A Teacher Evaluation System Designed to Improve Student Achievement

Research and Classroom Experience-Based Recommendations by Teachers

By Teachers United’s Teacher Evaluation Policy Team, 2012

BACKGROUND + RESEARCH
How the issue was selected and how the research was conducted.
PAGE 3

RECOMMENDATIONS
How real classroom teachers think they should be evaluated in order to increase student achievement.
PAGE 5

CONCLUSION
A short summary of the policy team’s recommendations.
PAGE 12
The mission of Teachers United is to grow a movement of teachers who build an education system worthy of our students' potential.

We are working toward the day when 100% of our students in Washington State graduate high school ready for college or their career.

Educators who took part on the Teacher Evaluation Policy Team:

Deanna Barrett
Emily Elasky
Ritchie Garcia
Katie Kribbs
Elizabeth Ogden
Suzanne Righi
Katie Virga

Katie Cryan Leary
Pam Faulkner
Jen Hamilton
Michelle Lewis
Marquita Prinzing
Jenna Velozo
Jeremy Whiting

A special thanks to our Policy Director, Erin Gustafson, for assimilating this immense amount of knowledge and experience; and our former Field Organizer, Lindsey Sailors, for working to assemble this amazing group of educators.

Our Board of Directors
Kirby Green, Board President, Teacher, Hawthorne Elementary, Seattle
Kristin Bailey-Fogarty, Board Vice-President, Teacher, Eckstein Middle, Seattle
Gena Poling, Board Secretary, Teacher, Chinook Middle, Highline
Nathan Gibbs-Bowling, Board Treasurer, Teacher, Lincoln High, Tacoma
Todd Hausman, Teacher, Lowell Elementary, Bellingham
Christopher Eide, Executive Director, Substitute Teacher, Seattle

Produced by Teachers United, 2012. All rights reserved.
210 South Hudson Street, Suite 326, Seattle, Washington 98134
At a network meeting in May, teachers affiliated with Teachers United discussed and voted on three policy issues to study in the upcoming year. One of these issues was how teachers should be evaluated and how staffing decisions such as layoffs and placements should be made. Any interested teacher was invited to join the “Evaluations Policy Team” and fourteen teachers did. These teachers were particularly interested in making the evaluation process both fair and an excellent professional development tool. Above all, they were concerned that it would increase student achievement.

Beginning in the spring, this group of teachers read evaluation case studies from ten districts across the country, nine research papers on the elements of evaluations, and nine articles and opinion pieces on evaluations and how to use them. They attended a presentation by district and union leaders from Anacortes Public Schools on the new state evaluation system being piloted in their district. They also attended a meeting at The Bill and Melinda Gates Foundation to discuss successful evaluation components and systems with national experts studying systems across the country.
The teachers on the policy team initially shared what they felt was and wasn’t working about their own evaluation system. Teachers agreed the new four-tiered rating system is a dramatic improvement over the previous “satisfactory/unsatisfactory” evaluation system, and that the instructional frameworks adopted by the state were sound. They felt that there is great potential for a teacher’s growth with the frameworks, but that challenges and weaknesses are being exposed in the implementation of the evaluation system. This conversation about areas where improvements could be made led to the team breaking down the evaluation process into components. A curriculum was then formed around this teacher-created list, and reading and interviews began. (For the list of readings and interviews conducted by the policy team, please visit www.teachersunitedwa.org).

As the research progressed over seven months, seven major areas to address emerged:

1) Observer reliability
   *The evaluation rubric may be great, but it is only as great as the evaluator.*

2) Frequency of observations
   *In order to improve, teachers need observations and feedback.*

3) Peer observers
   *Who should be able to observe, give feedback, evaluate?*

4) Individualized professional development
   *If teachers are asked to improve their craft, how should it happen?*

5) Student perception surveys
   *Students are in schools every day; should they evaluate teachers?*

6) Student growth measures
   *Students must progress in their academics, but how do we measure that?*

7) Staffing decisions
   *How should we use teacher evaluation data?*
1. Observer Reliability

One important component of a successful evaluation system is trust in the system itself. Teachers agreed that the system must be as fair and accurate as possible. Concerns were expressed about whether all principals were either qualified or had ample time to conduct observations and provide relevant feedback so that teachers may grow and improve their practice. There was widespread agreement that this was not happening consistently. This has lead to inequities in the system and thus, an overall lack of trust and lost opportunities for growth.

In a research report outlining case studies of teacher evaluation in ten districts across the country, requiring some sort of certification process for evaluators was a common theme. At many of the sites this process included extensive training on the evaluation system and how to provide accurate feedback as well as some sort of assessment process to ensure observers had reached a minimum level of rating reliability.¹ The Measures of Effective Teaching (MET) study has concluded that this process of building observer capacity to conduct accurate observations and provide relevant feedback is essential to any successful evaluation system.² Locally, the Anacortes School District (one of Washington State’s Teacher and Principal Evaluation Pilot [TPEP] sites) requires observers to receive a certain minimum score on rating reliability in order to conduct observations.

Given what they learned from this research and in talking to practitioners, the policy team concluded that they do need to know their observers have been adequately trained and have proven themselves to be accurate and reliable in their observation ratings.

Recommendations:

1) Require principals and peer observers to receive evaluation certification to ensure they meet a minimum level of rating reliability.

2) Principals and peer observers will do formal observations only when they are certified.

3) If principals and peer observers are not certified, only a certified evaluator will conduct formal observations.


2. Frequency of Observations

More observations are better. While our research showed a wide variety in the number of observations teachers received in different systems, the Measures of Effective Teaching study concluded that evaluation reliability increases with more observations. This certainly resonated with the policy team. With fewer observations, more is riding on each one. Being able to average scores over more observations ensures that an atypical observation does not unduly affect a teacher’s overall rating. More observations also mean more opportunity to observe a wide variety of lessons and practices. This results in more data and thus, more opportunities for feedback to improve practice.

Feedback is perhaps the most important element. In addition, the team agreed that although the process of planning for a formal observation was a valuable one – including the pre-observation conference and goal setting – evidence collected in informal “drop-in” observations would also lead to valuable and potentially different feedback that ought to be included in the evaluation process. The goal should be to increase data and feedback for teachers and a combination of formal and informal observations achieves this.

Recommendations:

1) Each teacher shall receive a minimum of two formal observations with pre- and post observation conferences with the observer.

2) In addition to formal observations, each teacher shall receive a minimum of four informal “drop-in” observations. Immediate feedback shall be provided following these informal observations in writing.

3) Evidence collected during informal observations relating to a teacher’s growth goals shall also factor into the teacher’s summative evaluation rating.

4) From a capacity standpoint, we recommend each certified evaluator have a caseload of no more than 15 teachers.

Kane, T. “Capturing the Dimensions of Effective Teaching,” EducationNext, Vol. 12, No. 4 (Fall 2012)
3. Peer Observers

Reliability is incredibly important. The policy team’s research showed that having multiple observations from multiple observers greatly increases the reliability of observations. With only one observer who may bring either overly positive or negative bias into an observation setting, reliability is potentially compromised. In addition, many of our teachers felt that given their other responsibilities, principals often had limited time to dedicate fully to a robust observation/feedback process for each teacher in their buildings. Principals also may not always have the capacity, grade level expertise or content area knowledge necessary to provide the best feedback for teachers.

Solving capacity issues and providing career ladder options.
Including peer observers in the evaluation system is a way of solving these capacity problems, increasing the reliability of observations, and ensuring that teachers receive the valuable feedback they need to adjust their practice to improve student outcomes.

Recommendations:

1) Peer observers must receive certification to perform formal evaluations.

2) A peer observer is an educator from outside the building, preferably with grade-level or content area knowledge matched with that of the teacher being observed.

3) Peer observers should have recent and successful classroom experience and a willingness to be in this full-time position for 2-3 years.

4) Formal and informal observations may be divided up among principals and peer observers depending on building and teachers’ needs.

5) There is a pool of potential peer evaluators that might be comprised of career ladder teachers, master teachers or coaches.

6 While introducing extra observers into the process certainly costs money, there may be opportunities to repurpose career ladder positions and resources to do this work.
4. Targeted Professional Development

One of the best ways to improve practice is through targeted, individualized professional development that is connected to areas where teachers need to improve. When teachers experience well-designed professional development, these skills can then be applied to improving the content of instruction and instructional practice. This leads to increased student learning, which is the goal.

Unfortunately, professional development has often a required one-size-fits-all approach that does not address teacher needs, wastes time and does not lead to improved practice. This is a loss in terms of money for the district; but more importantly a lost opportunity for teachers to hone their craft. The policy team described many examples of professional development that did not relate at all to what they needed to improve their practice.

In order to accurately assess the usefulness of district professional development, teachers should be surveyed and professional development offerings should be adjusted or eliminated based on the data collected. If most teachers are strong in classroom management for example, there should be fewer professional development offerings related to that. Responding to areas of actual need will create efficiency and financial savings that can then be potentially invested into more career ladder positions and more relevant professional development.

There was agreement that career ladder positions should be expanded to every school, as they are a way to begin providing that job-embedded professional development to teachers. The coaching that career ladder teachers can give is essential for teachers to incorporate professional development content into their actual practice.

Recommendations:

1) Connect professional development to the adopted instructional framework.
2) Every school should have a career ladder teacher available to provide coaching and resources connected to areas needing improvement on a teacher’s evaluation observation or student survey results.
3) Professional development funds should be directed to career ladder/coaching positions since they are - by definition of their work - providing job-embedded professional development to teachers.
4) Require district to collect data and track trends about areas where teachers need more or less professional development. This will allow the district to focus professional development on actual areas of need and not require professional development that does not address that. Use summative data from end-of-year evaluations to identify these trends.
5) Compile list of career ladder teachers who are willing to provide professional development district-wide that teachers can choose from. Survey teachers and collect data on the quality/usefulness of these professional development offerings. In this way, highly rated professional development will rise to the top.
5. Student Perception Surveys

Research shows that student perception surveys, when used in conjunction with student growth data and observations, can be a more accurate and reliable measure of how teachers are doing and will do in the future than any one measure on its own. More importantly, student perception surveys can help show exact areas where teachers need to improve. Unlike a test that shows what students did and didn’t learn, student perception surveys can show the underlying causes for why a student may or may not be learning. This is valuable data to help inform a teacher’s practice.

In addition, student perception surveys provide a “check and balance” on the other components of a teacher’s evaluation (student growth data and observation results). For teachers of non-tested subjects who currently do not receive a student growth score, student perception surveys are an important component to balance observation ratings.

Trust in the survey itself is important. Teachers need to know that the questions do not constitute a “popularity contest” and that they really do get at important teaching and classroom practices. After looking at a set of student perception questions from the TRIPOD survey, teachers felt more confident that the questions really were designed to help provide feedback to adjust and improve classroom instruction.

The policy team felt that one way to gain trust in the use of student perception surveys was to have a committee of teachers study surveys being used nationally for potential use in a district pilot. They felt this emphasis on a more teacher-led process of development and implementation over several years would lead to the successful incorporation of student perception data into the evaluation system in the long run.

Recommendations:

1) In the first year, a committee of teacher-leaders should help develop the system and implementation plan.
2) In the second year, teachers should pilot student perception surveys in the district on an opt-in basis.
3) Conduct focus groups of teachers who are piloting the survey to get feedback for those developing the system.
4) Once the system is developed, a 1-2 year district-wide pilot of system would be implemented. To gain trust in the system, student perception surveys would not count towards evaluation ratings in the district-wide pilot year(s).
5) Focus groups would continue to happen in this district-wide pilot year(s) to get feedback, refine, and gain trust in system.
6) At the end of the district-wide pilot year(s), negotiate if and how surveys would factor into evaluation ratings.

---

7 Kane, T. and Staigner, D. “Gathering Feedback for Teaching: Combining High-Quality Observations with Student Surveys and Achievement Gains,” The Bill and Melinda Gates Foundation (2012)
8 Kane, T. and Staigner, D. “Gathering Feedback for Teaching: Combining High-Quality Observations with Student Surveys and Achievement Gains,” The Bill and Melinda Gates Foundation (2012)
6. Student Growth Measures

Multiple measures of growth must be used. Washington state law (SB 5895) has dictated that student growth must be a “significant factor” in a teacher’s evaluation rating. However, teachers felt it should by no means be the majority factor and that they should have a say in what assessments are used to measure student growth. There was a fair amount of concern among the members of the policy team that the MAP (formative assessment used in many districts in Washington) isn’t well aligned with state standards, and that because the MSP (annual state test) is only given once a year, it does not provide teachers with rich enough data to meaningfully adjust their practice.

Although it was agreed that there ought to be at least one common assessment used to measure student growth, teachers felt that multiple measures are necessary to increase reliability and to provide more data points on student growth during the year.

Recommendations:

1) Student growth data should be a component of a teacher’s evaluation rating, but by no means the majority factor.

2) Multiple measures should be used to measure student growth to increase reliability and should include at least one common state or district-based assessment to ensure consistency.

3) Teachers should be able to choose additional measures of student growth, particularly those they feel align with the content they are required to teach. These can be interim assessments designed to collect data and guide practice.

4) Prioritize development of assessments for non-tested subjects as is being done in districts like Anacortes, WA and Hillsborough County, FL.

5) Committees of teachers need to develop useful common benchmark assessments that can be used district-wide.

---

7. Staffing Decisions

We must ensure that the right teachers are in the right schools. The policy team universally disliked the practice of forcibly placing teachers into schools. They felt this was bad for teachers, building leaders and school communities as a whole. When a teacher is placed in a building and doesn't fit the mission of the school, this creates difficulty for that teacher as well as the rest of the building staff.\(^{11}\) It was agreed that mutual consent between the teacher and the building leader must be the guiding principle on placements.

The policy team also felt that we should move to a system where performance is the top factor in staffing decisions like layoffs, placements and displacements. They agreed that laying off teachers based on seniority with no regard for performance was demoralizing and a potential disincentive to entering the profession in a time when we need more highly qualified teachers.

That said, defining what “performance” means is important and there must be transparency in how that is determined. Performance must be determined using multiple measures and the team felt that using the summative evaluation was a way to do that effectively and fairly.\(^{12}\) In a situation where teachers have the same evaluation score, using seniority as a tiebreaker is a way of honoring the experience that comes with more time in the profession.

Recommendations:

1) The raw score generated by the summative evaluation should be the top factor in staffing decisions like layoffs, placements and displacements.

2) Teachers must receive the information necessary to have a good understanding of how that raw score is calculated.\(^{13}\)

3) Seniority acts as a tiebreaker in situations where evaluation ratings are equal.

4) Practice of forced placements should end. Teachers and principals should have mutual consent about placement to ensure the right fit with a school.


\(^{13}\) See for deeper explanation: http://tpep-wa.org/wp-content/uploads/TPEP_8%2009_webinar_w_links%20%281%29%26w%20%26%20slides%20%26%20pdf%20page%29.pdf
CONCLUSION

The driving purpose of an evaluation system should be to provide teachers with the feedback and data they need to improve their practice and ultimately, increase student learning. Washington State has selected three frameworks, each providing an excellent foundation upon which to build this important work. It is a beginning. However, to be able to carry out this work with fidelity, changes need to be made in how the system is being implemented.

We must ensure that observers are capable and that teachers receive enough observations with useful feedback from multiple observers. We must make sure that student growth is measured in a fair and useable way, and that students provide their teachers with feedback on their classroom experiences that can enhance and balance out other components of the evaluation system. When all these pieces are in place and implemented well, we can be confident that the staffing decisions made based on this information will be fair and in the best interest of students, teachers and school communities.

Teachers are the ones to lead the way in this work. They know what they need in order to refine their practice and the profession as a whole. Improving the way they are evaluated is a huge step in that direction.