectively and in a way that maintains high expectations for students and mitigates the possibility of manipulation. In the early grades, teachers and districts must be mindful about using appropriate assessments to measure student growth and should take advantage of the opportunities SLOs offer to gauge a student's growth outside of the oftentimes limited domains of math and literacy.

A LOOK AT STUDENT SURVEYS

Peer into a college classroom near the end of a semester and you may encounter a perplexing sight: students hunched over answer sheets, bubbling in answers and scribbling notes, seemingly unsupervised and undirected. But these students aren't taking a test. They are completing student surveys, a tool used by colleges and universities to evaluate professors. Now this practice, long a part of higher education, is being tried in K–12 classrooms as part of efforts to reform the way teachers are evaluated.

The reasoning is intuitive. Throughout the year, students spend countless hours with their teachers, while principals, who have long been the primary evaluators of classroom teaching quality, may visit a classroom a few times a year. Students know whether their teacher is able to manage the classroom and whether they feel supported as learners. It shouldn't come as a surprise, then, that research shows that student perceptions of teacher quality may actually be more accurate than those of a principal.[52]

In the changing teacher evaluation landscape, student surveys have surfaced as a way to collect feedback from those who know teachers best.

Using student surveys to effectively evaluate teachers doesn't come without its fair share of challenges, however. Surveys rely on students being able to understand and answer questions about their teachers. Older students seem to have no problem doing this. The Bill & Melinda Gates Foundation's Measures of Effective Teaching (MET) project found that survey results from students in grades four through eight can predict student achievement gains, are more consistent than results from classroom observations, and provide teachers specific feedback to improve their practices.[53]

But what about using student surveys for children in pre-K through second grade? Asking young children to evaluate their teachers can prove more challenging for a host of reasons. There are concerns about their ability to understand questions or give unbiased feedback.[54] Current research on the consistency and accuracy of student survey results is far from robust, but a few studies have found that pre-kindergarteners and kindergarteners can reliably rate their teachers.[55] Still, many state and local policymakers are still deciding whether surveys in the youngest grades make sense.

Whether surveys are appropriate at any age depends largely on how they are designed and implemented. An effective survey must ask questions in a way that yields meaningful results.[56] For instance, rather than asking students to respond to the statement, "My teacher is clear," a high-quality survey might ask students to respond to, "My teacher has several

good ways to explain each topic that we cover in this class." Surveys should look different in pre-K through second grade than they do later on because of literacy challenges and developmental differences. A six-year-old probably wouldn't be able to read the sentence above and might not even comprehend it correctly if read aloud. And, if administered aloud, would a child feel comfortable providing negative feedback? If administered in a small group, would children feel pressured to agree with their peers?

> **In the changing teacher evaluation landscape, student surveys have surfaced as a way to collect feedback from those who know teachers best.**

Although most states do not use surveys in the early grades to inform teacher evaluation, a handful of states and districts have either experimented with or implemented K–2 student surveys. Their experiences offer a glimpse into what this evaluation tool could look like for others. Some of these states and districts have undertaken the design process on their own, but many others use companies, like Tripod or Panorama Education, which have expertise in student survey design. Most of the leading companies designing student surveys, however, do not offer them for students in kindergarten through second grade.

Tripod is one of the few that does.[57] Tripod's surveys, which span from kindergarten through grade 12, were first developed in 2001 as a partnership between Harvard University professor Ronald F. Ferguson and Ohio educators. These surveys, now used in nearly 30,000 classrooms in 24 states, have been updated in response to feedback from research, field experience, and input from key stakeholders to ensure that "survey items at the Early Elementary level are specifically geared for students at this [developmental] level." Tripod says its K–2 surveys are "read aloud to students, in small groups, by someone other than the classroom teacher," and students "only need to be able to recognize the words 'No,' 'Maybe,' and 'Yes'" to respond. Tripod's analysis from "thousands of classrooms," claims that kindergarten students are able to distinguish between these words. It's unclear whether this analysis found whether students can fully comprehend the questions asked. However, Tripod says that its surveys "have generated results that meet rigorous tests of reliability," and that early elementary surveys "have been shown to predict achievement gains as well as other outcomes such as classroom-level behavioral engagement, emotional engagement, and motivational mindsets."[58]

Panorama Education also offers surveys in the early grades.[59] A representative from Panorama says the surveys the company has designed for students in grades 3–12 have been validated, but "there is some research that K–2 students may have a difficult time conceptualizing and comprehending some items around climate, valuing of school, classroom engagement—making them less reliable measures." It still offers early elementary surveys when requested, though. On Panorama's survey, students choose between a smiling face, neutral face, and frowning face, rather than "yes," "maybe," and "no" due to the varied and often limited reading abilities of young children. When we asked Tripod if it considered giving students the option of pictorial responses, such as a smile for "yes" and a frown for "no" the company responded: "pictorial responses can be more open to interpretation and do not always align to the thoughts and feelings we are trying to capture through the survey items." Whether pictorial or word-based responses best capture students' views remains to be proven.

Tulsa Public Schools

Tulsa Public Schools is one district that recently started using Tripod surveys in K–2, but there has been backlash from dozens of early grade teachers who expressed concerns about the appropriateness of the surveys for the age group and their potential impact on the student-teacher relationship.[60]

In response, the superintendent of Tulsa Public Schools created a task force to study the use of the surveys, in the early grades.[61] Tripod has been working with the task force to answer questions about the surveys, create resources to support survey administration, and even offered suggestions for reformatting the surveys to address teacher concerns.

Hawaii

The state of Hawaii implemented Tripod surveys in the early grades statewide but ultimately decided against using them. Hawaii started using the Tripod student survey in the 2012–13 school year to constitute 10 percent of a teacher's overall evaluation score for all grade levels. Teachers and teacher unions lobbied to remove surveys from early grades, and as of the 2014–15 school year, kindergarten through second grade teachers did not use surveys at all. Despite the data provided by Tripod, educators felt that students at this age did not have the capacity to comprehend the purpose of the survey. They also raised concerns over the amount of time and resources it took to implement the surveys, as they had to be read aloud in as many as 14 different languages to accommodate the state's diverse student population.[62]

> **Despite the data provided by Tripod, educators felt that students at this age did not have the capacity to comprehend the purpose of the survey.**

Georgia

Georgia had a related experience. The state experimented with using student surveys to inform teacher evaluation in 2012. According to Cindy Saxon, associate superintendent of teacher and leader effectiveness for the Georgia Department of Education, there was feedback from teachers and parents after the pilot that led to the discontinuation of the survey in K–2. Georgia experienced particular difficulties in the early grades, where young students struggled to log into and navigate a complex online survey platform. Saxon said that "what should have taken ten to fifteen minutes might have taken an hour and a half" in younger grades. These logistical barriers, and teachers' perception that "students couldn't give an accurate view of their classroom" were enormous obstacles to survey use.[63]

Additionally, the state evaluated the pilot survey's results and found that K–2 students tended to rate their teachers significantly higher across the domains, and K–2 results were the least consistent with other evaluation measures.[64] In the most recent legislative session, lawmakers removed surveys from Georgia's evaluation system for all grade levels. Saxon felt that, unfortunately, the surveys were often perceived as another assessment to be used punitively against teachers, and there was concern about the time and resources they took to administer.

Massachusetts

Massachusetts is another state working to determine how best to incorporate student feedback into teacher evaluations at all grade levels. Massachusetts' overall evaluation system provides significant flexibility to districts; the state doesn't assign specific weights or point values to different components of the evaluation. Rather, it describes three broad categories of evaluation and recommends pieces of evidence to collect in each one. One of these categories requires evidence of "student and staff feedback."

The state has also decided that surveys for early grade students are not the answer. In 2013, Massachusetts officials began collaborating with districts to develop and pilot standardized student surveys in all grades. But according to Matthew Holloway, an educator effectiveness specialist at the Massachusetts Department of Elementary and Secondary Education, teachers in the early grades gave feedback after the initial pilot that "some of the barriers around literacy, reading, and writing,

and more general barriers" made a more traditional survey "an ineffective way to get feedback."[65] With this in mind, the state worked with educators to create a set of "Discussion Prompts" that capture "a broad range of educator effectiveness standards" in what it feels is a more developmentally-appropriate manner.[66] The state recommends that teachers or a third party use the discussion prompts with small groups of students to solicit feedback. For example, the teacher might try to gauge whether the following statement related to informal assessments is true for students: "When I am stuck, my teacher wants me to try again before she or he helps me."

Using the state-developed discussion prompts is not mandatory; districts and school districts can choose their own form of surveys or alternative evidence. Holloway stressed that giving districts flexibility in how they collect student feedback "allows them to emphasize their priorities." He noted that he has seen the most success with student feedback in districts that promote the process as a way for teachers to "align student feedback to their own self-assessment and goal setting processes," where teachers work to improve their practice in a way that is informed by students. Massachusetts is still in the early stages of implementing student feedback as part of teacher evaluations and is examining what is most beneficial to districts, teachers, and students.

It is possible that student survey responses can shed light on teacher performance and encourage teachers to tailor their instruction to better serve their students, but there are real concerns states and districts have about their usefulness in the early grades. States that have attempted to incorporate surveys in the early grades have found them to be very time- and resource-intensive to implement. Further, early grade teachers have expressed concerns over their students' ability to understand the purpose of the task.

Some states do appear to be eager to find ways to integrate student feedback into the evaluation system. As one state official put it, "kids deserve the opportunity to [take student surveys]... and our kids can be empowered if they think they have a voice." While the research on whether young children can effectively evaluate teachers appears inconclusive, there is little doubt that teachers can benefit from knowing what their students think about their classroom experience. But using surveys to inform evaluation is still new and inconsistent. If used solely to inform teacher practice and professional development, reluctance on the side of the teacher may be reduced. More research, however, is needed to determine if surveys should be incorporated into high-stakes decisions for teachers of young children.

Notes

1 Raj Chetty, John Friedman, and Jonah Rockoff, "Great Teaching: Measuring its effects on students' future earnings," Education Next 12, no. 3 (Summer 2012): http://educationnext.org/great-teaching/

2 Daniel Weisberg, Susan Sexton, Jennifer Mulhern, and David Keeling, The Widget Effect: The National Failure to Acknowledge and Act on the Differences in Teacher Effectiveness (New York: The New Teacher Project, 2009), http://tntp.org/publications/view/evaluation-and-development/the-widget-effect-failure-to-act-on-differences-in-teacher-effectiveness.

3 Kaylan Connally and Melissa Tooley, Beyond Ratings: Re-envisioning State Teacher Evaluation Systems as Tools for Professional Growth (Washington, DC: New America, 2016), https://www.newamerica.org/education-policy/policy-papers/beyond-ratings/.

4 U.S. Department of Education, "Race to the Top Fund," 2016, http://www2.ed.gov/programs/racetothetop/index.html.

5 Anne Hyslop, It's All Relative: How NCLB Waivers Did-and Did Not-Transform School Accountability (Washington, DC: New America, December 2013), https://www.newamerica.org/education-policy/policy-papers/its-all-relative/.

6 Joy Resmovitz, "States Escaping No Child Left Behind Can Get More Time On Teacher Evaluations," Huffington Post, August 21, 2014.

7 Kathryn M. Doherty and Sandi Jacobs, State of the States: Evaluating Teaching, Leading, and Learning (Washington, DC: National Council on Teacher Quality, November 2015), http://www.nctq.org/dmsView/StateofStates2015.

8 Laura Bornfreud, An Ocean of Unknowns: An Ocean of Unknowns: Risks and Opportunities in Using Student Achievement Data to Evaluate PreK–3rd Grade Teachers (Washington, DC: New America, May 2013), https://www.newamerica.org/education-policy/policy-papers/an-ocean-of-unknowns/

9 Mark Weakland, Captain Kidd's Crew Experiments with Sinking and Floating, (Picture Window Books, 2012).

10 Gathering Feedback for Teaching (Washington, DC: Bill & Melinda Gates Foundation, 2012), http://k12education.gatesfoundation.org/wp-content/uploads/2016/06/MET_Gathering_Feedback_for_Teaching_Summary1.pdf.

11 Lisa Guernsey and Susan Oschorn, Watching Teachers Work Using Observation Tools to Promote Effective Teaching in the Early Years and Early Grades (Washington, DC: New America, November 2011), http://fcd-us.org/sites/default/files/Watching_Teachers_Work_0_0.pdf).

12 QRIS Compendium, "State Profiles," http://qriscompendium.org/view-state-profiles.

13 Lori Connors-Tadros and Michelle Horowitz, How Are Early Childhood Teachers Faring in State Teacher Evaluation Systems? (New Brunswick, NJ: CEELO, March 2014), http://ceelo.org/wp-content/uploads/2014/03/CEELO_policy_report_ece_teachereval_march_2014.pdf.; The Danielson Group, "The Framework for Teaching," 2013, https://www.danielsongroup.org/framework/.

14 Lisa Hood (Senior Policy Analyst and Researcher Center for the Study of Education Policy), interview with Abbie Lieberman, September 2016

15 Crosswalk Between Pennsylvania Core Knowledge Competencies, NAEYC Standards for Professional Preparation Programs, NBPTS Early Childhood Generalist Standards, and Charlotte Danielson's Framework for Teaching (Harrisburg, PA: Pennsylvania Office of Child Development and Early Learning, October 2013); Illinois State University Center for the Study of Education Policy, "Danielson CLASS (Classroom Assessment Scoring System) Crosswalk," http://teecc.illinoisstate.edu/downloads/Danielson_CLASS_Crosswalk_2014.pdf.; U.S. Department of Health and Human Services, "Head Start Policy and Regulations," https://eclkc.ohs.acf.hhs.gov/policy?language_content_entity=en.

16 Lisa Hood (Senior Policy Analyst and Researcher Center for the Study of Education Policy), interview with Abbie Lieberman, September 2016

17 Lisa Hood, Debra Kasperski, Erika Hunt, Lizanne DeStefano, Saraí Coba Rodriguez, Gabriela Garcia, and Amber Kirchoff, Studying the Danielson Framework for Teaching in PreK-3rd Grade Classrooms (Normal, IL: Illinois State University, September 2015), https://education.illinoisstate.edu/downloads/csep/Final_Danielson%20white%20paper.pdf.

18 Charlotte Danielson, 2013 Framework for Teaching Evaluation Instrument, (2012), http://teecc.illinoisstate.edu/downloads/Pages%20from%20Final%20Revised%20IL_Danielson%20Framework%20with%20Early%20Learning%20Examples-1.pdf.

19 Illinois Principals Association, "PD Overview," http://www.ilprincipals.org/professional-development/pd-overview.; Illinois Board of Education, "Guidance on Building Teacher Evaluation Systems for Teachers of Students With Disabilities, English Learners, and Early Childhood Students," (August 2014), http://isbe.net/peac/pdf/guidance/14-3-teacher-eval-sped-ell-preschool.pdf.

20 Emma Brown, "D.C. Public Schools, closely watched for its reform efforts, is overhauling teacher evaluation and training," Washington Post, February 10, 2016, https://www.washingtonpost.com/local/education/dc-public-schools-to-overhaul-teacher-evaluation-and-training/2016/02/10/bdb9ed2a-cf41-11e5-b2bc-988409ee911b_story.html.; District of Columbia Public Schools, "IMPACT: An Overview," http://dcps.dc.gov/page/impact-overview.

21 District of Columbia Public Schools, "Early Childhood Teachers IMPACT," (August 18, 2016), http://dcps.dc.gov/sites/default/files/dc/sites/dcps/publication/attachments/2a.%20Early%20Childhood%20Education%20Teachers.pdf.

22 District of Columbia Public Schools, "Grades 1-2 General Education Teachers IMPACT," (August 18, 2016), http://dcps.dc.gov/sites/default/files/dc/sites/dcps/publication/attachments/2b.%20Grades%201-2%20General%20Education%20Teachers.pdf.

23 District of Columbia Public Schools, "2016-2017 IMPACT Guidebooks," (August 2016), http://dcps.dc.gov/node/1182935.

24 Emma Brown, "D.C. Public Schools, closely watched for its reform efforts, is overhauling teacher evaluation and training," Washington Post, February 10, 2016, https://www.washingtonpost.com/local/education/dc-public-schools-to-overhaul-teacher-evaluation-and-training/2016/02/10/bdb9ed2a-cf41-11e5-b2bc-988409ee911b_story.html.

25 Stephanie Shultz and Michelle Hudacsko (IMPACT, District of Columbia Public Schools), interview with Abbie Lieberman, August 2016

26 Laura Bornfreund, Connecting Pre-K and the Early Grades: Principals on Transitions (Washington, DC: New America, May 2016), https://static.newamerica.org/attachments/13105-principals-corner/5.4Early-Ed-Principal-4.761d732ad93840d4af845cf8be2d13dc.pdf.

27 Kathryn M. Doherty and Sandi Jacobs, State of the States: Evaluating Teaching, Leading, and Learning (Washington, DC: National Council on Teacher Quality, November 2015), http://www.nctq.org/dmsView/StateofStates2015.

28 Kaylan Connally and Melissa Tooley, Beyond Ratings: Re-envisioning State Teacher Evaluation Systems as Tools for Professional Growth (Washington, DC: New America, 2016), https://www.newamerica.org/education-policy/policy-papers/beyond-ratings/.

29 Laura Bornfreud, An Ocean of Unknowns: An Ocean of Unknowns: Risks and Opportunities in Using Student Achievement Data to Evaluate PreK–3rd Grade Teachers (Washington, DC: New America, May 2013), https://www.newamerica.org/education-policy/policy-papers/an-ocean-of-unknowns/

30 Natalie Lacireno-Paquet, Claire Morgan, Daniel Mello, How states use student learning objectives in teacher evaluation systems: a review of state websites (Washington, DC: U.S. Department of

Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northeast & Islands, March 2014), https://ies.ed.gov/ncee/edlabs/regions/northeast/pdf/REL_2014013.pdf.

31 Moira McCullough, Brittany English, Megan Hague Angus, Brian Gill, Alternative student growth measures for teacher evaluation: Implementation experiences of early-adopting districts (Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Mid-Atlantic, July 2015), http://ies.ed.gov/ncee/edlabs/regions/midatlantic/pdf/REL_2015093.pdf.; State of New Jersey Department of Education, "2013-14 Final Educator Evaluation Implementation Report," http://www.nj.gov/education/AchieveNJ/ resources/201314AchieveNJImplementationReport.pdf.

32 State of New Jersey Department of Education, "2013-14 Final Educator Evaluation Implementation Report," http://www.nj.gov/education/AchieveNJ/resources/201314AchieveNJImplementationReport.pdf.; William J. Slotnik, Daniel Bugler, Guodong Liang, Real progress in Maryland: Student learning objectives and teacher and principal evaluation (San Francisco, CA: Mid-Atlantic Comprehensive Center at WestEd, September 2014), https://www.wested.org/wp-content/files_mf/1413394919RealProgressinMD_Report.pdf.

33 Lisa Lachlan-Haché, The Art and Science of Student Learning Objectives: A Research Synthesis (Washington, DC: American Institutes for Research, April 2015), http://www.air.org/sites/default/files/downloads/report/Art-and-Science-of-Student-Learning-Objectives-April-2015.pdf.

34 Austin Independent School District Department of Research and Evaluation, "AISD Reach Program Update, 2010-2011 Texas Assessment of Knowledge and Skills Growth and Student Learning Objectives," (Fall 2011), https://www.austinisd.org/sites/default/files/dre-reports/rb/10.84_AISD_Reach_TAKS_and_SLOs_2010-2011.pdf.; William J. Stolnik, Maribeth D. Smith, Barbara J. Helms, Zhaogang Qiao, It's More than Money: Teacher Incentive Fund – Leadership for Educators' Advanced Performance Charlotte-Mecklenburg Schools (Boston, MA: Community Training and Assistance Center, February 2013), http://ctacusa.com/wp-content/uploads/2013/11/MoreThanMoney.pdf.; William J. Stolnik, Maribeth D. Smith, Roberta J. Glass, Barbara J. Helms, Catalyst for Change: Pay for Performance in Denver Final Report (Boston, MA: Community Training and Assistance Center, January 2004), http://www.ctacusa.com/wp-content/uploads/2013/11/CatalystForChange.pdf.; Austin Independent School District Department of Program Evaluation, "AISD REACH Year 2: Evaluation Report II, 2008–2009," (April 2010), https://www.austinisd.org/sites/default/files/dre-reports/08.97_AISD_Reach_Year2_Evaluation_ReportII_2008_2009.pdf.

35 Daniel Weisberg, Susan Sexton, Jennifer Mulhern, and David Keeling, The Widget Effect: The National Failure to Acknowledge and Act on the Differences in Teacher Effectiveness (New York: The New Teacher Project, 2009), http://tntp.org/publications/view/evaluation-and-development/the-widget-effect-failure-to-act-on-differences-in-teacher-effectiveness.

36 Moira McCullough, Brittany English, Megan Hague Angus, Brian Gill, Alternative student growth measures for teacher evaluation: Implementation experiences of early-adopting districts (Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Mid-Atlantic, July 2015), http://ies.ed.gov/ncee/edlabs/regions/midatlantic/pdf/REL_2015093.pdf.

37 State of New Jersey Department of Education, "2013-14 Final Educator Evaluation Implementation Report," http://www.nj.gov/education/AchieveNJ/resources/201314AchieveNJImplementationReport.pdf.

38 Matt Barnum, "Putting the 'Achievable Over the Ambitious': The Problem When Teachers Grade Themselves," The 74, July 12, 2016, https://www.the74million.org/article/putting-the-achievable-over-the-ambitious-the-problem-when-teachers-grade-themselves.

39 Laura Bornfreund and Clare McCann, "Principals Central to Student Learning Objectives," EdCentral (blog), New America, February 27, 2014, https://www.newamerica.org/education-policy/edcentral/principals-central-student-learning-objectives/.

40 Matt Barnum, "Putting the 'Achievable Over the Ambitious': The Problem When Teachers Grade Themselves," The 74, July 12, 2016, https://www.the74million.org/article/putting-the-achievable-over-the-ambitious-the-problem-when-teachers-grade-themselves.

41 Brian Gill, Julie Bruch, Kevin Booker, Using alternative student growth measures for evaluating teacher performance: what the literature says (Washington, DC: Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Mid-Atlantic, September 2013), http://ies.ed.gov/ncee/edlabs/regions/midatlantic/pdf/REL_2013002.pdf.

42 Gwen Stevens and Karen DeBord, "Issues of Assessment in Testing Children Under Age Eight," The Forum for Family and Consumer Issues 6, no. 2 (2001): https://ncsu.edu/ffci/publications/2001/v6-n2-2001-spring/issues.php.

43 Laura Bornfreund and Clare McCann, "Principals Central to Student Learning Objectives," EdCentral (blog), New America, February 27, 2014, https://www.newamerica.org/education-policy/edcentral/principals-central-student-learning-objectives/.

44 Delaware Department of Education, "2015-2016 Measure B – Internal Assessments," http://www.doe.k12.de.us/cms/lib09/DE01922744/Centricity/Domain/375/Internal%20Measures%202015-16.pdf.; Delaware Department of Education, "2015-16 DPAS-II Component V Student Growth Measures External Measures: Core Content Areas – Approved," (Spring 2016), http://www.doe.k12.de.us/cms/lib09/DE01922744/Centricity/Domain/375/External%20Measures%202015-16%2020160509.pdf.

45 Laura Schneider (Director, Educator Effectiveness, Delaware Department of Education), interview with Abbie Lieberman and David Loewenberg, September 2016

46 Teacher-Assessed Student Achievement Data (TAS) Guidance (Washington, DC: District of Columbia Public Schools, 2015), http://www.wtulocal6.org/usr/2015/pdfs/SY2015-2016%20TAS%20Guidance.pdf.

47 Stephanie Shultz and Michelle Hudacsko (IMPACT, District of Columbia Public Schools), interview with Abbie Lieberman, August 2016

48 Lee Shearer, "Change in law means less time testing, more time in class for Georgia students," Athens Banner-Herald, August 6, 2016, http://onlineathens.com/mobile/2016-08-06/change-law-means-less-time-testing-more-time-class-georgia-students.

49 Quality Basic Education Act, SB 364, Georgia General Assembly, 2015-2016 Regular Session (2016), http://www.legis.ga.gov/legislation/en-US/Display/20152016/SB/364.

50 Cindy Saxon (Associate Superintendent of Teacher and Leader Effectiveness Georgia Department of Education), interview with Abbie Lieberman and Reed DesRosiers, August 2016

51 Stephen Sawchuk, "ESSA Loosens Reins on Teacher Evaluations, Qualifications," Education Week, January 5, 2016, http://www.edweek.org/ew/articles/2016/01/06/essa-loosens-reins-on-teacher-evaluations-qualifications.html.

52 Student Perception Surveys and Teacher Assessments (Washington, DC: Hanover Research, February 2013), https://dese.mo.gov/sites/default/files/Hanover-Research-Student-Surveys.pdf.

53 Gathering Feedback for Teaching (Washington, DC: Bill & Melinda Gates Foundation, 2012), http://k12education.gatesfoundation.org/wp-content/uploads/2016/06/MET_Gathering_Feedback_for_Teaching_Summary1.pdf.; Asking Students about Teaching: Student Perception Surveys and Their Implementation, (Washington, DC: Bill & Melinda Gates Foundation, September 2012), http://k12education.gatesfoundation.org/wpcontent/uploads/2015/12/Asking_Students_Practitioner_Brief.pdf.

54 Sarah Butrymowicz, "Student surveys to be used to rate teachers in pilot program — even in kindergarten classes," Hechinger Report, May 14, 2012, http://hechingerreport.org/student-surveys-to-be-used-to-rate-teachers-in-pilot-program-even-in-kindergarten-classes/.

55 John Fullman, "Elementary Public School Pupil Rating of Teacher Effectiveness," Child Study Journal 25, no. 1 (1995): 57-78, http://eric.ed.gov/?id=EJ508797.

56 Asking Students about Teaching: Student Perception Surveys and Their Implementation, (Washington, DC: Bill & Melinda Gates Foundation, September 2012), http://k12education.gatesfoundation.org/wp-content/uploads/2015/12/Asking_Students_Practitioner_Brief.pdf.

57 Tripod Education Partners, "Tripod," http://tripoded.com/about-us-2/.

58 Tripod Education Partners, interview via email with Reed DesRosiers, August 2016

59 Panorama Education, "Panorama Student Survey," http://www.panoramaed.com/panorama-student-survey.

60 Andrea Eger, "Tulsa teachers meet with CEO of student surveys to address concerns," Tulsa World, January 8, 2015, http://www.tulsaworld.com/newshomepage3/tulsa-teachers-meet-with-ceo-of-student-surveys-to-address/article_87f8a26e-768a-55a1-87ae-c81fd8f942e3.html.

61 Andrea Eger, "Superintendent announces task forces to address testing, assessment concerns," Tulsa World, November 26, 2014, http://www.tulsaworld.com/news/education/superintendent-announces-task-forces-to-address-testing-assessment-concerns/article_37c36c53-a7da-5d81-99ae-29262074fc4d.html.

62 Sean Arai (Hawaii Department of Education), interview via email with Reed DesRosiers, August 2016

63 Cindy Saxon (Associate Superintendent of Teacher and Leader Effectiveness Georgia Department of Education), interview with Abbie Lieberman and Reed DesRosiers, August 2016

64 Georgia Department of Education, "Overview to the 2012 TKES/LKES Pilot Evaluation Report," (December 12, 2012), https://www.gadoe.org/School-Improvement/Teacher-and-Leader-Effectiveness/Documents/Pilot%20Report_Overview%20and%20Report%20Combined%201-10-13.pdf.

65 Matthew Halloway and Donna Traynham (Massachusetts Department of Elementary and Secondary Education), interview with Abbie Lieberman and Reed DesRosiers, August 2016

66 Massachusetts Department of Elementary and Secondary Education, "Educator Evaluation," February 2015, http://www.doe.mass.edu/edeval/feedback/.