

**QUALITY TEACHER EVALUATION IN ALASKA:  
VOICES FROM THE FIELD**

**POLICY BRIEF 2**

**Prepared by**  
Martin Laster  
University of Alaska Southeast School of Education

**October 2013**



Center for Alaska Education Policy is part of the Institute of Social and Economic Research at UAA

### INTRODUCTION

This brief focuses on the results of research regarding teacher evaluation policy and practice among a pilot group of Alaska Superintendents. The results of this report are intended to guide policy makers on creating and supporting policies which enable school leaders to effectively evaluate and support classroom teachers, helping to elicit the best from teachers on behalf of Alaska students. As Darling-Hammond (1999) states, “Despite conventional wisdom that school inputs make little difference in student learning, a growing body of research suggests that schools can make a difference, and a substantial portion of that difference is attributable to teachers.” It is critical that we identify how to enable teachers to increase desired performance for every student.

### BACKGROUND

In the last several years, states have been revising their education statutes and regulations, primarily to meet the requirements for Race to the Top funding or for waivers for elements of No Child Left Behind (NCLB). Including Alaska, there are over twenty states with new evaluation systems that require student performance data be included as part of educator evaluations.

In his opening remarks on February 28, 2013, Rep. Todd Rokita (R-IN), Chairman of the U. S. House of Representatives Subcommittee on Early Childhood, Elementary, and Secondary Education, cited the shortfalls of current elementary and secondary education law, stating, “No Child Left Behind’s rigid ‘Highly Qualified Teacher’ provisions require educators to have a bachelor’s degree, hold a state certification or license, and be able to demonstrate knowledge of the subject matter they plan to teach. That all sounds great in theory, but in reality it meant schools were forced to value an educator’s credentials over his or her ability to effectively and successfully teach our children.” While elements of this law expanded the focus of the education community to previously underserved subpopulations, it failed in its focus on growing more capable teachers for individual students.

At the same hearing, the Measures of Effective Teaching (MET) study received a great deal of attention. Dr. Steven Cantrell, Co-Director of the project, stated that research confirms teacher evaluation systems incorporating multiple measures are a superior way to gauge teacher quality. “Preliminary MET findings demonstrated that three measures – student assessments, classroom observations, and student surveys – helped predict whether teachers would raise the performance of future groups of students,” he said. “Indeed, the combination of these measures does a far better job predicting which teachers will succeed in raising student performance than master’s degrees and years of teaching experience... Better evaluation and feedback systems are essential to improving teaching and learning.”

Alaska recently revised its educator evaluation system. Many of the changes in the new Alaska regulations are aligned with measuring growth with multiple indicators over time as advocated by the META Study cited above. This brief reports on research with Alaska school district superintendents about quality teacher evaluation in Alaska, specifically on the elements of an effective teacher evaluation system and the use of student performance data for evaluating teachers.

## **CURRENT RESEARCH ON EVALUATION SYSTEMS**

The MET Study identified three measures of teacher effectiveness that predicted student learning when used together: classroom observations, student perception surveys or student assessments of whether they leaned in the class and student achievement gains (Cantrell and Kane, 2013). Using all three measures to produce a weighted estimate provided more stable measures of teacher quality than using prior student achievement alone; and adding a second observer increased reliability of the estimated quality, even using the same number of observations. Hanover Research (2012) in their study of using Best Practices for Including Multiple Measures in Teacher Evaluation found similar results but emphasized the lack of use of perception instruments and the need for the commitment of time and resources. They also expand the list of instruments: “It is common ... to include student achievement (value-added) data, classroom observations, administrator reports, parent surveys, student reports, portfolios, examples of work and, sometimes, peer reviews.” Additionally, Leahy (2012) has focused attention on the need to train evaluators for reviewing teaching effectiveness. Burris and Welner (2011) identify four areas that need to be present for effective teacher evaluation systems: summative data, formative data, working conditions and incentives. While the research addressed all four of these areas, the brief only reports on summative and formative assessments.

In general, traditional evaluation systems tend to focus on classroom observations, but with fewer observations than research recommends. In addition, they are usually limited to one observer and infrequent use of perception data. Darling-Hammond (2010) notes, “Virtually everyone agrees that teacher evaluation in the United States needs an overhaul. Existing systems rarely help teachers improve or clearly distinguish those who are succeeding from those who are struggling. The tools that are used do not always represent the important features of good teaching.”

## **ALASKA SUPERINTENDENT INTERVIEWS**

### **Methodology**

I interviewed six experienced superintendents for this study just prior to the State Board of Education adopting new educator evaluation regulations in March 2013. These superintendents work in very diverse settings: two are from urban districts, two from rural districts and two from single site districts. They have all been administrators for ten years or more and each has distinguished him or herself as an instructional leader through their credibility with colleagues and the student performance in their districts.

I asked them two broad questions:

- 1) *What are the key elements of a quality teacher evaluation system?*
- 2) *In what ways does student performance factor into teacher evaluations in your district?*

A steering committee reviewed the interview protocol for face validity. The superintendents' responses were recorded, transcribed and analyzed for common themes as well as notable differences. While each superintendent had his or her own view, they had a number of commonalities that are shared below. The findings were reviewed by the superintendents to assess the accuracy and reliability of the reporting and analysis. I've omitted names, to protect the confidentiality of the respondents.

## **Findings: Key Elements of a Quality Teacher Evaluation System**

In our discussions, superintendents provided extensive ideas about what needs to be included in a quality teacher evaluation system. Seven themes emerged from the interviews:

- Frequency of Observations
- Clarity
- Professional and Content Standard Driven Observations
- Student Centered Continuous Improvement
- Ownership by Stakeholders
- Significant-Data Driven
- Professionalism

### *Frequency of Observations*

The superintendents defined this as multiple (more than the required two), real observations not a “dog and pony show.” They cited the need for consistency. The process should include clear communication so no surprises result since the process has been addressed at the start of the year. Superintendents also identified the benefit of multiple observers as a means to improve reliability.

### *Clarity*

The need for clarity was a constant theme across all interviews. It involves having clearly defined expectations and procedures, as well as roles, responsibilities and involvement for all stakeholders involved in the evaluation process. Superintendents also spoke to the need for clear, concise instruments with descriptors, protocols, and rubrics guiding the ratings. They also stated that the process must not be overwhelming. It must be practical, meaningful and well communicated. There is an absolute need for common vocabulary.

### *Student-centered Professional and Content Driven Standards*

Instruments need to be tied to both the content area and professional educator standards, so as to clearly measure teachers’ subject knowledge or content expertise and their ability to share content/subject expertise with every student.

### *Student-Centered Continuous Improvement*

Administrators’ work with teachers needs to be personalized and engaging. Teachers need to demonstrate the 3 R’s: Relevancy, Reflection and Refinement. For there to be professional growth, there needs to be a trusting process for improvement and no ‘gotchas.’ For example, while it is important to have informal, unannounced observations, the net result should not be one of a climate of finding fault. Observations should reveal skilled knowledge and application of instructional/assessment techniques. The process needs to be focused on improving instructional quality and student performance. Systematic lesson design should be evident. The professional conversation should be based on a trusting relationship and focused on what can we do better. The culminating outcomes should be aligned with preparing each student for success the following year. Both instruction and observation should be targeted at student performance and progress. Teaching and learning should be the center of the dialogue between the administrator and teacher.

### *Stakeholder Ownership*

The superintendents spoke about the importance of involving stakeholders in the design, development and implementation of the evaluation process. They did caution about the appropriate inclusion of stakeholders in the process and the ability to calibrate feedback for

accuracy and reliability. As one superintendent noted, “It should be *meaningful* versus *mean* feedback for growth to occur.”

### *“Significant-Data” Driven*

It is not enough to be data-driven; the superintendents indicated the data must be meaningful for improvement and must be defensible -statistically. It must involve multiple indicators over time, “not a snapshot but a movie.” Superintendents maintained that we must use the students’ stories to assess student success and “we must be able to flip the lens” or individualize education to obtain a clearer view of every student. They felt strongly that it must be fair to teachers and principals. They were concerned that Alaska’s new state regulations only include a partial year of data and so won’t be able to assess the full year of teaching and learning. Finally, they questioned whether reliable assessment tools would be available for alternative (non-English and math) content areas.

### *Professionalism*

Superintendents felt strongly that this was a key element of quality teacher evaluation and performance. However, to successfully achieve this, teachers need specifically targeted protocols. Superintendents specified items to be included in these protocols, including: participation in staff collaboration activities, promoting a positive learning environment, support for student activities in the community and community building. The superintendents identified these items as representative of teacher professionalism outside the classroom.

## **Findings: Use of Student Performance Data in Teacher Evaluations**

After researching teacher evaluation tools and building questions that addressed these strategies, the responses fell into five categories:

- General Observations
- Summative Data
- Formative Data
- Student Products
- Student Satisfaction Indicators

### *General Observations*

Superintendents reported that at present, student performance data is not used directly or quantifiably in teacher evaluations. However, they unanimously saw a value in its use. They saw it as positive for growth measurement, continuous improvement and teacher recognition. They wanted to insure that it did not displace the importance of “soft skill” observations such as attendance, promptness, work ethic and time on task. They also saw its value for informing discussions regarding classroom observations. However, to be useful the data must be clearly aligned with and meaningful to teacher improvement and student performance.

Trust is essential in this process, built on relationships and stakeholder ownership of the process. Superintendents emphasized the critical need for multiple measures including statewide, nationwide and local data. They were concerned that the process be systematic and that adequate technology and expertise for data collection and analysis be available to all districts. The data should be standards-based, rigorous, content-focused and address common instructional targets. The big picture should incorporate the long view and use a moving average. It should be longitudinal and include follow up studies of graduates and non-graduates. This emphasis on student learning data should recognize the impact of social/behavioral factors. For teacher evaluation to be successful, resources need to be available, especially for time and training in the evaluation system. Professional development needs to focus on inter-rater reliability, evaluation

techniques, core standards, assessment techniques, and continuing commitment to employees.

#### *Summative Data*

The superintendents wanted access to statewide and nationally-normed measures. They did not want this to be the sole assessment; they universally called for multiple measures, preferably including some that are culturally sensitive, given the diversity of Alaska. They expressed concern that the summative assessments used might not be timely because they are out of sync with the evaluation timeline. This would be unacceptable to the superintendents. They felt strongly that statewide design and reporting must address this challenge. They looked to the new Alaska standards to increase academic rigor and raise the performance bar for students and teachers. At the same time, they felt it was imperative to communicate the impact of increased rigor at the state level to all stakeholders. They support a growth model, using an apples to apples approach over time or the comparison of the same student rather than this year's fourth grade to last year's fourth grade.

They offered the following cautions: 1) Data representing small numbers of students may not be accurate; and 2) there must be a way to factor in the impact of adversity on test results, for instance a plane crash in a small village during testing week.

#### *Formative Data*

The superintendents saw promise in formative assessments. They called for these to be consistent over time, rigorous, reliable and multiple. They cited current approaches and data sources such as MAP, AIMSWEB, Discovery and Chapter Tests. Superintendents expressed concern regarding non-English and Math content areas while recognizing the use of chapter tests. They also cautioned against the use of instructionally designed formative assessments for personnel evaluation when the assessments were not designed with that purpose in mind.

#### *Student Products*

Superintendents spoke of the value of standards-based binders for each student but fewer than 25% used this assessment of student progress. Student led conferences were far more common. Superintendents saw project-based assessments as meaningful, real-life, personalized and engaging. However, they also saw them as ambiguous and fuzzy requiring considerable focus on inter-rater reliability. Superintendents also saw promise in student portfolios that look at graduate outcomes but cited that staff would require training for consistency and reliability.

#### *Student Satisfaction Indicators*

Superintendents saw student ratings of teachers as valuable but also problematic for teacher evaluation. It is essential that the protocol used be clear, reflective of the desired information and fair, not a popularity contest. As one superintendent noted, "It can be a valuable part of the testimony but not the verdict."

### **DISCUSSION**

As Burris and Kaplan (2011) state, "Evaluations can be powerful interventions. High quality, thoughtful evaluation carries the potential to improve schooling. Misguided evaluation approaches, however, have a corresponding potential to harm our schools. It is essential, therefore, that we understand how to design evaluation systems that have the greatest likelihood of improving rather than undermining school performance." As Alaska implements a new teacher evaluation system incorporating student outcomes data, the recommendations presented in this brief can contribute to development of a robust and successful system that helps guide improved practice.

Development of this new system is happening at the same time as three other major initiatives: new content standards, assessment and accountability systems. To accomplish these four initiatives simultaneously will be challenging, and policy makers can help by a) minimizing additional new initiatives; b) supporting actions that lessen staff turnover, and c) providing funding to support the districts' capacity to fully implement these initiatives. Additionally, the Alaska Department of Education needs to be supported by the State Board of Education and legislature so that it can assist districts. A key tool for districts will be the development of templates or examples of strategies to address the new educator evaluation regulations.

Finally, a concerted effort must be made with and by all stakeholders, as this is a change that will require the long view and dynamic, continuous improvement.

## REFERENCE LIST

- Burris, C. C., & Welner, K. G. (2011). "Conversations with Arne Duncan: Offering Advice on Educator Evaluations." *Phi Delta Kappan*, 93(2), pp. 38-41.
- Cantrell, Steve and Kane, Thomas J., (2013). *Ensuring Fair and Reliable Measures of Effective Teaching*, Seattle, WA. Bill and Melinda Gates Foundation.
- Darling-Hammond, L. (1999). *Teacher Quality and Student Achievement: A Review of State Policy Evidence Center for the Study of Teaching and Policy*, Seattle: WA. University Of Washington.
- Darling-Hammond, L. (2010). *Evaluating Teacher Effectiveness: How Teacher Performance Assessments Can Measure and Improve Teaching*, Washington, DC. Center For American Progress.
- Darling-Hammond, L. (2012). *Creating a comprehensive system for evaluating and supporting effective teaching*. Stanford, CA. Stanford Center for Opportunity Policy in Education.
- Hanover Research (2012). *Best Practices for Including Multiple Measures in Teacher Evaluations* Washington: D.C. Hanover Research.
- Laster, M., (1984) *Analysis of Differing District Type on Providing Gifted Services in Colorado*, Athens: Ohio. Ohio University Press.
- Laster, M., (2013). *Quality Teacher Evaluation in Alaska*, Superintendent's Spring Conference Juneau: Alaska.
- Leahy, C., (2012). *Teacher Evaluator Training: Ensuring Quality Classroom Observers* Denver: CO. Education Commission of the States.
- Chairman Rep. Rokita , (February 28, 2013) Chairman of The Subcommittee on Early Childhood, Elementary, and Secondary Education, Washington: D.C.
- Race To The Top (2011) *State Comparison* Washington, DC. Department of Education.