

MEASURING TEACHER EFFECTIVENESS

A LOOK “UNDER THE HOOD” OF
TEACHER EVALUATION IN 10 SITES

A CONNCAN, 50CAN, AND PUBLIC IMPACT REPORT



MEASURING TEACHER EFFECTIVENESS

A LOOK “UNDER THE HOOD” OF
TEACHER EVALUATION IN 10 SITES

BY DANIELA DOYLE AND JIYE GRACE HAN,
PUBLIC IMPACT

A CONNCAN, 50CAN, AND PUBLIC IMPACT REPORT

ACKNOWLEDGMENTS

ConnCAN, 50CAN, and Public Impact are grateful to the many partners and collaborators from the sites featured in this report: Sarah Coon of Achievement First; Madeleine Bayard of the Rodel Foundation; Diane Donohue of the Delaware Department of Education; David Steele of Hillsborough County Public Schools; Ellen Hur of The New Teacher Project; Michele Sherban-Kline of New Haven Public Schools; Sam Franklin, Abigail Loughrey, and Erika Gentry of Pittsburgh Public Schools; Brent Maddin, Thackston Lundy, Norman Atkins, and Kaycee Salmacia of the Relay Graduate School of Education; Melissa Cabral of the Rhode Island Department of Education; Emily Barton and Sara Heyburn of the Tennessee Department of Education; Kim Levensgood of District of Columbia Public Schools; Greg Sommers of the Bill & Melinda Gates Foundation; and Sandi Jacobs of the National Council on Teacher Quality. We also would like to thank Public Impact, including Bryan Hassel who advised on the report and Sharon Kebschull Barrett for copy-editing. Additional thanks go to Jennifer Alexander and Alex Spurrier from ConnCAN and Gretchen Guffy from 50CAN for their contributions to this report. This report was made possible by the generous support of the H. A. Vance Foundation.

PLEASE CITE THIS REPORT AS:

Doyle, D., & Han, J.G. (2012). *Measuring teacher effectiveness: A look “under the hood” of teacher evaluation in 10 sites*. New York: 50CAN; New Haven, CT: ConnCAN; and Chapel Hill, NC: Public Impact. Retrieved from <http://www.conncan.org/learn/research/teachers/measuring-teacher-effectiveness>

©2012 50CAN, ConnCAN, and Public Impact

The authors encourage the free use, reproduction, and distribution of this paper for noncommercial use. **We require attribution for all use.** For more information and instructions on the commercial use of our materials, please contact us at www.publicimpact.com.

TO ORDER ADDITIONAL COPIES:

Please contact ConnCAN at info@conncan.org or 203 772 4017

ConnCAN
85 Willow Street
New Haven, CT 06511

www.conncan.org
[@conncan](https://www.instagram.com/conncan)

CONTENTS

FOREWORD	1
INTRODUCTORY BRIEF	3
CROSS-SITE ANALYSIS	17
SITE PROFILES	25
DELAWARE	25
RHODE ISLAND	36
TENNESSEE	46
HILLSBOROUGH COUNTY, FL	53
HOUSTON, TX	60
NEW HAVEN, CT	69
PITTSBURGH, PA	78
WASHINGTON, D.C.	86
ACHIEVEMENT FIRST (CMO)	96
RELAY GRADUATE SCHOOL OF EDUCATION	105
MORE INFORMATION	115



3

FOREWORD

PATRICK RICCARDS, CONNCAN
MARC PORTER MAGEE, 50CAN

2

1

INCH

0

Research and experience make this clear: Great teachers change lives. They inspire and motivate students, and set them on a path for future success. By contrast, just one underperforming teacher can have a lasting negative impact on a student.

Given this reality, significant time and attention has rightly been focused on ensuring that all children have outstanding teachers at the front of their classrooms. This includes improving how teacher performance is evaluated and using evaluations to provide training, support, and recognition for outstanding performance. It also means ensuring swift and fair dismissal for the small number of teachers who do not improve after receiving support.

A number of publications have reported on state and district policies in this area and offered useful guidance for policy development. Little has been reported to date, however, about what happens once these policies are in place and the difficult work of design and implementation begins. Leaders, advocates, educators, and others must answer complex technical questions about how evaluation systems will work in practice. For example: “How should we evaluate a teacher in a non-tested grade or subject?” or “What mechanisms exist to ensure validity?”

Answering these questions can be a daunting task, but this work does not require reinventing the wheel. Those considering embracing rigorous teacher evaluations can learn from the states, districts, and education organizations that are already engaged in this work. That’s why we developed this report. In partnership with Public Impact, and with generous support from the H.A. Vance Foundation, Measuring Teacher Effectiveness looks “under the hood” at the evaluation systems being implemented by 10 leading sites, including states, districts, a charter management organization, and a graduate school of education.

None of these systems is perfect, but they do show us possible paths forward. Each site continually—and rightly—refines and improves its evaluation system. As other states and districts take on this work, we hope this information will help education leaders develop their own paths forward, and will provide a clearer picture for educators of the next generation of evaluation and development systems. As advocates for great public schools for every child, ConnCAN and 50CAN also hope that advocates can use this information as proof that we can—and must—move forward with smart, balanced, and fair evaluations of a teacher’s impact on student performance and growth. Such evaluation systems are fundamental to ensuring great teachers for every child. Because great schools, and great teachers, change everything.



INTRODUCTORY BRIEF

In the past, we have had, at best, a vague sense of how our teachers were performing and their impact on student learning. After just one or two observations, often lasting fewer than 80 minutes in all, most teachers received some sort of satisfactory rating—as did 94+ percent of their peers.¹ Afterward, we had no better idea of teacher quality, how to best develop teachers’ talents, or how to address ineffectiveness to support students’ academic needs.

Today, all of that is changing. States, school districts, charter school networks, and schools of education across the country are rethinking the way they measure teacher effectiveness, and using that information to improve teacher practice and student outcomes. As part of that effort, these systems are adopting new educator evaluations that include multiple measures, including teacher observation, student performance, student perceptions, community involvement, and other factors that foster the conditions students need to achieve at high levels. Sites are using these measures not only to differentiate between multiple levels of teacher effectiveness, but also as a tool to inform important decisions about teacher professional development, retention, dismissal, and pay, all of which are aimed at increasing overall teacher effectiveness and improving student outcomes. The best way to make these changes and others, however, is neither easy nor obvious. With this challenge in mind, ConnCAN and 50CAN approached Public Impact for this report.

REPORT GOALS

This report highlights 10 of the most advanced and talked-about teacher evaluation systems nationally: Delaware; Rhode Island; Tennessee; Hillsborough County, FL; Houston, TX; New Haven, CT; Pittsburgh, PA; Washington, DC (referred to throughout just as Washington); Achievement First (a charter management organization, or CMO); and the Relay Graduate School of Education in New York City. Together, these systems serve more than 1.6 million students each year. These are not the only systems taking on this work, but they are trailblazers. None of the sites we feature would say they have fully “figured out” teacher evaluation. Nor do we hold them up as examples of perfection. Yet they have all worked long and hard to carefully and thoughtfully tackle the most difficult challenges related to developing and implementing a high-quality teacher evaluation system.

As more states, districts, school systems, and schools of education strive to develop and implement teacher evaluation systems, they will wrestle with these same challenges. This report is designed to help share information and lessons learned about this difficult, yet essential, work. Our report therefore strives to “get under the hood” of the teacher evaluation systems at these 10 sites. We have collected information about these emerging systems in a single place, using a consistent format that allows users to look across sites and see where organizations are proceeding in similar ways, and where approaches diverge. Although we recognize that there are many uses for data related to teacher effectiveness, and that many sites are grappling with how best to use this information to improve teacher quality, this report focuses specifically on evaluation practices.

Other organizations, including the National Council for Teacher Quality, The New Teacher Project, the Aspen Institute, and Bellwether Education Partners have provided invaluable information about the emerging state of teacher evaluation policy; but these policies inevitably leave holes to be worked out during implementation. We do not know of any other report that collects as much

¹ Weisberg, D, Sexton, S., Mulhern, J., & Keeling, D. (2009). *The widget effect*. The New Teacher Project. Available: <http://widgeteffect.org/downloads/TheWidgetEffect.pdf>

detailed technical information as this report to describe how states, districts, and charter organizations are implementing these evaluation policies. We hope that this report serves as a “go-to” tool for policymakers working to reform an evaluation system or create one anew.

We must also note what this report does not do. It does not assess the merits of the evaluation systems; instead, it aims to describe the processes sites are using at a particular time and explain, when possible, why they decided to go in that direction. Nor does this report aim to include all of the possible options for addressing the challenges highlighted, as many other works have already contributed much to that conversation.² Instead, this report focuses on the methods used across the featured sites. The report also does not address how other school staff members are evaluated. Although we recognize that other school staff, especially the school leadership team, are central to reform, and that many sites have developed or are developing robust systems to evaluate them, a survey of that work went beyond the scope of this report.

REPORT COMPONENTS

This report consists of four components:

1. This **brief**, which provides an overview of the report and background on the topics addressed in the other documents, including key questions, implementation options and trade-offs, and key terms
2. A **cross-site analysis** that looks across all 10 sites and summarizes key components of their evaluation systems
3. **10 detailed profiles** of the teacher evaluation systems at our featured sites
4. A **library of documents** that are the building blocks of the 10 systems

These components build on one another, with each providing increasing detail. This brief offers a logical starting point for understanding key issues and terms and getting an overview of teacher evaluation in the 10 sites. The cross-site analysis provides a high-level overview of the sites side by side, enabling users to compare systems and key topics. The site profiles offer details on each of the sites, covering all of the aspects highlighted in the cross-site analysis and more. Finally, the library of documents offers links to documents and reports by and about the sites.

Because all of these sites continue to develop their evaluation systems, this work represents a snapshot in time and is current as of March 15, 2012. We plan to update this report over time to reflect new developments.

² See, for example: The New Teacher Project. (2010). *Teacher evaluation 2.0*. Available: <http://tntp.org/ideas-and-innovations/view/teacher-evaluation-2.0>; Public Impact. (2009.) *How should states define teacher effectiveness?* Available: http://www.publicimpact.com/publications/PublicImpact-How_Should_States_Define_Teacher_Effectiveness.pdf; National Comprehensive Center for Teacher Quality. *Database on teacher evaluation policies*. Available: <http://resource.tqsource.org/stateevaldb/>; Curtis, R., & Wiener, R. (2012, March). *Means to an end: A guide to developing teacher evaluation systems that support growth and development*. Washington, DC: The Aspen Institute Education and Society Program. Available: <http://www.aspeninstitute.org/publications/means-end-guide-developing-teacher-evaluation-systems-support-growth-development>

METHODOLOGY

We spent several months identifying sites, determining which information to collect, scouring public documents, conducting site interviews, soliciting feedback, and revising our materials. Below we describe each of the steps we took to create the documents included in this report:

- Select sites.** We set out to identify 10 sites to feature for this report. We sought sites that, while still evolving, were advanced enough to describe in detail. Our initial list consisted of about two dozen sites, including states that had recently adopted new legislation related to teacher evaluation or won Race to the Top grants; districts that had received grants through the Bill & Melinda Gates Foundation or the Teacher Incentive Fund to develop their teacher evaluation systems; sites that had recently been featured in other research, news reports, or education blogs; and sites recommended by experts with whom we consulted. Using that list, we conducted preliminary research to determine whether the evaluation system was in effect as of the 2011–12 school year, and to gauge how much information we could collect on the sites. Based on our research, we narrowed our list to 10 sites: three states, five districts, a charter school network, and a graduate school of education.

FEATURED SITES		
Delaware	Hillsborough County, FL	Achievement First (CMO)
Tennessee	Houston, TX	Relay Graduate School of Education
Rhode Island	New Haven, CT	
	Pittsburgh, PA	
	Washington, D.C.	

- Develop a framework for collecting the data.** We drew on recent research to identify the most common and vexing challenges sites face as they attempt to establish or revise their teacher evaluation systems. Next, we developed a series of questions related to those challenges. We also consulted with several experts to review and revise our questions, and we continued to tweak the framework as we gathered additional information or when more appropriate questions came to light.
- Review public documents.** For each site, we scoured publicly available documents describing the evaluation system and how it works. These included official documents from the site, as well as previous reports and news stories describing the evaluation system. Many of these documents are hyperlinked in the site profiles or available in the document library for this report.
- Conduct site interviews.** We interviewed at least one representative from each of the sites, and often interviewed two or more to answer remaining questions. These interviews included a mix of state and district personnel and consulting groups involved in designing and implementing the evaluation system.
- Review process.** After we completed the report documents and reviewed them internally, at least one representative from each site reviewed all of that site’s materials.

IMPLEMENTATION CHALLENGES

The rest of this brief describes major implementation challenges identified in our site research and interviews, and the methods sites are using to address those challenges. The challenges generally fell into five categories:

- Student achievement measures
- Classroom observations
- Other nonacademic measures
- Accuracy, validity, and reliability
- Reporting and using evaluation results

Our goal here is not to describe each category or challenge in full—which would take many more pages than we have here—but to highlight and discuss a set of critical issues that any district, state, charter organization, or school of education is likely to face. Throughout this brief, we therefore provide references to other works that dive into these issues in more detail.

STUDENT ACHIEVEMENT MEASURES

A teacher's primary job is to improve student learning. In the past, most school systems have not included student achievement as a factor in a teacher's evaluation. That tendency is changing, but fairly and accurately tying student growth to a particular teacher has proved difficult.³ To name just a few challenges: Student assessment data does not tell us everything about a teacher; many teachers teach untested grades or subjects; and teachers work with students starting from different points and facing different challenges.⁴ All of the sites profiled in this report have concluded, however, that including student achievement measures is worth the effort, because student learning progress is the core result for which all of the schools and districts strive. In addition, researchers have reached positive conclusions about common measures of teachers' contributions to student learning in core subjects, such as the correlation with students' mastery of higher-order skills and the year-to-year consistency of scores.⁵

Not all of the sites we featured have rolled out their systems' student achievement component yet, but they have all at least drafted a plan to do so. All of the sites have therefore wrestled with such tough questions as: how to measure student performance in untested grades and subjects; methods for calculating student growth; whether and how to adjust scoring for untested grades and subjects; and how heavily to weigh student performance in a teacher's final rating. Each site's decisions and experiences with these issues can offer lessons to others taking on the same challenges.

³Weisberg, D, Sexton, S., Mulhern, J., & Keeling, D. (2009). *The widget effect*. The New Teacher Project. Available: <http://widgeteffect.org/downloads/TheWidgetEffect.pdf>

⁴For more, see: Stumbo, C., & McWalters, P. (December 2010/January 2011). Measuring effectiveness: What will it take? *The Effective Educator*. Vol. 68, No. 4. Available: <http://www.ascd.org/publications/educational-leadership/dec10/vol68/num04/Measuring-Effectiveness@-What-Will-It-Take%2%A2.aspx>; Goe, L., Bell, C., & Little, O. (2008, June). *Approaches to evaluating teacher effectiveness: A research synthesis*. Washington, DC: National Comprehensive Center for Teacher Quality. Available: <http://www.tqsource.org/publications/EvaluatingTeachEffectiveness.pdf>

⁵The Bill & Melinda Gates Foundation. (2012). *Gathering feedback for teaching*. Available: http://metproject.org/downloads/MET_Gathering_Feedback_Practioner_Brief.pdf; Glazerman, S. et al. (2010). *Evaluating teachers: The important role of value-added*. Washington, DC: Brookings Institution. Available: http://www.brookings.edu/reports/2010/11/17_evaluating_teachers.aspx

MEASURING STUDENT ACHIEVEMENT IN UNTESTED GRADES AND SUBJECTS

Most of the sites we reviewed use the results from state standardized tests to measure student achievement when those data are available. Most teachers do not teach a class for which there is a state standardized test, however.

Several sites are expanding their student assessment options by creating new standardized assessments in untested grades and subjects. Pittsburgh Public Schools, for example, is using locally developed assessments to complement statewide tests, as well as working to expand its assessment portfolio over time. In Hillsborough County Public Schools, the district has built out its assessment portfolio to include tests for every grade and subject the district offers.

Other sites use standardized test results from a class or subject that the teacher's work supports. In Tennessee, for example, state law requires that student growth data count for 35 percent of a teacher's overall rating. For untested grades and subjects, the state has provided guidance on appropriate school-level value-added scores that districts should use.

Most sites, however, are choosing alternate measures of student achievement for untested grades and subjects, including the following:

- **Existing assessments.** Many of the sites are vetting and approving assessments used elsewhere. In grades K–2, for example, our featured sites consistently approved using several nationally normed literacy assessments for teacher evaluation, including the Developmental Reading Assessment (DRA) 2 and the Fountas and Pinnell benchmark assessment system.
- **School or teacher-created assessments.** Most sites allow teachers to create their own assessments if an approved assessment is not available.
- **Portfolio.** In some grades and subjects that are particularly difficult to test, sites are assessing student performance using a portfolio, which includes student work samples and other materials that demonstrate growth. At Achievement First, for example, teachers of art, music, physical education, dance, and theater decide on their individual program goals and assessments with a network achievement director at the beginning of the year. At the end of the year, a content expert reviews teachers' portfolios of student assessments to determine their progress toward their goals and their impact on student achievement.

In addition, several sites are using *multiple measures* to assess student achievement, even when standardized test data are available. According to those interviewees, the rationale is that no single measure is perfect, but combining multiple measures diminishes the weaknesses of any particular measure. In Houston, the district plans to require every teacher to have at least two measures of student performance by 2012–13. Similarly, Delaware plans to measure student growth using three measures—a state assessment-based measure, an approved external/internal assessment measure, and a growth goal measure. Meanwhile, Pittsburgh served as the pilot site for the Measures of Effective Teaching (MET) project through the Bill & Melinda Gates Foundation, and is considering whether and how to include different measures that research has identified as highly correlated with student achievement results.

METHODS FOR CALCULATING GROWTH

After sites collect student performance data, they must still calculate how much learning growth students make with a given teacher. Value-added models and student growth percentiles consider a student's prior performance to determine whether they are making as much growth as they should be:

- **Value-added model.**⁶ Value-added models use a statistical formula to predict how much growth a student will make based on the amount of growth that similar students—including students performing at a similar level at the start of the year—make, or have made historically. Some models even aim to parse out individual teacher effects among teachers who team-teach.⁷ A teacher's effectiveness is therefore measured in terms of the growth her students make compared to the amount of growth the model predicts that those students will make.

Several sites, including Tennessee, Washington, Pittsburgh, and Achievement First use teacher-level value-added data to assess student growth.⁸ Two sites, Tennessee and Washington, also include school-level value-added data, which looks at student results across the entire school, rather than at the teacher level. Pittsburgh is considering including a school-level value-added measure in teachers' summative ratings as well.

- **Student growth percentile.**⁹ A student growth percentile shows student progress in comparison to his academic peers—students whose performance was similar on previous assessments. Each child receives a percentile rank, indicating the percentage of his academic peers that he out-grew. Rhode Island, the only site we looked at that uses student growth percentiles to measure growth, uses the median (middle) Student Growth Score (percentile rank) to summarize student growth across a teacher's class. This number captures the point at which half the students had higher levels of growth and half the students had lower levels of growth. If the median is greater than 50, the teacher has exceeded expected growth, and vice versa.

Sites also used other methods for calculating growth that do not include comparable data, such as progress towards growth goals and mastery of standards:

- **Growth from pre-test to post-test.** Several sites set a growth target, then measure student performance against that target using a pre-test and post-test. In New Haven, teachers work with administrators to develop growth goals for their students; for example, students will grow, on average, at least four units on the elementary reading assessment from the beginning of the year to the end. Teachers are then evaluated on the progress students make towards those growth goals.

⁶For more on value-added models, see: McCaffrey, D., Lockwood, J., Koretz, D., & Hamilton, L. (2003). *Evaluating value-added models for teacher accountability*. RAND Corporation. Available: http://www.rand.org/pubs/monographs/2004/RAND_MG158.pdf; Mathematica Policy Research. *Using value-added growth models to track teacher and school performance*. Princeton, NJ: Author. Available: http://www.mathematica-mpr.com/education/value_added.asp; Braun, H. (2005). *Using student progress to evaluate teachers: A primer on value-added models*. Princeton, NJ: Educational Testing Service. Available: <http://www.ets.org/Media/Research/pdf/PICVAM.pdf>

⁷Of the sites we reviewed, the value-added models used by Hillsborough County and Pittsburgh Public Schools aim to parse out individual teacher effects. In Rhode Island, the state weights teacher results to reflect the time each teacher spends with a particular student.

⁸Teacher-level value-added results do not currently contribute to a teacher's summative rating in Pittsburgh, but will beginning in the 2013–14 school year.

⁹For more on student growth percentiles, see: Betebenner, D. (2008). *A primer on student growth percentiles*. Dover, NH: National Center for the Improvement of Educational Assessment. Available: <http://www.cde.state.co.us/cdedocs/Research/PDF/Aprimeronstudent-growthpercentiles.pdf>

- **Mastery of standards.** In some grades and subjects, especially high school elective classes such as economics or journalism, growth goals are not appropriate because students enter the class with no or limited background in the subject. Some sites therefore evaluate teachers based on students' mastery of the standards instead. At Relay Graduate School of Education, teachers must demonstrate that their students, on average, have made at least a year's worth of growth or have mastered at least 70 percent of the grade-level standards to earn a degree.

Overall, our interviewees expressed that although student growth measures and our methods for calculating growth are imperfect, they remain useful and are constantly improving. In the long term, interviewees told us, they hope to find and create more reliable and valid assessments for non-tested grades and subjects that are independently administered and graded, but they acknowledge that such measures are not available in the short term. They also noted that the formulas underlying these scores must be robust, and the data systems in which test information is collected and analyzed must have the capacity to ensure the data are accurate and secure.

WEIGHTING STUDENT PERFORMANCE

At the end of the day, the evaluation systems we researched produce, or will soon produce, a summative score estimating the effectiveness of each teacher, which requires sites to combine all of the components by weighting them relative to their importance. In all of the systems we examined, student performance is a major factor in a teacher's final score. The sites differ, however, in how they weight student performance. Here are some of the approaches we saw:

- **Flat percentage.** Five sites use a flat weight for student performance. The weights fall between 40 percent (Hillsborough County and Achievement First) and 55 percent (Washington). Some sites lower that weight for teachers of untested grades and subjects, however. If value-added data are not available, student performance counts for just 15 percent of a teacher's final rating in Washington, and 20 percent at Achievement First.
- **Matrices.** Several sites are using or are planning to use a matrix that combines ratings on different scales into a single rating. New Haven, for example, rates each teacher on two factors: (1) student learning growth, and (2) instructional practice and professional values. A matrix then assigns a final rating to the teacher based on the combination of (1) and (2). For example, a teacher rated "strong" in growth and "effective" in practice and values would receive a final rating of "strong."
- **Rating ceiling.** Delaware places a ceiling on the final rating a teacher can receive based on her student's growth. Teachers must demonstrate satisfactory growth or better to be deemed effective or highly effective.
- **Providing a substitute measure.** At some sites, the evaluation system substitutes a teacher-level value-added score with another growth measure as needed. For example, if value-added data are not available for a particular teacher in Washington, that teacher receives a score for "teacher-assessed achievement data," a measure of student progress toward a target. Similarly, Houston will use teacher-level value-added data for any teacher for whom they are available, but has identified alternate measures for teachers of untested grades and subjects.

CLASSROOM OBSERVATIONS

In addition to measuring teacher effectiveness by using student performance data, all of the sites include classroom observation. There was a surprising amount of consensus around what to look for during observations. Most of the sites use Charlotte Danielson's *Framework for Teaching* to assess teacher practice, or have built their own rubric based on the Danielson model, which looks across four domains of teaching: planning and preparation, the classroom environment, instruction, and professional responsibilities.¹⁰ Who conducts those observations and how often teachers are observed varies from site to site.

OBSERVERS

The sites we spoke to identified several considerations related to choosing observers.

- Who understands the school and classroom context in which the teacher works?
- Who has the subject matter and grade-level expertise to provide a valid and meaningful evaluation?
- Who has the time to conduct all of the evaluations needed?
- Who is sufficiently objective to follow a rubric reliably?
- What approaches are affordable within our budgets?

Based on these questions, the sites generally chose among administrators, a third-party evaluator unaffiliated with school, and teacher peers within the school to serve as observers. As we describe below, each type of evaluator has pros and cons, so a particular type may be more appropriate in a specific situation:

- **Administrators and coaches.** Administrators and coaches see teachers every day and understand the conditions under which teachers work. Administrators may not have subject matter expertise, however, and may not have enough time in their schedules to get into the classroom and evaluate teachers as needed. And they may have trouble making objective determinations, especially as they get to know teachers over time. Administrators serve as observers in every school system we reviewed.
- **Third-party evaluators unaffiliated with the schools.** Third-party evaluators are usually subject-matter and grade-level experts with a history of high performance in the classroom. Their primary job is to go from school to school conducting teacher evaluations. They have the benefit of being impartial, because they do not know the teacher or the school, and their results serve as a check on results from other observers. But third-party evaluators also lack an understanding of the school and the special challenges a particular classroom may face. Third-party evaluators are paid for this work, increasing total evaluation costs. Third-party evaluators are a key component of several systems we reviewed, including Hillsborough County, Washington, and New Haven.

¹⁰ See the cross-site analysis and individual fact sheets for examples where sites did not use Charlotte Danielson's *Framework for Teaching*.

- **Peers within the school.** Like administrators, other teachers in the school understand the conditions under which the teacher works. They are also likely to have subject-matter expertise. While some districts have found peers to be “tougher” evaluators than administrators, peers arguably may have trouble making objective determinations when they have close relationships with the teachers being evaluated.¹¹ Both Pittsburgh Public Schools and Achievement First use peer evaluators from within the school, but their observations do not factor into teachers’ summative ratings.¹²

Sites do not necessarily use just one type of observer. Although administrators conducted observations at every site, sites sometimes used Third-party evaluators as well, allowing teachers to be observed more frequently when needed. At Achievement First, for example, where internal school leaders observe alongside a network instructional expert (regional superintendents or an achievement director with expertise in a particular content area), the use of multiple observers also ensures adherence to a network-wide standard of excellence and provides meaningful professional development for school leaders.

FREQUENCY OF OBSERVATIONS

Some sites observed all teachers an equal number of times; other sites observed teachers with varying frequency based on what would be most needed or helpful, given limited resources and a teacher’s evaluation results. In some systems, this broke down by the following types of teachers:

- **Low-performing teachers.** Additional observations can be a way to provide additional support for low-performing teachers. Such is the case in Delaware, where the average teacher receives just one observation annually, but low-performing teachers receive two.
- **High-performing teachers.** The frequency with which high-performing teachers were observed varied across sites. In New Haven, both low-performing and high-performing teachers receive additional observations. A third party evaluator conducts an additional observation to verify that the teacher truly performs at a high level, because such teachers are eligible to become a lead teacher, mentor, or coach, and may be called upon to help other teachers improve. In other systems, such as Washington’s, high-performing teachers have the option to receive fewer observations as a way to recognize some of the district’s best teachers and allow principals to spend more time supporting struggling teachers.
- **Novice teachers.** Additional observations can help provide more opportunities for feedback and development early in a teacher’s career, when data indicate teachers achieve the most professional growth.¹³ In Tennessee, for example, novice teachers receive six observations per year, while other teachers receive four. Interestingly, none of the sites we reviewed require fewer observations for tenured teachers.

¹¹ Papay, J., & Johnson, S. (2011). Is PAR a good investment? Understanding the costs and benefits of teacher peer assistance and review programs. *NGT Working Paper*. Available http://www.gse.harvard.edu/~ngt/new_papers/PAR%20Costs%20and%20Benefits%20-%20January%202011.pdf

¹² Teachers filling the Instructional Teacher Leader 2 (ITL2) position at PPS will conduct observations at their own schools in their first year as an ITL2, and those observations will be used for formative purposes, only. In their second and third years, ITL2s will observe teachers in other schools, serving more as third party evaluators, and those observations will be factored into teachers’ summative ratings. See fact sheet for more detail. At Achievement First schools, peer observations do not factor into a teacher’s final observation score.

OTHER NONACADEMIC MEASURES

Many of the sites we researched use other nonacademic measures in addition to classroom observation. By using a combination of academic and nonacademic measures, they aim to capture different aspects of effective teaching. Some of the other nonacademic measures sites are using or exploring include:

- **Student perceptions.** This year, Pittsburgh began a district-wide administration of the Tripod student survey, which measures a variety of nonacademic student outcomes. Preliminary results show a strong correlation between some elements of students' perceptions demonstrated on these surveys and teachers' value-added scores. The district has not yet decided, however, how and whether to use the results as a factor in teachers' overall ratings. Washington is also running a small pilot of the Tripod survey, but has no immediate plans to include student surveys as part of its teacher evaluation system. Rather, the district hopes to make the results available to teachers as a tool for improvement.
- **Student character.** Both Achievement First and Relay Graduate School of Education score teachers based on growth in student character. At Achievement First, the student character component is based on student and parent surveys about relationships and communication with students and families.
- **Peer ratings.** At Achievement First, a teacher receives a score for "core values and contribution input," which is based on a peer survey that assesses a teacher's core values and contribution to the AF mission.
- **Commitment/contribution to school community.** Every site we looked at included a measure of professionalism, such as collegiality and high expectations for students. Washington, however, includes an extra component: 10 to 15 percent of a teacher's final rating reflects her commitment and contribution to the school community. Administrators measure that commitment using a rubric that scores a teacher's support of local school initiatives, support of special education and English language learner programs, high expectations, partnership with families, and instructional collaboration.

VALIDITY AND ACCURACY OF DATA

All of the sites recognized that teacher evaluation results are valuable only if the measures used are accurate and valid. Sites are generally early in their efforts to assess data accuracy and validity, but are taking some steps to do so, including:

- **Evaluator training.** All of the sites train their evaluators. The length and intensity of the training differ by site, however, as do the amount and types of ongoing support. In Houston, for example, evaluators participated in a four-day training session over the summer this year. In Hillsborough County, training lasted for seven to 10 days and included conducting

¹³ See for example: Nye, B., Konstantopoulou, S., Hedges, L. (2004). "How Large Are Teacher Effects." *Educational Evaluation and Policy Analysis*. Vol. 26, No. 3. Pp 237-257. Available <http://steinhardt.nyu.edu/scmsAdmin/uploads/002/834/127%20-%20Nye%20B%20%20Hedges%20L%20%20V%20%20Konstantopoulou%20S%20%20%20%282004%29.pdf>; Clotfelter, C., Ladd, H., Vigdor, J. (2007). "Teacher Credentials and Student Achievement in High School: A Cross-Subject Analysis with Student Fixed Effects." Calder Center. Available Clotfelter, C.T., Ladd, H.F., Vigdor, J.L., "Teacher Credentials and Student Achievement in High School: A Cross-Subject Analysis with Student Fixed Effects," Calder Center, October 2007; Harris, D., Sass, T (2007), "Teacher Training, Teacher Quality, and Student Achievement," Calder Center. Available http://www.caldercenter.org/PDF/1001059_Teacher_Training.pdf

paired observations with another evaluator. In Pittsburgh, all evaluators participate in the Instructional Quality Assurance Certification (IQA-C) Process, which includes two certification levels focusing on rating accuracy and instructional feedback and support.

- **Evaluator tests.** At several sites, including Delaware, Tennessee, and Houston, evaluators must pass a test before they can evaluate teachers.
- **Evaluators' ratings.** Evaluators are themselves evaluated, based in part on the accuracy and validity of their ratings. This is most common at sites that use third-party evaluators.
- **Third party evaluators used.** The use of a third-party evaluator serves as a check on administrator ratings in the building.
- **Procedure to investigate misalignment between measures.** In the case that the score on different measures, such as student performance and observation, are significantly different, several sites will conduct a review. In Rhode Island, for example, the district will review the evaluation data for a particular teacher if a teacher has an extremely high score for student achievement, but an extremely low score for professional practice and responsibilities, adjusting as needed.
- **Regular data checks and follow-ups.** In Rhode Island and a handful of other sites, district officials are continually monitoring evaluation data. In instances when the data seem “off,” such as results from a particular school or evaluator being very different from others, district officials will share the results with those involved to determine the cause of the discrepancy.
- **Reviews of student performance results.** Every year, Relay Graduate School of Education randomly reviews the student performance results of about 5 to 10 percent of its teachers. During the review, Relay GSE reviews all documentation and has a conference with the teacher about his or her students' work.
- **Principal discretion combined with superintendent sign-off.** At Achievement First, teachers wanted principals to have more discretion, rather than less. Principals can review student data and use other evidence of teacher performance to add points to the “student achievement outcome” component. In those instances, a regional superintendent reviews the principals' decisions.

REPORTING AND USING EVALUATION RESULTS

REPORTING RESULTS

Once teacher evaluations are complete, sites must decide what to do with the results. At every site, teachers and school leaders received their scores. Most sites also made public the distribution of ratings across all schools in the system. For example, Washington released the percentage of teachers falling into each performance category. In Rhode Island, the state department of education also plans to publish report cards for all educator preparation programs in the state based on teacher evaluation results beginning at the end of the 2014–15 school year.¹⁴

There seems to be little effort or plan to report scores in any further detail. Sites continue to struggle with the question of how to share evaluation information, but generally hesitate to report the data in a way that makes it possible for the public to link results to a specific teacher.

USING RESULTS

All the sites in this report use teacher evaluation results to incentivize and reward different behaviors and outcomes that pay off for students, and to help administrators make decisions about individual teachers to improve a school's performance. Across the sites, states, districts, and schools are using the results to determine:

- **Professional development.** Almost all of the sites reported using teacher evaluation data to create targeted professional development. Several sites also reported using evaluation results to identify teachers for an improvement plan.
- **Bonuses and salary increases for top performers.** Based on their performance, effective teachers at Achievement First schools can earn significantly more than current salary scales within the surrounding traditional public school districts. In Washington, teachers deemed “highly effective” can receive bonuses of up to \$25,000. If they earn “highly effective” ratings for multiple years, they can increase their base salary by as much as \$27,000.
- **Promotion/increased responsibility.** Several systems offer or plan to offer opportunities for teachers to take on additional responsibility or new roles. In Hillsborough County, for example, top-performing teachers can apply to become a peer evaluator. In New Haven, exemplary teachers are eligible to be a lead teacher, mentor, or coach.
- **Student assignment.** Rhode Island plans to implement a policy in which no student will have a low-performing teacher two years in a row.
- **Tenure.** In Tennessee, new teachers can receive tenure only if they teach in the district for five years and attain a rating in the top two evaluation categories for the previous two years. In Pittsburgh, teachers must earn six satisfactory ratings before receiving tenure.
- **Licensure/Certification.** In Delaware, new teachers must earn a satisfactory rating at least twice within three years to earn a continuing license. Similarly, teachers enrolled at Relay Graduate School of Education cannot earn a degree unless they demonstrate measureable growth in the classroom. And in Rhode Island, even experienced teachers will not be able to renew their certification if they are deemed ineffective for five years. Rhode Island is also the only site that uses differentiated certification (initial, professional, and advanced) based on evaluation results.

¹⁴ Louisiana has long been the leader in matching teacher evaluation results to teacher preparation programs. See, for example: <http://www.regents.doa.louisiana.gov/assets/docs/TeacherPreparation/RegentsReceipt11FINAL.pdf>

- **Dismissal.** Several sites consider evaluation results in dismissal decisions. In Hillsborough County, for example, teachers who receive a rating of 1 or 2 for two consecutive years could be recommended for dismissal. Teacher dismissal procedures in Hillsborough County and several other sites we reviewed, however, are often set by state statute and have not changed as a result of new evaluation systems.

CONCLUSION

The additional sections of this report provide more detail on the 10 evaluation systems we reviewed. As we mentioned at the start of this brief, none of these systems claims to have cracked the code for teacher evaluation. Nonetheless, we consistently heard that the perfect should not be the enemy of the good. As one interviewee described his system's latest student performance measure: "This is an imperfect measure. It's the best one we ever had."

Although they are still works in progress, the evaluation frameworks in place at these 10 sites successfully use multiple measures to gauge teacher effectiveness. Gathering data on teacher effectiveness is only the first step, however. What sites do with the data is just as critical. The sites we reviewed use the data they collect to differentiate teacher quality, which allows them to reward excellence, remove poor performers, use talent more effectively, provide targeted professional development, and elevate the teaching profession. Most important, these evaluation systems and the measures collected provide a key instrument for improving student achievement for more than 1.6 million children every year.



CROSS-SITE ANALYSIS

This section of *Measuring Teacher Effectiveness* looks across all 10 sites and summarizes key components of each evaluation system. It provides a high-level overview of the sites to enable side-by-side comparisons and allow readers to zero in on the most relevant sites and topics. For more detail on any of the systems, please see the site profiles.

The tables in this section compare and contrast the following major components of the teacher evaluation systems we reviewed:

- Background on each evaluation system
- Student achievement measures
- Classroom observations
- Other nonacademic measures
- Accuracy, validity, and reliability
- Reporting and using evaluation results

		STATES			DISTRICTS						OTHER	
		DELAWARE	RHODE ISLAND	TENNESSEE	HILLSBOROUGH COUNTY, FL	HOUSTON, TX	NEW HAVEN, CT	PITTSBURGH, PA	WASHINGTON, D.C.	ACHIEVEMENT FIRST (CMO)	RELAY GRADUATE SCHOOL OF EDUCATION	
BACKGROUND ON EVALUATION SYSTEMS	Number of students	130,610	143,793	933,703	192,547	202,773	20,759	25,000	45,000	6,200	21,000	
	Number of schools	208	300	1,736	250	298	45	60	125	20	93	
	Number of teachers	8,594	14,260	64,229	12,468	12,829	1,850	2,000	4,000	585	420	
	Most recent evaluation system implemented systemwide in...	2011-12	2011-12	2011-12	2010-11	2011-12	2010-11	2010-11	2009-10	2010-11	2011-12	
	Other staff evaluated using an aligned process	Principals	X	X	X	X	X ¹⁵	X	X ¹⁶	X	X	n/a
		Non-teaching staff	X		X	X		X		X		n/a
	How often are teachers evaluated?	Teachers in general	1x/2yrs	1x/yr	1x/yr	1x/yr	1x/yr	1x/yr	1x/yr	1x/yr	1x/yr	Once at end of year 2
		High-performing teachers	1x/2yrs	Evaluation intervals do not change based on teacher performance								n/a
		Low-performing teachers	1x/yr	Evaluation intervals do not change based on teacher performance								n/a
		Novice teachers	1x/yr	Evaluation intervals do not change based on tenure / experience						2x/yr	Evaluation intervals do not change based on tenure/ experience	n/a
Tenured/ experienced teachers		Varies (1x/1-2 yrs)	Evaluation intervals do not change based on tenure / experience						1x/yr	n/a		

¹⁵ Houston is currently designing an aligned principal appraisal system.

¹⁶ Pittsburgh's principal evaluation system is being updated to align with the district's teacher evaluation system.

			STATES			DISTRICTS					OTHER			
			DELAWARE	RHODE ISLAND	TENNESSEE	HILLSBOROUGH COUNTY, FL	HOUSTON, TX	NEW HAVEN, CT	PITTSBURGH, PA	WASHINGTON, D.C.	ACHIEVEMENT FIRST (CMO)	RELAY GRADUATE SCHOOL OF EDUCATION		
STUDENT ACHIEVEMENT MEASURES	Student assessments used for evaluation	Standardized state assessment	X	X	X	X	X	X	X	X	X			
		Existing assessment	X	X	X	X	X	X	X	X	X	X		
		School- or teacher-created	X	X	X		X	X		X	X	X		
		Portfolio	X		X		X	X		X	X			
	What is the student achievement measure for tested grades/subjects?	Teacher-level value-added measure				X	X			X	X	X		
		Student growth percentile		X										
		Growth from pre-test to post-test	X	X			X	X					X	
		Mastery of standards	X ¹⁷	X	X								X	
		Other (e.g., achievement)												
	What is the student achievement measure for untested grades/subjects? ¹⁷	Value-added measure from a related subject or grade				X	n/a all students are tested			TBD				
		Student growth percentile		X										
		Growth from pre-test to post-test	X	X	X ¹⁸			X	X			X	X	X
		Mastery of standards	X ¹⁹	X	X			X				X		X
		Other (e.g., teacher-assessed achievement data)				X						X		
Does the system include a school-level student achievement measure?				X				TBD	X					
Weight of student achievement component in tested grades/subjects		Varies / Uses Rating ceiling	Varies / Uses matrix	50%	40%	Varies / Uses matrix	Varies / Uses matrix	TBD	55%	40%	45%			
Weight of student achievement component in untested grades/subjects		Varies / Uses Rating ceiling	Varies / Uses matrix	50%	n/a	Varies / Uses matrix	Varies / Uses matrix	TBD	15%	20%	45%			

¹⁷ Represents all possibilities, although not all measures are used for all teachers. See individual site profiles for more detail.

¹⁸ Currently being piloted for untested grades and subjects.

¹⁹ The Delaware Department of Education plans to look at external assessments that include proficiency/attainment until enough data is gathered to calculate growth.

			STATES			DISTRICTS					OTHER		
			DELAWARE	RHODE ISLAND	TENNESSEE	HILLSBOROUGH COUNTY, FL	HOUSTON, TX	NEW HAVEN, CT	PITTSBURGH, PA	WASHINGTON, D.C.	ACHIEVEMENT FIRST (CMO)	RELAY GRADUATE SCHOOL OF EDUCATION	
CLASSROOM OBSERVATIONS	Who are the observers who rate teacher effectiveness?	Administrators or coaches	X	X ²⁰	X	X	X	X	X	X	X		
		Third-party evaluators unaffiliated with school				X		X	X	X	X		
		Peers within the school											
		Other (e.g., regional superintendent, achievement director)			X						X	X	
	How do teachers receive feedback after an observation?	Conference	X	X	X	X	X	X	X	X	X	X	X
		Written feedback	X	X	X	X	X	X	X	X	X	X	X
	How often are teachers observed ²¹	Teachers in general	1x/yr	4x/yr	4x/yr	5x/yr	4x/yr	1x/yr ²²	4x/yr	5x/yr	3x/yr ²²	2-3x/yr	
		Low-performing teachers	2x/yr		4x/yr	>5x/yr ²³		4x/yr ²²	4x/yr	5x/yr			
		High-performing teachers	1x/yr		4x/yr	5x/yr		3x/yr ²²	4x/yr	2x/yr			
		Novice teachers	3x/yr		6x/yr	n/a		1x/yr ²²	8x/yr	5x/yr			
		Tenured teachers	Varies		4x/yr	n/a		Varies	4x/yr	5x/yr			
	Are formal observations announced?	Yes				X						X	
		No					X ²⁴				X		
		Sometimes	X	X	X			X	X	X			
What is the basis for the observation rubric?	Charlotte Danielson	X			X			X			X		
	Other (e.g. TAP, Teaching as Leadership, etc.)			X					X	X	X		

²⁰ Although most formal observations are conducted by principals or assistant principals, LEAs may determine who primary evaluators will be, as long as they receive training.

²¹ These groupings are the authors' creation. Sites may differentiate observations based on different teacher categories or terms.

²² Includes formal observations only. Number of informal observations not set, but "frequent."

²³ Hillsborough County, FL, conducts up to 11 observations, including 7 formal and 4 informal.

²⁴ Houston refers to observations as "longer" and "shorter" rather than "formal" and "informal."

		STATES			DISTRICTS					OTHER	
		DELAWARE	RHODE ISLAND	TENNESSEE	HILLSBOROUGH COUNTY, FL	HOUSTON, TX	NEW HAVEN, CT	PITTSBURGH, PA	WASHINGTON, D.C.	ACHIEVEMENT FIRST (CMO)	RELAY GRADUATE SCHOOL OF EDUCATION
What additional measures are included?	Student perceptions			X ²⁵				X ²⁶		X	
	Student character									X	X
	Peer ratings									X	
	Other (e.g., completion of class modules, contribution to school community—see profiles for details)		X						X		X

²⁵ Piloted this year in Memphis, but will expand next year.

²⁶ Pittsburgh Public Schools is administering a district-wide student survey this year, but has not committed to using the results to determine a teacher's evaluation score.

			STATES			DISTRICTS					OTHER		
			DELAWARE	RHODE ISLAND	TENNESSEE	HILLSBOROUGH COUNTY, FL	HOUSTON, TX	NEW HAVEN, CT	PITTSBURGH, PA	WASHINGTON, D.C.	ACHIEVEMENT FIRST (CMO)	RELAY GRADUATE SCHOOL OF EDUCATION	
ACCURACY, VALIDITY, & RELIABILITY	What steps is the site taking to ensure that results are accurate and valid?	Evaluator training	X	X	X	X	X	X	X	X	X	X	
		Evaluator tests	X		X	X	X		X				
		Evaluator ratings			X	X					X	X	
		Third-party evaluators used				X		X	X	X	X	X	
		Procedure to investigate misalignment between measures		X					X	X		X	X
		Regular data checks and follow-ups (e.g. district/school follow-ups)		X	X	X	X			X	X	X	
		Review of student performance results		X	X								X
		Other (e.g., peer survey, principal discretion)				X				X		X	

		STATES			DISTRICTS					OTHER		
		DELAWARE	RHODE ISLAND	TENNESSEE ²⁷	HILLSBOROUGH COUNTY, FL	HOUSTON, TX	NEW HAVEN, CT	PITTSBURGH, PA	WASHINGTON, D.C.	ACHIEVEMENT FIRST (CMO)	RELAY GRADUATE SCHOOL OF EDUCATION	
REPORTING AND USING RESULTS	How many rating categories are there?	4	4	5	5	4	5	4	4	5	n/a (scale 1-100)	
	How are results reported and used?	Teachers and administrators receive score	X	X	X	X	X	X	X	X	X	X
		Results are used to evaluate teacher preparation programs			X		X					X
	Results influence...	Professional development	X	X	X	X	X	X	X	X	X	n/a ²⁹
		Bonuses and salary increases for top performers	X		X	X			X	X	X	n/a ²⁹
		Promotion/increased responsibility		X	X	X		X	X	X	X	n/a ²⁹
		Student assignment		X	X							n/a ²⁹
		Tenure	n/a ²⁸		X		n/a ²⁸		X	n/a ²⁸	n/a ²⁸	n/a ²⁹
		Licensure	X	X	X							X ²⁹
		Dismissal		X	X	X	X	X		X		
Other (e.g., commendation, school-based appreciations, special recognition, etc.)	X		X			X		X	X	X ²⁹		
Teachers can appeal their ratings	X	X	X	X	X	X	X	N/A ³⁰	³¹	X	X	

²⁷The First to the Top Act requires that annual evaluations be a factor in personnel decisions, including promotion, retention, tenure, and compensation. The details of such decisions, however, are left to the district except with respect to tenure.

²⁸Delaware does not award "tenure," so teachers are either considered "novice" or "experienced" (holding a valid and current continuing or advanced license). Similarly, Houston, Washington, D.C., and Achievement First do not award "tenure."

²⁹Relay GSE does not issue a teaching license, but recommends teachers for certification if they complete the Relay GSE program. In addition, teachers receive a degree only if they earn a high enough score. All other "n/a"s to not apply to Relay GSE because it is a graduate school of education.

³⁰The district and union have appeal processes.

³¹Through the chancellor's appeals process, District of Columbia Public Schools provides teachers who earn ineffective or minimally effective ratings with an opportunity to submit an appeal if they do not believe the evaluation procedures were followed appropriately (e.g. a post-observation conference was not provided within 15 days). Teachers may appeal the evaluation process only, rather than observation scores themselves. The Washington Teachers' Union also provides teachers with an opportunity to submit appeals.



SITE PROFILES

DELAWARE

DELAWARE PERFORMANCE APPRAISAL SYSTEM FOR TEACHERS (DPAS II-REVISED)

OVERVIEW

Delaware’s Performance Appraisal System (DPAS) has undergone many changes since first being piloted in 1986. DPAS I, in effect until 2006–07, evaluated teachers on four components: planning and preparation, classroom environment, instruction, and professional responsibilities. DPAS II, implemented in 2007–08, added student growth as a fifth component in teacher evaluations. Then in 2010, Delaware’s Race To The Top application outlined revisions strengthening the student growth component.

In the state’s current evaluation system (DPAS II-Revised), the first four components of the evaluation are based on teacher observation. Delaware uses expert evaluators from the state and development coaches—former administrators and teacher leaders—to train administrators, calibrate observation results, and model observations and conference skills. Planning and preparation, classroom environment, instruction, and professional responsibilities have equal weight in a teacher’s rating, but a teacher cannot be deemed effective or highly effective if she does not produce at least a satisfactory rating on student growth, regardless of her ratings for the other components.

The Delaware Department of Education (DDOE) continues to work with stakeholder groups to refine the student growth component, including determining how much growth teachers must make to be deemed effective. In its current interim plan, DDOE uses multiple measures to evaluate student growth and provides teachers with options for calculating each measure. DDOE is also convening working groups of teachers to develop pre- and post-assessments for all grades and subjects.

ABOUT THE SCHOOL SYSTEM

<i>Number of students</i>	130,610
<i>Number of schools</i>	20,759
<i>Number of teachers</i>	8,594

BACKGROUND

<i>Corresponding legislation</i>	Race To The Top (RTTT) application revised DPAS-II
<i>Year legislation passed</i>	2010
<i>Was the system piloted?</i>	Yes
<i>In effect since</i>	<ul style="list-style-type: none"> • DPAS II: 2007–08 • DPAS II-Revised: Interim year 2011–12 (includes revised student growth component, but does not factor into overall rating; also ties rewards to evaluation results)
<i>Are there plans for additional phases / components?</i>	<ul style="list-style-type: none"> • 2011–12 is considered an “interim year,” and does not include the student growth component • Beginning in 2012–13, both positive and negative consequences will be tied to evaluation results, which will include the student growth component
<i>Who gets evaluated?</i>	All part-time and full-time teachers, specialists, and administrators

SITE PROFILES

<p><i>How often do teachers receive a rating?</i></p>	<ul style="list-style-type: none"> • Novice teacher: Summative evaluation once a year • Experienced teachers with most recent overall rating of highly effective: minimum of one summative evaluation every two years • Experienced teachers with most recent overall rating of effective, plus satisfactory ratings on at least four of the five evaluation components: minimum of one summative evaluation every two years • Experienced teachers who have not earned overall ratings of highly effective or effective: once-a-year summative evaluation <p>Note: Delaware does not award “tenure,” so teachers are either considered “novice” or “experienced” (holding a valid and current Continuing or Advanced License, respectively).</p>
<p><i>Can LEAs devise their own effectiveness measures where not already defined in legislation?</i></p>	<ul style="list-style-type: none"> • Yes, all educators must be evaluated annually using the DPAS II-Revised evaluation tool. However, LEAs may administer other evaluations in addition to DPAS II-Revised, as long as these are reported to DDOE • If an LEA wishes to use a locally developed evaluation system that is the result of a collective bargaining process, it may apply for a waiver of the DPAS II-Revised provisions. The DDOE will review the system to ensure that it: <ul style="list-style-type: none"> - is educationally sound and rigorous - measures student growth using multiple measures - has mechanisms in place to certify evaluators and validate results - does not apply to educators holding an initial license • The state also mandates that all districts use a state-created assessment system (DCAS), but enables districts to propose other models that meet state requirements such as: <ul style="list-style-type: none"> - LEA-wide diagnostic assessments (such as Reading inventory, Gates McGinitie) - LEA-wide common assessments (such as written curriculum-based assessments, performance-based assessments, student portfolios)
<h3>STUDENT ACHIEVEMENT MEASURES</h3>	
<p><i>Assessments to measure student achievement</i></p>	<ul style="list-style-type: none"> • Delaware Department of Education’s (DDOE) State assessment: <u>Delaware Comprehensive Assessment System (DCAS)</u> <ul style="list-style-type: none"> - DCAS math and reading assessments for grades 3–10 <ul style="list-style-type: none"> » Frequency: three times a year (fall, winter, spring) - DCAS Alternate Assessment (DCAS-Alt1)
<p><i>Are any new assessments in development to use with teacher evaluations?</i></p>	<p>Delaware is developing the internal measures that can be used by districts for the student growth component of the evaluation system. These measures must be approved by the Department of Education, and include:</p> <ul style="list-style-type: none"> • Internal pre- and post-assessments for non-DCAS content/subject teachers who report student grades • Growth goals for non-DCAS content/subject teachers who do not report student grades

<p><i>What academic/achievement criteria are included?</i></p>	<p>The student growth component is still in development. In 2011–12, only DCAS-based student growth will be calculated for DCAS-tested educators to determine highly effective educators. The 2012–13 school year plan for measuring student growth includes a DCAS measure, internal or external measures, and growth goals.</p> <ul style="list-style-type: none"> • DCAS Measure <ul style="list-style-type: none"> - Student DCAS fall-spring growth score (called “instructional score”) in reading and/or math for all of teacher’s students - This measure applies to all DCAS subject and grade teachers (grades 3–10, reading and math) • Internal or External Measures <ul style="list-style-type: none"> - Includes two options: <ul style="list-style-type: none"> » External measures: Non-DCAS assessments <u>approved by the DDOE</u> that are standards-based and show fall-to-spring growth³² <ul style="list-style-type: none"> › Examples: district-wide Measures of Academic Progress (MAP assessments) or IEP progress › Internal measures: statewide pre- and post-assessment created by educators <ul style="list-style-type: none"> › All are approved by the DDOE and meet criteria of <u>internal measures rubric</u> - Measures are selected by teacher and evaluator between August and October of the school year - This measure applies to all DCAS subject and grade teachers and non-DCAS teachers who give grades (e.g., science or social studies teachers) • Growth Goals <ul style="list-style-type: none"> - Each educator uses 15 indicators to measure growth, which include statewide and district-wide indicators that measure what educators are doing to assist in academic growth of students based on available data, job descriptions, and standards - This measure applies to non DCAS-subject teachers and teachers who do not give student grades (e.g., counselors and nurses)
<p><i>How much do the student achievement measures count in a teacher’s final rating?</i></p>	<ul style="list-style-type: none"> • Beginning in 2012–13, teachers will be able to earn an “effective” rating or better only if their students make satisfactory growth or better (an amount to be determined) • The various growth measures of student achievement will be weighted in the following ways: <ul style="list-style-type: none"> - For DCAS subject and grade teachers: <ul style="list-style-type: none"> » DCAS Measure: 50% » Internal or External Measures: 50% - For non-DCAS subject teachers: <ul style="list-style-type: none"> » Internal or External Measures: 50% » Growth Goals: 50% - For non-subject teachers who do not give student grades: <ul style="list-style-type: none"> » Growth Goals: 100%

³²DDOE plans to look at external assessments that include proficiency/attainment until enough data is gathered to calculate growth.

SITE PROFILES

NONACADEMIC MEASURES	
<i>What nonacademic evaluation criteria are included?</i>	<p>There are four nonacademic components in DPAS II-Revised: Planning and preparation, classroom environment, instruction, and professional responsibilities (see p. 7-34 of guide).</p> <ul style="list-style-type: none"> • Component 1: Planning and Preparation <ul style="list-style-type: none"> - Criteria are: selecting instructional goals; designing coherent instruction; demonstrating knowledge of content and pedagogy; demonstrating knowledge of students; and designing student assessments • Component 2: Classroom Environment <ul style="list-style-type: none"> - Criteria are: managing classroom procedures; managing student behavior; creating an environment to support learning; and organizing physical space • Component 3: Instruction <ul style="list-style-type: none"> - Criteria are: engaging students in learning; demonstrating flexibility and responsiveness; communicating clearly and accurately; using questioning and discussion techniques; and using assessment in instruction • Component 4: Professional Responsibilities <ul style="list-style-type: none"> - Criteria are: communicating with family; recording data in a student record system; growing and developing professionally; and reflecting on professional practice
<i>How much do the nonacademic measures count in a teacher's final rating?</i>	<p>Nonacademic components are weighted equally. Teacher receives one summative rating for these components (see p. 79 and 92 of guide) of satisfactory or unsatisfactory.</p>
CLASSROOM OBSERVATIONS	
<i>Who are the observers who rate teachers' effectiveness?</i>	<p>Certified evaluators: Usually principals or other administrators in the school administration.</p>
<i>What training do observers receive?</i>	<ul style="list-style-type: none"> • All evaluators must complete four training modules online, and pass all online quizzes, which total approximately six hours. If an evaluator does not pass a quiz after three tries, he will not be certified for the upcoming year • Evaluators earn a certification from DDOE that is valid for five years, and renewable upon completion of DDOE's DPAS II-Revised professional development requirements • Expert evaluators and development coaches receive further training from DDOE and the Delaware Academy for School Leadership (DASL). <p>For more information on administrator, evaluator, and teacher training, see here.</p>

<p><i>Are observers compensated for this work?</i></p>	<p>No, observations are part of the administrator's job description/salary.</p>
<p><i>How often are teachers observed?</i></p>	<ul style="list-style-type: none"> • New teachers: minimum of 3x/year • Experienced teachers with most recent overall rating of highly effective: minimum of 1x/year • Experienced teachers with most recent overall rating of effective, plus ratings of satisfactory on at least four of the five components: minimum of 1x/year • Experienced teachers who are not rated highly effective or effective: minimum of 2x/year
<p><i>Is there a time frame within which observations must be conducted?</i></p>	<p>Observations occur after the first five full school days of the year, and before the last five days of the school year.</p> <p>Suggested target dates for completion of observations:</p> <ul style="list-style-type: none"> • New teachers: <ul style="list-style-type: none"> - Observation 1: by October 31 - Observations 2 and 3: by March 31 • Experienced teachers: <ul style="list-style-type: none"> - Observation: by January 31
<p><i>Are observations announced or unannounced?</i></p>	<ul style="list-style-type: none"> • New teachers: two announced and one unannounced • Experienced teachers rated highly effective or effective: one announced • Experienced teachers not rated highly effective or effective: one announced, one unannounced
<p><i>What is the basis for the observation rubric?</i></p>	<ul style="list-style-type: none"> • DDOE based the observation rubric on Charlotte Danielson's <i>Framework for Teaching</i>.
<p><i>What is included in the observation scoring rubric?</i></p>	<p>Evaluators use the <u>Components rubrics (pp. 114-131 of guide)</u>, based on <u>Charlotte Danielson's <i>Framework for Teaching</i></u>, during observations for nonacademic components:</p> <ul style="list-style-type: none"> • Component 1: Planning and Preparation • Component 2: Classroom Environment • Component 3: Instruction • Component 4: Professional Responsibilities
<p><i>How are observers held accountable for their scoring?</i></p>	<ul style="list-style-type: none"> • Administrators go through an online training program • DDOE is developing a credentialing process
<p><i>Do teachers have an opportunity to debrief with observers?</i></p>	<ul style="list-style-type: none"> • All announced observations must be preceded by a pre-observation conference • All observations are followed by a post-observation conference within 10 days of the conference. Evaluators give <u>formative feedback (see p. 70 of guide)</u> with expectations and recommendations for teacher improvement • All teachers participate in a summative evaluation conference that discusses all previous observations

SITE PROFILES

ACCURACY, VALIDITY, AND RELIABILITY	
<i>Is there a process to validate non-standardized assessments?</i>	<ul style="list-style-type: none"> • Non-DCAS assessments <u>approved by the DDOE</u> that are standards-based and show fall-to-spring growth or proficiency • Teacher-selected pre- and post-assessments must meet criteria of <u>internal measures rubric</u>
<i>Are there procedures to ensure that observation scores are valid?</i>	<ul style="list-style-type: none"> • State development coaches spend four hours a week training administrators to ensure proper evaluation of teachers and administrators <ul style="list-style-type: none"> - The position of development coach was created through Delaware’s Race to the Top application to work with districts and administrators across the state to ensure DPAS II- Revised is being implemented correctly - Development coaches do not observe or rate teachers • District expert evaluators (district office administrators, lead mentors, or development coaches) will model effective evaluation and monitor observations and evaluations
<i>Is there a procedure to validate that observation scores are correlated with student outcomes?</i>	No
<i>Is there a procedure to validate that other nonacademic measures are correlated with student outcomes?</i>	No
USING AND REPORTING EVALUATION RESULTS	
<i>What are the rating categories?</i>	<p>Teachers will receive an overall rating of highly effective, effective, needs improvement, or ineffective.</p> <ul style="list-style-type: none"> • Highly effective: teacher earns satisfactory ratings on at least four of the five components, including rating of “exceeds” on growth component (component 5) • Effective: teacher earns satisfactory ratings on at least three of the five components, but does not receive a rating of “exceeds” on growth component (component 5) • Needs improvement: teacher earns one or two satisfactory ratings on five components, including growth component (component 5); OR, teacher earns three or four satisfactory ratings on five components, plus rating of “unsatisfactory” on growth component (component 5) • Ineffective: teacher earns zero to two satisfactory ratings on five components, plus rating of “unsatisfactory” on growth component (component 5); OR, teacher’s overall rating is needs improvement for three consecutive years
<i>How are results reported?</i>	<ul style="list-style-type: none"> • Individual teacher results are reported to teachers and administrators only • RTTT commitment is to report evaluation results publicly by district, school, and teacher preparation program in the near future

<p><i>Are there consequences tied to evaluation results? If yes, what are they?</i></p>	<p>Yes, evaluation results are or will be tied to pay, professional development, and licensure. Other consequences are delivered at the district's discretion.</p> <ul style="list-style-type: none"> • Professional development/improvement plan <ul style="list-style-type: none"> - A teacher will receive an improvement plan for any of the following situations: <ul style="list-style-type: none"> » Teacher does not meet expectations set by evaluators during conferences » Teacher receives "unsatisfactory" rating on a nonacademic component during a lesson observation » Teacher is rated unsatisfactory on any of the nonacademic components (Components 1–4) on overall evaluation » Teacher receives an overall rating of ineffective or needs improvement. Teacher receives a rating of unsatisfactory on growth component (Component 5) (will not apply during interim year 2011–12) • Pay <ul style="list-style-type: none"> - Beginning in 2012–13, RTTT attraction and retention bonuses given to teachers who maintain highly effective rating in a high-needs school or transfer to a high-needs school - In year one, only DCAS reading and math teachers will be eligible • Licensure <ul style="list-style-type: none"> - A teacher who began teaching in the 2010–11 school year must have at least two years of "satisfactory" ratings in growth component (Component 5) within a three-year period to secure her continuing license. The interim 2011–12 school year will not count as part of the three-year period. • Other <ul style="list-style-type: none"> - Teachers who excel in any criterion or element of evaluation, or new teachers who demonstrate substantial professional growth, are eligible for a "commendation," additional comments in the evaluation that recognize outstanding achievement - If there is a "pattern of ineffective teaching" (see p. 58 of Guide) a teacher may be dismissed at the local school board's discretion - Rewards such as promotions or increased responsibilities are delivered at district's discretion
<p><i>Can teachers appeal their ratings? If yes, how?</i></p>	<ul style="list-style-type: none"> • Teacher may submit a written challenge to the evaluator's supervisor, an administrator, for: <ul style="list-style-type: none"> - Any lesson observation rating - Any component or overall rating on final evaluation • Teacher must submit specific information to evaluator's supervisor within 15 working days of teacher's receipt of evaluation • Supervisor must hold a challenge hearing with teacher to review and discuss challenge and evaluation record within 15 working days of receipt of challenge • Supervisor must issue a written decision within 15 working days of challenge hearing. If challenge is denied, decision must state reasons for denial. Supervisor's decision is final <p>Note: If evaluator and teacher work in the same building, supervisor will be a district or charter management organization-level credentialed evaluator.</p>
<p><i>If performance results are available, what is the score distribution?</i></p>	<p>Data are currently unavailable. However, Delaware anticipates a distribution for overall rating (for 2013–14) of:</p> <ul style="list-style-type: none"> • 30% Highly effective • 50% Effective • 15% Needs improvement • 5% Ineffective

SITE PROFILES

TEACHER INVOLVEMENT AND CONTINUING IMPROVEMENT	
<p><i>What role did teachers play in developing the evaluation system?</i></p>	<ul style="list-style-type: none"> • More than 500 teachers participated in working groups to develop internal measures of student growth. These working groups are ongoing <ul style="list-style-type: none"> - Teachers were nominated by districts and unions - Teachers make high-level decisions, which are reviewed by DDOE and the Technical Advisory Group • DDOE hosted forums to gather teacher input throughout design and revision of DPAS II-Revised
<p><i>What role did the union play in developing the evaluation system?</i></p>	<ul style="list-style-type: none"> • Union involved in drafting regulations and ensuring compliance with RTTT commitments • Union makes recommendations to recruit teachers and administrators to develop aspects of DPAS II-Revised • Former president of union is now special assistant for DPAS II-Revised in the DDOE
<p><i>Is there a formal process for revising the evaluation system over time?</i></p>	<ul style="list-style-type: none"> • Every year since the inception of DPAS II-Revised, DDOE has contracted with Progress Education Corporation to conduct a full <u>evaluation of the system</u>. Progress studies current manuals, forms, and historical documents; administers surveys; conducts interviews; and facilitates focus groups to gather qualitative and quantitative data on: <ul style="list-style-type: none"> - Criteria used in the DPAS II-Revised system - Forms used for evaluations - Manageability of the entire system - Accuracy and reliability of the data being used in the system - Usefulness of the training sessions and manuals - Needed modifications - Efficacy of the DPAS II-Revised program in achieving quality assurance and professional growth • Using data collected, Progress develops a set of recommendations for revising the system for the next year

GLOSSARY	
<i>Component</i>	<ul style="list-style-type: none"> • One of the five areas of teacher practice and responsibility that teachers are evaluated against. • Components 1–4 are nonacademic • Component 5 is student growth
<i>Criterion</i>	<ul style="list-style-type: none"> • Subcategory of knowledge/skills within a component
<i>Development coach</i>	<ul style="list-style-type: none"> • Development coaches help deliver regional trainings to all administrators and work with administrators to norm and calibrate observation results. They do not observe or rate teachers. • Development coaches are typically former administrators or veteran teacher leaders. • There are nine coaches working in 60 schools across the state. Their contract is for two years. Next year, 10 coaches will serve 80 schools. • The position of development coach was created through Delaware’s Race to the Top application to work with districts and administrators across the state to ensure DPAS II-Revised is being implemented correctly
<i>Experienced teacher</i>	<ul style="list-style-type: none"> • A teacher holding a valid and current Continuing or Advanced License issued prior to August 1, 2003³³ • All other teachers are considered “novice”
<i>Expert evaluator</i>	<ul style="list-style-type: none"> • District office administrators, lead mentors, or development coaches that model effective evaluation and monitor observations and evaluations
<i>External measures</i>	<ul style="list-style-type: none"> • Non-DCAS assessments <u>approved by the DDOE</u> that are standards-based and show fall-to-spring growth • Examples: district-wide Measures of Academic Progress (MAP assessments) or IEP progress
<i>Internal Measures</i>	<ul style="list-style-type: none"> • May be internal measures: teacher-selected pre- and post-assessments • Must be approved by DDOE and meet criteria of <u>internal measures rubric</u> • Measures are selected by teacher and evaluator between August and October of the school year
<i>New (or Novice) teacher</i>	<ul style="list-style-type: none"> • Teacher holding a valid and current Initial License
<i>Pattern of Ineffective Teaching</i>	<ul style="list-style-type: none"> • See the <u>Guide, p. 58</u> for chart that shows possible consecutive ratings considered to be a pattern of ineffective teaching

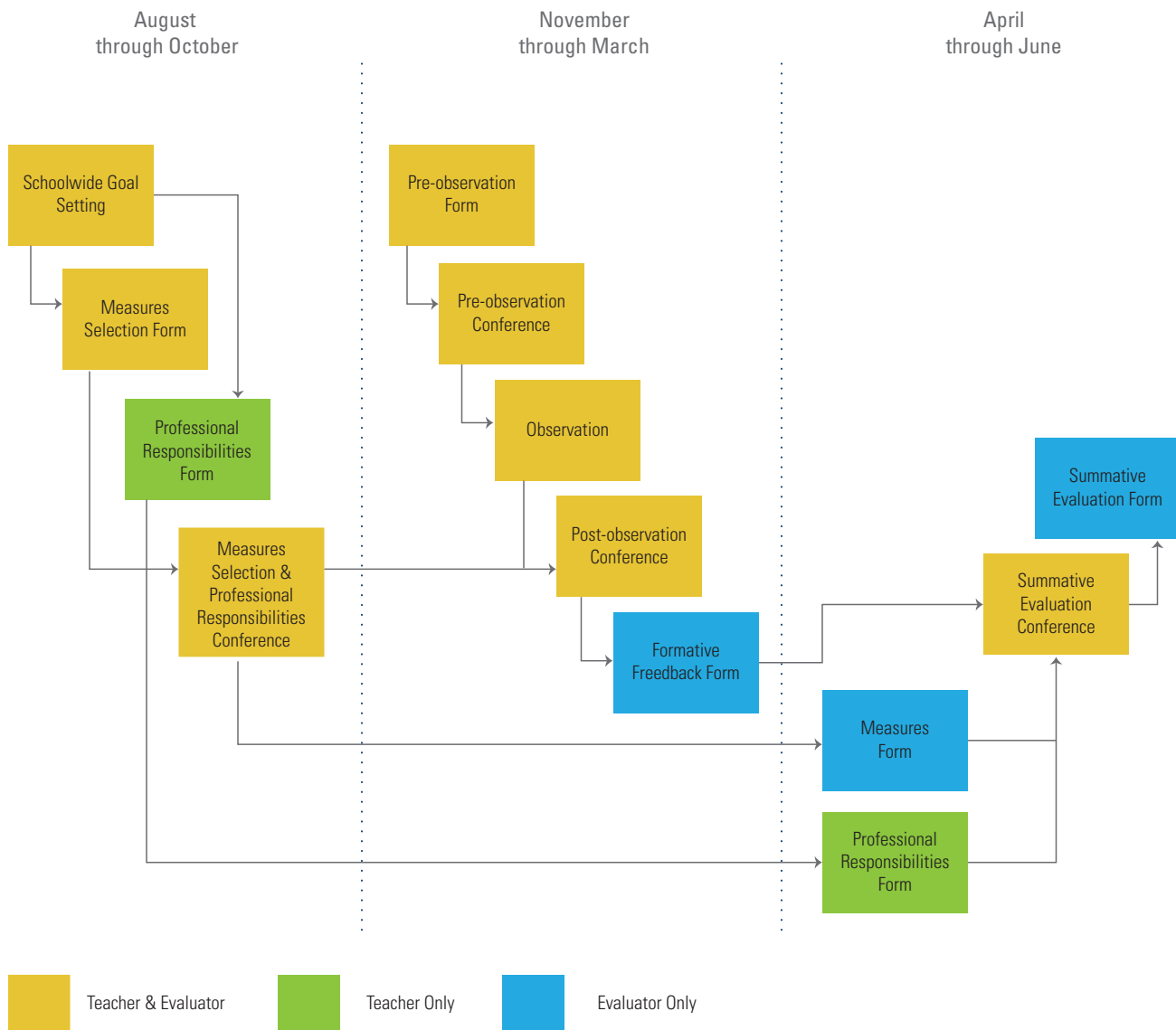
³³Delaware does not award “tenure,” so teachers are either considered “novice” or “experienced” (holding a valid and current Continuing or Advanced License).

SITE PROFILES

TIMELINE

DPAS II-REVISED - APPRAISAL CYCLE

This diagram illustrates a suggested timeline for the DPAS II evaluation process for teachers. This timeline is intended as a guide, not a mandate.



Note: Observation feedback process may be repeated using announced or unannounced observations

SOURCES

Delaware Department of Education. (2007). *Competitive sealed proposal for the evaluation of the Delaware Performance Appraisal System Second Edition (DPAS II)*. Retrieved May 1, 2012, from <http://www.doe.k12.de.us/rfp/DPASIIIRFP2008FINAL.pdf>

Delaware Department of Education. (2011). *DPAS II guide for teachers*. Retrieved November 20, 2011, from <http://www.doe.k12.de.us/csa/dpasii/ti/DPASII TeachFullGuide.pdf>

Delaware Department of Education. (2011). *DPAS II summative ratings for 2011–2012*. Retrieved November 20, 2011, from <http://www.doe.k12.de.us/csa/dpasii/summative2011final.shtml>

Delaware Department of Education. (2011). *Non-regulatory guidance for the DPAS II process for teachers and specialists*. Retrieved November 20, 2011, from <http://www.doe.k12.de.us/csa/dpasii/new/DPASII NonregGuidance-8-4-11.pdf>

Delaware Department of Education. "State Profile." Retrieved November 20, 2011 at: <http://profiles.doe.k12.de.us/SchoolProfiles/State/Account.aspx>

Delaware State Education Association. (2011). "Getting DPASII-REVISED right for 2011-2012." DSEA Action! Vol. 32, No. 1. Available: <http://digital.copcomm.com/issue/40928>

Progress Education Corporation. Annual DPAS II Evaluation Reports. Available: <http://www.doe.k12.de.us/csa/dpasii/default.shtml>

<h2 style="margin: 0;">RHODE ISLAND</h2> <h3 style="margin: 0;">RHODE ISLAND MODEL EDUCATOR EVALUATION SYSTEM</h3>									
<p>OVERVIEW</p> <p>Rhode Island began efforts to overhaul its teacher evaluation system in 2009. Those efforts received a boost in 2010 when the state won a Race to the Top grant, leading the state to roll out the Rhode Island Model Educator Evaluation System in 2011.</p> <p>The Rhode Island model will measure student growth using two different measures: a Growth Model score (available 2012–13) and a Student Learning Objective score. The Growth Model score will measure a student’s growth on the state exam in comparison to his/her academic peers by assigning students a growth percentile based on their relative growth. In addition, all teachers work with their administrators to set student learning objectives (SLOs), a long-term academic goal that is specific and measurable. Teachers then receive a score based on the degree to which students reached the SLO. If a Growth Model score is not available for a particular teacher, the SLO score will count as the only student growth measure. Teachers also receive scores for nonacademic measures, including professional practice and professional responsibility. Teachers receive ratings for each component, which are then put into a matrix to calculate a final evaluation score.</p> <p>The Rhode Island Department of Education (RIDE) has decided to implement a limited new evaluation system (fewer observations, no growth model rating, overall rating used for developmental purposes only) during the 2011–12 school year so district leaders can receive feedback and educators can have an adjustment period before tying consequences directly to ratings in 2012–13. RIDE faces the challenge of refining the system based on feedback from 2011–12 in time for full implementation in 2012–13. This timeline requires quickly incorporating numerous changes to the system by June 2012, when training begins for full implementation.</p>									
<p>ABOUT THE SCHOOL SYSTEM</p> <table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 30%;"><i>Number of students</i></td> <td>143,793</td> </tr> <tr> <td><i>Number of schools</i></td> <td>300</td> </tr> <tr> <td><i>Number of teachers</i></td> <td>14,260</td> </tr> </table>		<i>Number of students</i>	143,793	<i>Number of schools</i>	300	<i>Number of teachers</i>	14,260		
<i>Number of students</i>	143,793								
<i>Number of schools</i>	300								
<i>Number of teachers</i>	14,260								
<p>BACKGROUND</p> <table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 30%;"><i>Corresponding legislation</i></td> <td>No new legislation, but Board of Regents adopted into regulation new Educator Evaluation Standards in 2009</td> </tr> <tr> <td><i>Year legislation passed</i></td> <td>Not applicable</td> </tr> <tr> <td><i>Was the system piloted?</i></td> <td>Yes, in Spring 2011</td> </tr> <tr> <td><i>In effect since</i></td> <td>Gradual implementation began in all districts in 2011–12, with fewer observations and no RI Growth Model rating. Overall ratings for teachers are calculated, but they are used for developmental purposes only (see Guide, p. 8, for more details on gradual implementation requirements for 2011–12).</td> </tr> </table>		<i>Corresponding legislation</i>	No new legislation, but Board of Regents adopted into regulation new Educator Evaluation Standards in 2009	<i>Year legislation passed</i>	Not applicable	<i>Was the system piloted?</i>	Yes, in Spring 2011	<i>In effect since</i>	Gradual implementation began in all districts in 2011–12, with fewer observations and no RI Growth Model rating. Overall ratings for teachers are calculated, but they are used for developmental purposes only (see Guide, p. 8 , for more details on gradual implementation requirements for 2011–12).
<i>Corresponding legislation</i>	No new legislation, but Board of Regents adopted into regulation new Educator Evaluation Standards in 2009								
<i>Year legislation passed</i>	Not applicable								
<i>Was the system piloted?</i>	Yes, in Spring 2011								
<i>In effect since</i>	Gradual implementation began in all districts in 2011–12, with fewer observations and no RI Growth Model rating. Overall ratings for teachers are calculated, but they are used for developmental purposes only (see Guide, p. 8 , for more details on gradual implementation requirements for 2011–12).								

<p><i>Are there plans for additional phases/components?</i></p>	<ul style="list-style-type: none"> • Full implementation in 2012–13, with the exception of RI Growth Model rating; all teachers will receive an overall rating • Growth Model rating will be included in the overall rating in 2013–14 • Evaluation model for support professionals (social workers, nurses, counselors, school psychologists, and central office staff) in development
<p><i>Who gets evaluated?</i></p>	<p>All teachers and administrators</p>
<p><i>How often do teachers receive a rating?</i></p>	<p>Once a year</p>
<p><i>Can LEAs devise their own effectiveness measures where not already defined in legislation?</i></p>	<ul style="list-style-type: none"> • The state provides an evaluation model, but allows districts to propose other models that meet state requirements • Each district will have a District Evaluation Committee (DEC) that will make implementation decisions related to: <ul style="list-style-type: none"> - Teacher appeals - Improvement plans - Local assessments and rubrics
<p>STUDENT ACHIEVEMENT MEASURES</p>	
<p><i>Assessments to measure student achievement</i></p>	<p><u>New England Common Assessment Program (NECAP)</u></p> <ul style="list-style-type: none"> • Grades 3–8 • Reading and math
<p><i>Are any new assessments in development to use with teacher evaluations?</i></p>	<p><u>Partnership for Assessment of Readiness for College and Careers (PARCC)</u></p> <ul style="list-style-type: none"> • PARCC will provide K–12 assessments aligned to Common Core State Standards in English language arts and math • Assessments throughout the year will include summative and non-summative tests, a speaking and listening component, and variety of constructed response items, performance-based tasks, and computer-scored items • PARCC assessments will begin in 2014–15 • Once enough data are available, RIDE will use PARCC results to measure proficiency and growth for grades 3–11

SITE PROFILES

<p><i>What academic/achievement criteria are included?</i></p>	<p>The student learning component of the evaluation is based on two criteria: a teacher's Growth Model score (available 2013–14 and her Student Learning Objective score:</p> <ul style="list-style-type: none"> • The <u>Rhode Island Growth Model</u> (RIGM) score measures student growth on the state exam in comparison to their academic peers by assigning students a growth score. <ul style="list-style-type: none"> - Teachers in tested grades and subjects: The RIGM score is based on student growth on the NECAP reading and math tests for grades 3–7 - Contributing educators <ul style="list-style-type: none"> » All teachers who contribute to the literacy or mathematics development of students (e.g., special educators, co-teaching educators, specialists) are considered “contributing educators” » RIDE provides a <u>policy guide</u> to districts that offers guidelines for identifying contributing educators and suggestions for how growth should be attributed to them - All other teachers. Teachers who do not contribute to the literacy or mathematics development of students in tested grades and subjects do not receive an RIGM score - Growth score. RIGM uses median growth percentiles, rather than a value-added measure³⁴ <ul style="list-style-type: none"> » Each student's annual growth on his NECAP score is compared to the growth of his or her academic peers (students with a similar test score history) » Each student then receives a Student Growth Score (SGS) based on his relative performance » Teachers receive a score of Low Growth, Typical Growth, or High Growth based on the median SGS of her class over two years (cut scores for each level have not yet been determined) • <u>Student learning objectives</u> (SLOs) are specific and measurable goals aligned to state standards and based on available prior student data <ul style="list-style-type: none"> - All teachers work with evaluators and administrators to set two to four SLOs. Objectives must be <u>approved</u> based on three criteria (see <u>student learning guide, p. 17</u>): <ul style="list-style-type: none"> » Alignment to state standards » Rigor of target » Quality of evidence to measure progress - SLOs may measure either growth or mastery (See more examples <u>here</u>) <ul style="list-style-type: none"> » Example of mastery objective: <i>Students will demonstrate an understanding of United States government (local, state, national) by identifying elected leadership titles/basic role at different levels of government (e.g., mayor is the leader of a city, governor is the leader of the state, president is the leader of the country)</i> » Example of growth objective: <i>Students will improve their ability to compose informative/explanatory writing, including a clear thesis statement, organized body paragraphs with supporting evidence, and a concluding statement</i> - Teachers and evaluators will meet midyear to discuss progress toward their SLOs and revise them as needed - At the end of the year, evaluators will review results of assessments and other evidence to determine the extent to which the SLOs were met (<u>see rubric on student learning guide, p. 21</u>)
--	--

³⁴See the introductory brief for more information on student growth percentiles and value-added measures.

<p><i>How much do the student achievement measures count in a teacher's final rating?</i></p>	<ul style="list-style-type: none"> • Student Learning score is calculated using the Student Learning Matrix (Guide, p. 65), and includes: <ul style="list-style-type: none"> - Results from RI Growth Model for tested subjects and grades and “contributing educators” where defined (to be included in 2012–13) - Attainment of student learning objectives • Teachers of untested grades and subjects will receive a Student Learning score based solely on student learning objectives • For 2011–12, the Student Learning score will only include attainment of student learning objectives for all teachers • The final teacher rating is calculated using a matrix that combines the Student Learning score and Professional Practice and Professional Responsibilities score (Guide, p. 66) <ul style="list-style-type: none"> - Because the final ratings are calculated through a matrix, the weight of each component varies depending on the teacher’s performance on other pieces, but student performance accounts for roughly half of a teacher’s final evaluation score
<p>NONACADEMIC MEASURES</p>	
<p><i>What nonacademic criteria are included?</i></p>	<p>The Rhode Island model uses two nonacademic components: professional practice and professional responsibilities.</p> <ul style="list-style-type: none"> • Professional Practice criteria are planning and preparation; classroom instruction; classroom environment; and assessment, reflection, and improvement • Professional Responsibilities criteria are collaborating and contributing to the school community; belief in and advocating for students; creating a culture of respect; and exercising professional judgment and development
<p><i>How much do the nonacademic measures count in a teacher's final rating?</i></p>	<ul style="list-style-type: none"> • Professional Practice and Professional Responsibilities score is calculated using a PPxPR matrix (Guide, p. 64) • The final teacher rating is calculated using a matrix that combines the Student Learning score and Professional Practice and Professional Responsibilities score (Guide, p. 66) <ul style="list-style-type: none"> - Because the final ratings are calculated through a matrix, the weight of each component varies depending on the teacher’s performance on other pieces, but professional practice and responsibilities count for roughly half of a teacher’s final evaluation score
<p>CLASSROOM OBSERVATIONS</p>	
<p><i>Who are the observers who rate teachers' effectiveness?</i></p>	<p>There are two kinds of evaluators in the Rhode Island Model: primary and complementary.</p> <ul style="list-style-type: none"> • Primary evaluators: <ul style="list-style-type: none"> - Usually principals or assistant principals, but LEAs may determine who primary evaluators will be, as long as they receive training - Primary evaluators have sole responsibility for assigning evaluation ratings • Complementary evaluators: <ul style="list-style-type: none"> - May be a department head, content-area expert, or administrator from within the school who assists primary evaluator by conducting observations, gathering evidence, or providing feedback - May be an Intermediary Service Provider (ISP) that RIDE assigns to districts to train and guide principals during the first few years of implementation - ISPs are usually retired administrators, consultants, or district officials - There are approximately 25 ISPs working across all of the state’s 52 school districts

SITE PROFILES

<p><i>What training do observers receive?</i></p>	<p>All evaluators (primary and complementary) must attend a four-day academy of training sessions led by experienced teachers and administrators trained by RIDE.</p> <ul style="list-style-type: none"> • Each preliminary session focuses on one aspect of the evaluation system: <ul style="list-style-type: none"> - Student learning - Professional Growth Plans - Observations and feedback - Conferencing • Evaluators will also receive follow-up training provided by RIDE throughout the school year • Beginning in 2012–13, RIDE will set expectations that all evaluators must meet in order to successfully complete training
<p><i>Are observers compensated for this work?</i></p>	<p>ISPs are compensated for their work. Over time, the ISP position will be transitioned out as Race To The Top funding expires</p>
<p><i>How often are teachers observed?</i></p>	<p>Teachers receive at least four observations each year, including:</p> <ul style="list-style-type: none"> • At least one formal observation: at least 30 minutes • Informal observations: about 15 minutes
<p><i>Is there a time frame within which observations must be conducted?</i></p>	<ul style="list-style-type: none"> • Formal observations: at least one must be conducted before midyear conference • Informal observations: no time frame
<p><i>Are observations announced or unannounced?</i></p>	<ul style="list-style-type: none"> • Formal observations: announced • Informal observations: unannounced
<p><i>What is the basis for the observation rubric?</i></p>	<p>RIDE reviewed TAP, Danielson, CLASS, and IMPACT rubrics while designing the system's observation rubric</p>
<p><i>What is included in the observation scoring rubric?</i></p>	<p>Observation Rubric has criteria related to:</p> <ul style="list-style-type: none"> • Planning and Preparation • Classroom Instruction • Classroom Environment • Assessment, Reflection and Improvement • Professional Responsibilities
<p><i>How are observers held accountable for their scoring?</i></p>	<ul style="list-style-type: none"> • RIDE is monitoring the implementation of ratings during its gradual implementation year • Additional calibration tools will be introduced during the 2012–13 implementation year
<p><i>Do teachers have an opportunity to debrief with observers?</i></p>	<ul style="list-style-type: none"> • All observations followed by written feedback to teachers within two to three school days • Formal observations require post-observation conferences

ACCURACY, VALIDITY, AND RELIABILITY

<p><i>Is there a process to validate non-standardized assessments?</i></p>	<p>All assessments used to measure student learning objectives must be approved by evaluators using RIDE criteria (Guide, p. 50).</p>
<p><i>Are there procedures to ensure that observation scores are valid?</i></p>	<p>See below</p>
<p><i>Is there a procedure to validate that observation scores are correlated with student outcomes?</i></p>	<ul style="list-style-type: none"> • At the state level, RIDE will: <ul style="list-style-type: none"> - Periodically audit the evaluation process to ensure that evaluations are fair and accurate, and that they adhere to the Rhode Island Educator Evaluation System Standards • At the district level, District Evaluation Committees are responsible for: <ul style="list-style-type: none"> - Reviewing the accuracy and utility of the data produced - Reviewing the evaluation decisions made for fairness and consistency - Providing procedural safeguards to ensure the integrity of the system, including evaluation appeals - In the event that an evaluation process yields a contradictory outcome (e.g., a teacher has an extremely high Student Learning rating and an extremely low rating in Professional Practice and Professional Responsibilities), a review of the evaluation will be conducted at the district level • RIDE has committed to using RTTT funding to create an online Educator Performance Support System (EPSS) that maintains all evaluation data that RIDE can monitor
<p><i>Is there a procedure to validate that other nonacademic measures are correlated with student outcomes?</i></p>	<p>Yes, see above.</p>

SITE PROFILES

USING AND REPORTING EVALUATION RESULTS	
<i>What are the rating categories?</i>	Teachers receive a summative rating of highly effective, effective, developing, or ineffective.
<i>How are results reported?</i>	<ul style="list-style-type: none"> • Individual teacher results are reported to teachers and administrators only • Teachers and administrators will have access to evaluation results throughout the year on RIDE’s computer-based Educator Performance and Support System (EPSS) beginning in 2012–13
<i>Does the system allow teacher evaluation results to be reported by the training programs teachers attended?</i>	<ul style="list-style-type: none"> • RIDE will publish report cards for all educator preparation programs in the state based on teacher evaluation results • Report cards will be available to the public at the end of 2014–15 school year
<i>Are there consequences tied to evaluation results? If yes, what are they?</i>	<p>Yes, evaluation results are tied to promotion, professional development, and certification. RIDE plans to tie student assignment to evaluation results in the near future.</p> <ul style="list-style-type: none"> • Professional development/improvement plan <ul style="list-style-type: none"> - All teachers are required to have an annual professional development plan - Teachers who are rated developing or ineffective will be placed on an improvement plan for the following year • Promotion (effective January 2012) <ul style="list-style-type: none"> - RIDE issues differentiated teacher certificates (initial, professional, and advanced) that recognize different stages of development and accomplishment for teachers. Districts are encouraged to create career ladders for teachers. See regulations here • Student assignment (in development) <ul style="list-style-type: none"> - RI has committed to using RTTT funding to develop a system that ensures that a child will not have an ineffective teacher two years in a row • Certification (regulation effective January 2012) <ul style="list-style-type: none"> - Rhode Island educators who are rated highly effective, effective or developing on annual evaluations will be eligible for renewal of their certification - Experienced educators who receive evaluations of ineffective for five years in a row will not be able to renew their certification - New teachers with initial certification who receive evaluations of ineffective for three years in a row will lose their certification - RIDE is developing a process of training and supervised student teaching for teachers to be reinstated if they lose their certification - Certification is required to teach in all Rhode Island public schools • Performance pay, increased responsibility, tenure, and dismissal consequences are district-level decisions <ul style="list-style-type: none"> - If a teacher is ineffective for two years, RIDE recommends that districts move the teacher toward dismissal
<i>Can teachers appeal their ratings? If yes, how?</i>	Appeals will be handled at the district level in accordance with district policy and practice, collective bargaining agreements, and/or processes set by the District Evaluation Committee. State guidelines to govern district-level appeals are under development.
<i>If performance results are available, what is the score distribution?</i>	Data are unavailable.

TEACHER INVOLVEMENT AND CONTINUING IMPROVEMENT

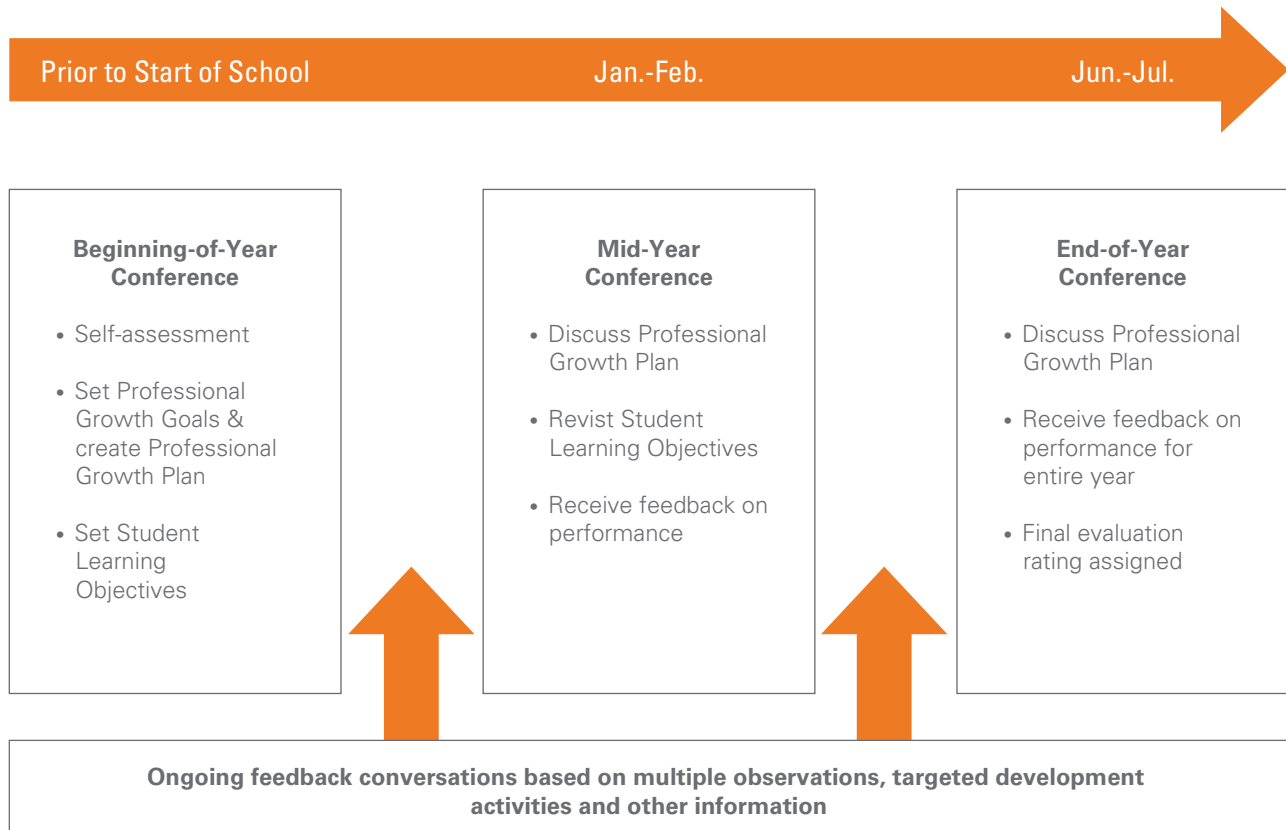
<p><i>What role did teachers play in developing the evaluation system?</i></p>	<ul style="list-style-type: none"> • During development of the evaluation system, RIDE created the Advisory Committee for Educator Evaluation Systems (ACEES) that <ul style="list-style-type: none"> - Was composed of parents, students, and educators - Advised RIDE on design of the RI Model - Worked alongside the Technical Advisory Committee • Teachers also sit on District Evaluation Committees, which make implementation decisions at the district level and provide feedback to RIDE
<p><i>What role did the union play in developing the evaluation system?</i></p>	<p>Union members sit on District Evaluation Committees, which make implementation decisions at the district level and provide feedback to RIDE.</p>
<p><i>Is there a formal process for revising the evaluation system over time?</i></p>	<p>The model will continue to be improved based on:</p> <ul style="list-style-type: none"> • Feedback from District Evaluation Committees, composed of union leaders, school committee members, parents, teachers and administrators • Feedback from the state Technical Advisory Committee (TAC), a group of national education and assessment experts who advised on methodologies and implementation • Formal reviews of evaluation data

SITE PROFILES

GLOSSARY	
<i>Complementary evaluator</i>	<ul style="list-style-type: none"> • May be a department head, content-area expert, or administrator from within the school who assists primary evaluator by conducting observations, gathering evidence, or providing feedback • May be an Intermediary Service Provider (ISP) whom RIDE assigns to districts to train and guide principals during the first few years of implementation; ISPs are usually retired administrators, consultants, or district officials
<i>District Evaluation Committee (DEC)</i>	<ul style="list-style-type: none"> • Oversees the implementation of educator evaluation in each local school system and ensures that the system is fairly and accurately administered
<i>Educator Performance Support System (EPSS)</i>	<ul style="list-style-type: none"> • RIDE’s online database that maintains all evaluation data • Teachers and administrators will have access to evaluation results throughout the year on EPSS • RIDE will use EPSS to monitor data for distribution and validity • EPSS will be implemented in 2012–13
<i>Intermediary Service Provider (ISP)</i>	<ul style="list-style-type: none"> • RIDE assigns ISPs to districts to train and guide principals during the first few years of implementation • ISPs are usually retired administrators, consultants, or district officials • There are approximately 25 ISPs working across all of the state’s 52 school districts • ISPs are compensated for their work • Over time, the ISP position will be transitioned out as Race To The Top funding expires
<i>Primary evaluator</i>	<ul style="list-style-type: none"> • May be principals, assistant principals, or department heads • Have sole responsibility for assigning evaluation ratings
<i>Student learning objectives</i>	<ul style="list-style-type: none"> • Learning goals set by teachers and approved by evaluators • Must be specific and measurable, aligned to state standards, and based on available prior student data • Objectives may measure either growth or mastery (see student learning guide for more information)
<i>Technical Advisory Committee (TAC)</i>	<ul style="list-style-type: none"> • Committee composed of national experts on assessment, performance management, and evaluation systems • Advises RIDE on all technical aspects of the model, including rating methodologies, Student Learning Objectives and the Rhode Island Growth Model

TIMELINE

TEACHER EVALUATION AND DEVELOPMENT PROCESS



SOURCES

Rhode Island Department of Education. *Office of Educator Quality and Certification*. <http://www.ride.ri.gov/educatorquality/educatorevaluation/Default.aspx>

Rhode Island Department of Education. *Rhode Island Growth Model*. <http://www.ride.ri.gov/assessment/RIGM.aspx>

Rhode Island Department of Education. (2011). *The Rhode Island Model: Guide to evaluating building administrators and teachers, 2011–2012*. Retrieved December 2, 2011, from <http://www.ride.ri.gov/educatorquality/educatorevaluation/Docs/RIModelGuide.pdf>

Rhode Island Department of Education. (2011). *Total populations of students, teachers and schools in Rhode Island*. Retrieved December 13, 2011, from <http://infoworks.ride.ri.gov/ri/statewide-data>