

The Evergreen Effect: Washington's Poor Evaluation System Revealed

When the state of Washington identified the bottom 5 percent of its lowest-performing schools in 2011, five of the 19 schools in Pasco School District made the list.¹ All elementary schools, they had failed to meet academic

performance goals every year since the state began measuring schools against them in 2006. Over a three-year period, only 30 percent of the students at these institutions tested at the proficient level in math and reading. The schools also suffered from chronic teacher absences—nearly half of the teachers took at least 10 days of vacation a year, a figure far above the national rate.²

And yet to judge by some of the actions of Pasco officials, all seemed well. District Superintendent Sandra Hill declined to take advantage of the federal government's School Improvement Grant program, which could have given the schools up to \$2 million each to turn themselves around.³ Even more remarkable, virtually no educator in any of the troubled schools had received a bad performance review. Under a perfunctory system in which employees were rated simply as satisfactory or unsatisfactory, 155 of the 156 teachers working at the schools, and all five of the principals, were rated satisfactory.⁴

Disparities like these have led states and districts around the country to make radical changes to their systems for evaluating teacher performance. And Washington is taking the lead. Laws recently passed by the state legislature have, among other things, increased the amount of time that third-year teachers—those on the cusp of earning tenure—must be observed, and they require that reviewers now consider objective data about student performance. Lawmakers acted because they had evidence that existing evaluations were virtually meaningless. It is only now, however, that we can determine just how bad they really were.

According to newly available data, released as part of a recent federal funding requirement, districts across the Evergreen State have used rudimentary evaluation systems like Pasco's that utterly fail to differentiate between truly great teachers and those who should be dismissed.⁵ What's more, the vague and often euphemistic labels that districts have used to describe performance suggest that they are uncomfortable even talking about poor instruction. This raises the obvious question: If districts can't identify their employees' strengths and weaknesses, how can they reward and promote top-performers and help the low-performers improve?



Since the Washington law was passed, there have been encouraging signs from pilot projects in the state that districts are becoming more discerning in how they judge teacher performance. But as these and other districts start implementing their new evaluations, they can learn a lesson from other states that have been early adopters of new systems: truly meaningful improvement—resulting in honest ratings—will require more than just tweaking requirements. It will take far more candor than what now characterizes conversations about teacher performance. The new data shows that nearly every school in the state has much work to do if it is to honestly appraise its educators and wisely use the results to make consequential decisions.

The Evergreen Effect: Most Rely on Simple Evaluations

The now-familiar “widget effect,” describing schools’ practice of treating teachers like interchangeable parts, was coined in 2009 by TNTP (formerly The New Teacher Project), the nonprofit that works to improve hiring practices in urban districts. Researchers reviewed 12 evaluation systems in four states and found that, like Pasco, they suffered from three common problems. One, most districts rated employees as simply satisfactory or not, with nothing above or between. Two, only tiny percentages of teachers were rated unsatisfactory. And three, districts by and large did not use the evaluation results to make important personnel decisions.⁶ The findings, which have subsequently been replicated elsewhere,⁷ prompted states and districts to push for evaluation systems that consider objective evidence of student growth and use evaluation results to reward good teachers, dismiss poor ones, and give struggling but promising teachers the help they need.

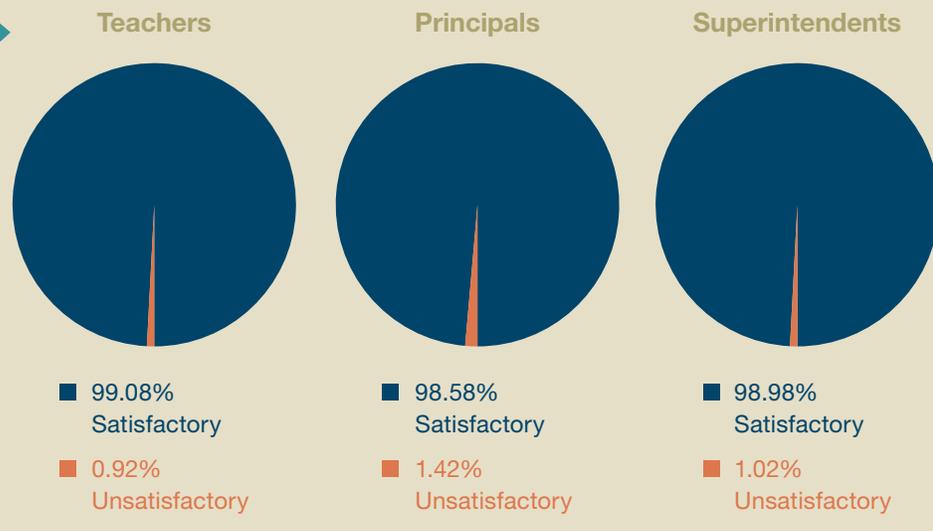
None of this, however, can happen without honest evaluations. The desire for more and better information about evaluation systems—what criteria were used, what the categories represented, how teachers were distributed along the performance scale—led the U.S. Department of Education to ask every state to make that information public. In exchange for federal money, in the form of allocations from the \$53.6 billion State Fiscal Stabilization Fund, every state agreed to report a number of factors for each district’s (or school’s) system. They include data on student achievement or growth and the number and percentage of teachers in each performance level.”⁸ The states were also asked to report on how districts evaluated principals, and to make all this information public for each school. Washington was the first state to release its data.

Washington’s share of the stabilization fund money totaled \$1 billion, and it has produced data that goes above and beyond the federal requirements. First, the state includes classes of education workers aside from teachers, such as superintendents and janitors. It also requires districts to demonstrate their capacity (or lack of it) to track not just a teacher’s overall rating but also how that individual

fared on each criterion (instructional skill, for instance, or classroom management) that added up to the total. It requires districts to report how often teachers and principals are formally observed, and to note who conducts those observations.⁹ And, as distinct from the original TNTP report, the Washington data gives us a picture of an entire state, not just a handful of districts.

The data confirm, not surprisingly, that Washington's educators have been treated like the proverbial widgets. Specifically, three out of every four districts in the state have relied on simple either/or evaluation systems, and, as **Chart 1** shows, very few teachers, principals, or superintendents received unsatisfactory ratings.¹⁰ Statewide in 2010–11, districts deemed the performance of a minuscule 0.92 percent of teachers unsatisfactory out of 54,781 public school teachers overall. Out of 2,470 principals, just 1.42 percent were rated unsatisfactory, and out of 293 superintendents, only three received unsatisfactory ratings. The numbers are similar for other school support staff, where 97.9 percent earned satisfactory ratings.

CHART 1
Very few
Washington
teachers,
principals,
and
superintendents
were rated
unsatisfactory in
2010–2011.



Washington's statewide averages tell an interesting and broad story, but the data also opens a window on individual schools. **Chart 2A** shows teacher evaluation ratings spread out over the 2,251 schools in the data set, and it reveals that teacher evaluations in those low-performing Pasco schools were far from unique: 1,905 Washington public schools did not identify a single teacher as unsatisfactory.¹¹

Even the schools that identify unsatisfactory teachers do so in small numbers, the data shows. There were 256 schools that each identified only one unsatisfactory teacher. Another 58 identified two, and only 32 schools gave three or more teachers unsatisfactory ratings. In other words, while the percentage of schools with an unsatisfactory teacher is higher than the percentage of unsatisfactory teachers

statewide, most schools are identifying no teachers, or only a tiny percentage of teachers, unsatisfactory.

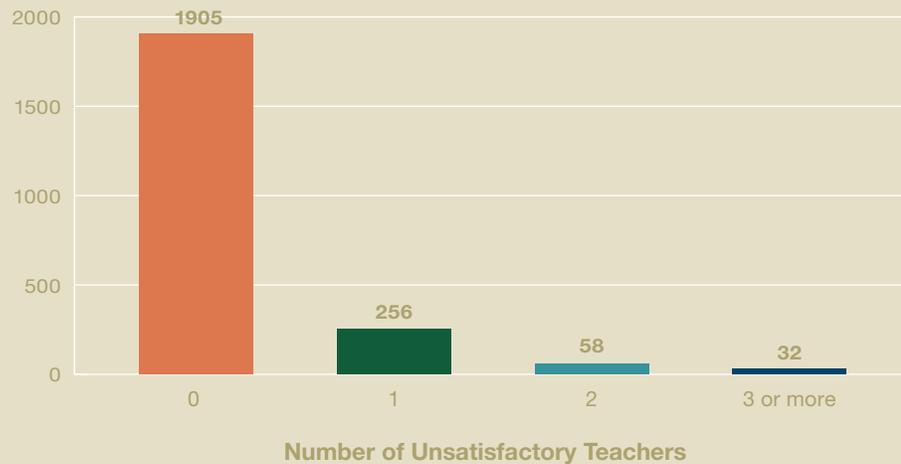
The same story applies in **Chart 2B**, which shows the districts identifying unsatisfactory principals, by the number of principals identified as such. Of the 261 districts in the dataset, the average employed 9.5 principals, but those numbers too are skewed because some districts are large. Seattle and Spokane, for example, reported more than 90 principals each, while many small and rural districts employed only a few. The vast majority of districts, 239, did not identify a single unsatisfactory principal. Eighteen districts identified one unsatisfactory principal, one district identified two, and only three districts identified three or more unsatisfactory principals.

CHART 2A

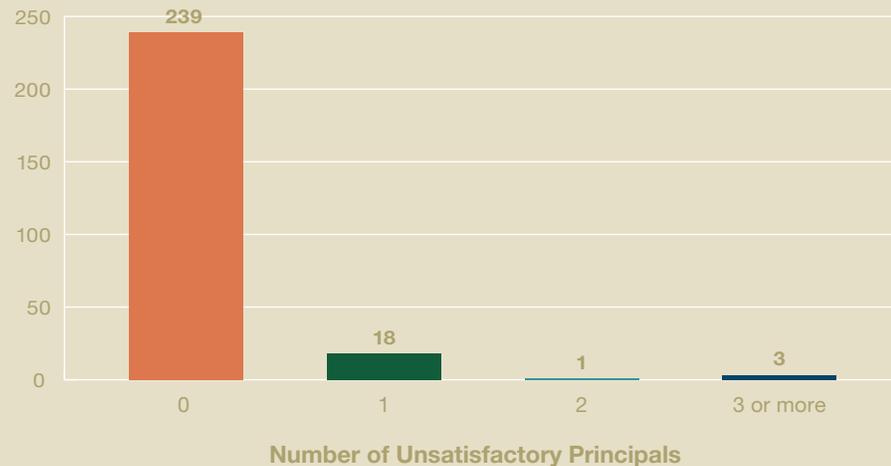
Most Washington schools didn't identify a single unsatisfactory teacher or principal in 2010–11.

CHART 2B

Number of Schools



Number of Districts



Many More Labels Describe Good Work Than Bad

In addition to providing the number of teachers rated in each category, districts with multi-category evaluation systems also provided the labels each district used for their performance levels and the number of teachers falling into each. These lists offer a revealing glimpse into how districts evaluate teachers and label their efforts, and they suggest that districts have a tough time even talking about unsatisfactory or ineffective performance.

To illustrate this point, **Chart 3** shows the labels that districts use, sized according to the frequency with which they appear in the districts' submissions. The author took some liberties in combining some closely related terms¹², but the results show that districts were much more willing to elaborate on and create new categories for satisfactory performance than they were for unsatisfactory performance.

The top of Chart 3 shows the 60 words that commonly appeared in district labels of satisfactory performance. "Satisfactory" is by far the most common because many districts used only an either/or determination. The list also includes superlatives like "strong," "excellent," "outstanding," "exceptional," "distinguished," and "commendable." Notably, districts gave overall satisfactory ratings to teachers with seemingly poor results. Districts gave positive summative evaluations to teachers described with words such as "unsatisfactory," "below," "fair," "minimally," and "basic."

The first thing to note from the bottom of Chart 3, which shows district labels for unsatisfactory ratings, is that there are simply fewer words than in the top portion: About half as many categories describe poor performance as describe good performance. Many of the negative words, such as "weak" or "fails" or "unacceptable," send a strong, clear message. But there is also quite a bit of overlap between positive and negative terms. Words like "emerging," "developing," and "adequate" convey progress or sufficiency, belying the fact that they are applied only to the bottom 1 percent of educators with unsatisfactory ratings.

CHART 3

Districts are much more willing to create categories for satisfactory performance than for unsatisfactory performance.



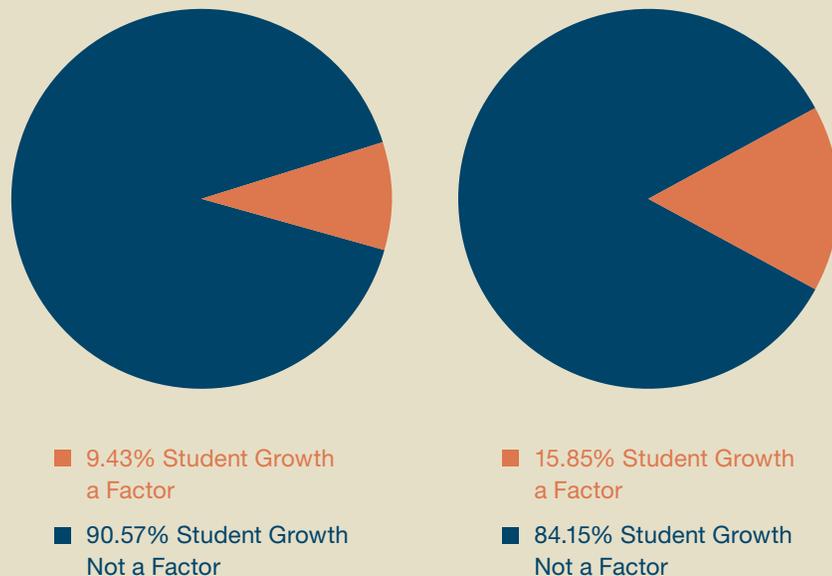
Few Held Teachers Accountable for Student Learning

The requirements of the stabilization fund were designed to inform the public how many districts evaluated teachers and principals based on how much their students learned. Four states had laws on the books that banned schools and districts from linking teachers with student performance data, but in most states teacher quality advocates were concerned that educator evaluations were too focused on a teacher's practice and disconnected from student outcomes.¹³ The TNTP study revealed a clear disconnect in places like Denver, where even in schools that failed to meet the state's academic benchmarks, more than 98 percent of tenured teachers received the highest evaluation rating.¹⁴

Washington districts, including Pasco, demonstrated the same disconnect, and they did not generally include objective information about how much students learned over the course of a year as a factor in teacher evaluations (**Chart 4**). Less than 10 percent of districts held teachers accountable for how much their students were growing academically, and only about one out of six districts were holding principals accountable for how much their students progressed.

CHART 4

Student growth was not a factor in most Washington teacher or principal evaluations in 2010–11.



Over the past few years, states have made big strides in upgrading their evaluation systems. According to the National Council on Teacher Quality, two-thirds of the states have adopted new ways to evaluate teachers. States also are increasingly using the evaluation results to make consequential decisions. In 2009 no state required districts to consider student learning when deciding on teacher tenure, but by 2012 sixteen states required districts to do so.¹⁵

The legislation that Washington passed in 2010 and 2012 requires districts to use four-level rating systems in evaluations and to factor in student growth. Beginning with the 2013–14 school year, teachers will be placed in one of three growth bands—high, average, or low—and those whose students show slow growth are automatically disqualified from receiving the highest rating. Tenured teachers who rank in the second-lowest of the four tiers for two consecutive years or twice in three years also must go on probation and develop a formal improvement plan. Failure to improve is then probable cause for dismissal.¹⁶

Although the legislative and regulatory changes aren't yet fully in place, 17 Washington districts have been participating in a pilot project developing more rigorous observation rubrics and rating teachers on a four-tier system. Of 14 participating districts with available data, 11 had been using the simple model of satisfactory-or-not. Now, all are using a four-tier continuum: unsatisfactory, basic, proficient, and distinguished.¹⁷

Such steps undoubtedly improve upon most existing systems, but districts still have a substantial amount of work to do if they are going to implement the new law with fidelity. Even in the pilot, no teachers were rated unsatisfactory, and only 2.8 percent were placed in the second-lowest level. Teacher participation in the pilot was voluntary, and districts didn't allow teachers with previously unsatisfactory ratings to participate, so the results may not be fully representative, but it's worth noting that the disparity between student performance and teacher practice may not be resolved simply by evaluating teachers in new ways.

The pilot evaluations do not yet consider student growth, but states that do demonstrate that even these objective measures cannot override a culture that doesn't value frank discussions of performance. Florida, for instance, requires that 50 percent of a teacher's rating come from his or her contributions to students' growth. Yet even with the addition of this data, 96.5 percent of Florida teachers were rated highly effective or effective in 2011–12.¹⁸ And in Tennessee, in the first year of a new evaluation system, 16.5 percent of teachers earned the lowest rating in student growth, but only 0.2 percent earned the lowest score on classroom observations.¹⁹

Only Three Districts Based Teacher Pay on Evaluations

Employee evaluations, by themselves, will not improve the quality of the workforce, but they do provide valuable information for both employee and employer. With solid information in hand, employers can make sound decisions about professional development, retention, and compensation. Without it, they are forced to make decisions blind, and to standardize policies for every employee no matter their unique situations. They use off-the-shelf professional development, for instance,

instead of tailoring it to individual principals and teachers or pairing up successful teachers with struggling ones. Research suggests that the traditional professional development model has failed to produce gains in student performance, but practices like informal mentoring and coaching, embedded in a teacher's job, have shown promising results.²⁰ These effective practices all start with effective evaluations.

Districts have also created compensation and retention systems that ignore information about performance. To be transparent and predictable, they have adopted single salary schedules that pay two teachers with the same level of experience and education the same, ignoring real differences in skill. Likewise, instead of taking proactive steps to influence whether a teacher stays or leaves, districts often leave staffing decisions up to the teachers themselves; in most districts principals have little control over who works in their school, and a recent survey found that districts and school principals did not make strategic efforts to retain their best employees.²¹

When Washington's Office of the Superintendent of Public Instruction surveyed districts on whether they used evaluations to make personnel decisions, the results varied considerably depending on the action (See **Table 1**). For example, nearly every district said it used evaluations to shape improvements in instruction and leadership. But only three districts out of 265 said that they base teacher compensation on evaluations, and only 17 said that performance was a factor in compensating principals. Slightly more districts indicated that they used evaluation results in decisions on teacher and principal promotion, but the numbers may be artificially low because most districts don't have many natural opportunities for promotion. In most places, teachers seeking promotion need to become principals, and principals must become district leaders.

A large percentage of Washington districts indicated that they used evaluations to decide whether to renew a teacher's contract (mainly teachers without tenure) and to determine probationary periods. But given the uniformity of satisfactory ratings, it's difficult to conceive of how districts would be able to make meaningful distinctions for these high-stakes purposes. A recent Bellwether Education Partners analysis of state evaluation laws praised the Washington legislation for linking tenure to effectiveness and for making poor evaluations a trigger for dismissal, but it also noted that the laws do not prohibit "last-in, first-out" layoffs that put seniority over quality.²² That flaw played out during the recent recession. In 2011–12, Washington districts employed 5.2 percent fewer teachers than they did three years ago.²³ Given the districts' failure to differentiate among teachers, it's unlikely they could have considered performance in those layoff decisions even if the law required it.

Failure to differentiate high- and low-performers also hurts students. According to research from the Washington-based Center for Education Data and Research,

basing lay-off decisions on seniority alone would lead to about 10 percent more teachers receiving pink slips. Because less-experienced teachers earn lower salaries, a district would have to lay off more of them to attain the same budgetary savings as an across-the-board cut. Additionally, the researchers found that quality-blind layoffs would result in a less-skilled workforce, leading to the equivalent of replacing an average teacher with a poor one.²⁴

TABLE 1

265 districts vary greatly in how they use evaluations for personnel decisions.

Personnel Decision	Number of Districts Using Evaluation Results to Make Decisions for Teachers:	Number of Districts Using Evaluation Results to Make Decisions for Principals:
Professional Development	262	247
Instructional/Leadership Improvement	254	231
Compensation	3	17
Promotion	22	39
Probable Cause for Non-Renewal of Contract ²⁵	244	214
Establishment of Probationary Period	227	163

Using the Data for Meaningful Evaluation

In the last few years, Washington has meaningfully addressed its failure to acknowledge, determine, and act on differences in educator effectiveness. The state's actions are particularly significant given its political culture. According to a recent analysis from the Thomas B. Fordham Institute, a think tank based in the "other" Washington, Washington state is home to one of the country's strongest teachers unions.²⁶ It is worth noting that the union, the Washington Education Association, is pushing for reform as part of a broad coalition that includes state officials, a bipartisan group of legislators, trade groups like the Association of Washington School Principals, and reform advocates such as Stand for Children and the League of Education Voters.²⁷

We're likely to get more data like Washington's as more states comply with the stabilization fund requirements. States granted waivers from NCLB, including Washington, are required to create and monitor comprehensive teacher and principal evaluation systems. These states should invite the public to measure how closely teacher evaluation ratings are related to school success.

As we follow this new data, it is important to remember that even under these new systems, decisions about what percentage of teachers should receive which ratings are ultimately judgment calls. Legislators in Florida have responded to the preliminary results there with calls that teacher evaluation ratings should match student proficiency rates. But, while this disconnect is problematic, states should resist the temptation to set quotas for unsatisfactory evaluations. As Andrew J. Rotherham, Sara Mead, and Rachael Brown warned in a recent report, “States should not mistake processes and systems as substitutes for cultural change.”²⁸ Instead, states should create meaningful evaluation systems that recognize and reward high-performers and identify and support low-performers. If schools and districts aren’t accomplishing that goal—particularly in schools where student achievement is persistently low—it strongly suggests that the adults in those schools are avoiding some difficult conversations.

Notes

1. Those five schools were Emerson, Longfellow, Robert Frost, Rowena Chess, and Virgie Robinson Elementary Schools. The full state list is available at: <http://www.k12.wa.us/StudentAndSchoolSuccess/SIG/default.aspx>.
2. Across the five schools, an average of 47 percent of teachers missed at least ten days of school. Source: *U.S. Department of Education, National Center for Education Statistics, 2009–10*. The national figure is 36 percent. Source: <http://www.americanprogress.org/issues/education/report/2012/11/05/40371/teacher-absence-as-a-leading-indicator-of-student-achievement/>.
3. Pasco applied for School Improvement Grant funding in 2010 but its application was not chosen by the state, and Superintendent Hill declined to re-apply in subsequent years. See, for example: <http://www.tri-cityherald.com/2011/01/24/1339334/pasco-kennewick-schools-wont-see.html>.
4. District-wide, all 19 principals and 804 out of 807 teachers earned satisfactory evaluation ratings. Even for other types of employees like paraeducators, secretaries, custodians, and those involved in campus security, technology, transportation, maintenance, grounds, and food service, only 2.4 percent earned unsatisfactory ratings. In addition to receiving her own formal satisfactory rating, Superintendent Hill was also named the 2013 Superintendent of the Year by the Washington Association of School Administrators: http://www.wasa-oly.org/AM/Template.cfm?Section=Press_Releases&Template=/CM/ContentDisplay.cfm&ContentFileID=3718.
5. The data are available for download at: <http://tpep-wa.org/resources/tpep-reports-studies-presentations-surveys/>.
6. <http://widgeteffect.org/downloads/TheWidgetEffect.pdf>.

7. See, for example: http://www.edtrust.org/sites/edtrust.org/files/publications/files/Strengthening%20Michigan's%20Teaching%20Force_0.pdf.
8. Washington's approved application under the State Fiscal Stabilization Fund is available at: <http://www2.ed.gov/programs/statestabilization/amended-applications/wa.pdf>.
9. From the state's perspective, this information did not cost a substantial amount of money. In a 2010 plan they provided to the U.S. Department of Education, Washington's Office of the Superintendent of Public Instruction estimated the development, execution, technology, supplies, and materials needed to conduct the survey amounted to only \$9,000. That does not include the cost of the district survey respondents' time. See Washington's plan at: <http://www2.ed.gov/programs/statestabilization/stateapps/phase-ii/wa-phase-2-approved.pdf>.
10. About one-quarter of the districts did report evaluation systems that used multiple performance levels. Those districts also identified which performance levels equated to an either/or satisfactory/ unsatisfactory categorization. The numbers cited in this report rely on these district-reported conversions.
11. This may be partially driven by the fact that before the recent legislation "unsatisfactory" was also a legal term. If a teacher was judged "unsatisfactory" it automatically triggered probationary procedures. The evaluation bills changed the terminology to "not judged satisfactory," which provides more flexibility to identify employee weaknesses without starting formal probationary procedures.
12. For example, "Sat" became "Satisfactory" and "Developing/ does not meet" and "Does not meet criterion" were combined into "Does Not Meet Expectations."
13. Those four were California, Nevada, New York, and Wisconsin. Source: http://tntp.org/assets/documents/TNTP_InterpretingR2T_2009.pdf?files/TNTP_InterpretingR2T_2009.pdf.
14. Ibid, *The Widget Effect*.
15. http://www.nctq.org/p/publications/docs/Updated_NCTQ_State%20of%20the%20States%202012_Teacher%20Effectiveness%20Policies.pdf.
16. For more on Washington's teacher evaluation laws and a comparison on how it stacks up with other states, see: <http://bellwethereducation.org/wp-content/uploads/2012/08/RSA-Teacher-Effectiveness.pdf>.
17. See http://tpep-wa.org/wp-content/uploads/Washington_TPEP_Cross_Case_Analysis.pdf and http://tpep-wa.org/wp-content/uploads/TPEP_Crosscase_Analysis_Summative_Ratings.pdf.
18. <http://www.voxxi.com/florida-vam-teacher-evaluation-system/#ixzz2lnAu5crZ>.
19. http://www.tn.gov/education/doc/yr_1_tchr_eval_rpt.pdf.
20. Richard Ingersoll and Michael Strong, "The Impacts of Induction and Mentoring

Programs for Beginning Teachers: A Critical Review of the Research.” *Review of Education Research* 81:2, 201–233.

21. <http://tntp.org/irreplaceables>.
22. <http://bellwethereducation.org/wp-content/uploads/2012/08/RSA-Teacher-Effectiveness.pdf>.
23. See <http://www.k12.wa.us/DataAdmin/pubdocs/personnel/positionandethnicity0809.pdf> and http://www.k12.wa.us/DataAdmin/pubdocs/personnel/2011_2012PersonnelByMajorPositionandRaceEthnicity.pdf.
24. [http://www.cedr.us/papers/working/CEDR%20WP%202011-1.2%20Teacher%20Layoffs%20\(6-15-2011\).pdf](http://www.cedr.us/papers/working/CEDR%20WP%202011-1.2%20Teacher%20Layoffs%20(6-15-2011).pdf).
25. According to state law, this should be every district. An evaluation that results in being judged not satisfactory must trigger probationary procedures, and not improving while on probation is probable cause for non-renewal: <http://apps.leg.wa.gov/RCW/default.aspx?cite=28A.405.100>.
26. The analysis is based on financial resources, size of membership, and number of policies determined by local collective bargaining. Available at: <http://www.edexcellencemedia.net/publications/2012/20121029-How-Strong-Are-US-Teacher-Unions/20121029-Union-Strength-State-Profiles.pdf>.
27. The 2012 legislation, for example, was championed by Democratic Rep. Eric Pettigrew and Republican Sen. Steve Litzow.
28. <http://bellwethereducation.org/the-hangover-thinking-about-the-unintended-consequences-of-the-nations-teacher-evaluation-binge/>.

ABOUT THE AUTHOR

Chad Aldeman is a Senior Policy Analyst at Bellwether Education Partners, a nonprofit organization working to improve educational outcomes for low-income students. He can be reached at chad.aldeman@bellwethereducation.org.

ACKNOWLEDGMENTS

Many thanks to those who offered invaluable feedback on drafts of this report, including Dave Powell, the Government Affairs Director at Washington Stand for Children, and Andy Rotherham, a co-founder and partner at Bellwether Education. Thanks also to the entire Education Sector team but especially Research Director Andrew Gillen and Managing Editor Susan Headden for their detailed engagement with the paper and their time spent improving it. It is undoubtedly better because of their contributions. Any errors or omissions are my own.

ABOUT EDUCATION SECTOR

Education Sector is an independent think tank that challenges conventional thinking in education policy. We are a nonprofit, nonpartisan organization committed to achieving measurable impact in education, both by improving existing reform initiatives and by developing new, innovative solutions to our nation’s most pressing education problems.