

Enhancing Teacher Evaluation: A Critical Lever for Improving Teaching and Learning

May 10–11, 2010

Doubletree Hotel Washington DC–Crystal City

Summary

Welcome and Keynote Address

Sabrina Laine, Ph.D., director of the National Comprehensive Center for Teacher Quality (TQ Center), launched the workshop by discussing the importance of building teacher evaluation systems that support and improve teacher effectiveness. She emphasized the need to include evaluation as part of a systemic approach to human capital management. Dr. Laine then reviewed the data from the 41 Race to the Top applications, highlighting the intensive efforts of states related to teacher evaluation. Specifically, 29 states reported recent changes to legislation or an intent to pass legislation, with most occurring in 2009–10.

Dr. Laine then introduced two teaching policy fellows from TeachPlus. **Christina Porter** is a National Board Certified Teacher of English and literacy coach to English language learners (ELLs) in Revere, Massachusetts, an urban fringe town near Boston. **Karen Smith** teaches third grade in Wayne Township, Indiana. She also works as a technology coordinator and was honored as a Teacher of the Year in 2010. Porter and Smith described the policy brief that Teach Plus teachers wrote, *Teachers' Views on Measuring Effectiveness: Principles of Implementation of State and District Reforms*. The brief presents the following five principles of which Porter and Smith provided examples from their varied experiences in their districts:

- **Improve student outcomes by improving teacher evaluation.** Improving teacher evaluation is a key lever to improving student achievement.
- **Link teacher effectiveness to student learning.** Student growth data are essential to teacher evaluation.
- **Ensure high-quality evaluators.** Evaluation will promote teacher efficacy only if evaluators are credible and competent.
- **Promote a culture of accountability.** School and district leaders must be held to the same high standards for promoting teacher growth as teachers are for promoting student growth.
- **Communicate, communicate, communicate.** Teachers need information on why their effectiveness matters to their students and how it will be measured.

Porter emphasized that teachers want to be evaluated. They want meaningful feedback on their teaching, as long as they are provided with clear expectations and guidelines and high-quality evaluation tools are used. Using a mix of formal and informal evaluations, differentiating evaluation for teachers at different stages in their career, and creating a culture of accountability

are also important. Smith emphasized that student success should be at the forefront of any evaluation system but cautioned that states have to use multiple assessments of student learning (not just the state test) and avoid systems that have hidden “gotchas” for teachers. Dr. Laine concluded the session by pointing out the critical need to include teacher voices in policy conversations.

Concurrent Session I: “Building a Quality Teacher Evaluation System”

Angela Minnici, Ph.D., associate director of the American Federation of Teachers Educational Issues department, moderated this session, which aimed to identify the key steps in the development of a teacher evaluation system at the state and district levels, with North Carolina and Chicago highlighted as examples.

Carolyn McKinney, executive director of the North Carolina Professional Teaching Standards Commission, described North Carolina’s implementation from 2008 to 2010 of a revamped teacher evaluation process. Steps that allowed for the successful development and rollout of the evaluation system included the following:

- Teachers, administrators, and schools of education had a voice in the conversation from the beginning.
- The evaluation system is aligned with the mission and goals of the North Carolina State Board of Education; the North Carolina Teacher Working Conditions Survey; 21st century skills; the standards for teachers, principals, and superintendents; professional development; mentoring; and program approval of schools of education.
- Stakeholders were trained in the process, which includes the self-assessment, pre-observation conference, observation, post-observation conference, summary evaluations, and professional development plans.

Sheri Frost Leo, director in the Office of Human Capital at Chicago Public Schools (CPS), and **Lauren Sartain**, a research analyst at the Consortium on Chicago School Research at the University of Chicago, then presented on Chicago’s plan to implement a new, more valid and reliable teacher evaluation system by 2012. Under the Performance Evaluation Reform Act of 2010, all districts in the state are required to use student growth as a measure of teacher effectiveness and use standards-based teacher observation practices. CPS is calling this new system the “Excellence in Teaching Project” and has developed a standards-based teacher observation rubric based on the Danielson *Framework for Teaching*. Also, under this act, CPS is required to enter into a 90-day window of collaboration with the teachers’ union; if no consensus is reached, CPS will implement the last best offer.

Lisa S. Johnson, Ed.D., senior research and policy associate at Learning Point Associates and the TQ Center, introduced participants to new TQ Center resources: the *Guide to Teacher Evaluation Products*, which highlights more than 75 teacher evaluation tools, and the *Critical Decisions Guide: Building Teacher Effectiveness Systems*, an interactive roadmap for states and districts to follow as they start down the road to reform.

Concurrent Session II: “Evaluating Teachers of At-Risk Populations”

This session was moderated by **Bill East, Ed.D.**, executive director for the National Association of State Directors of Special Education. The presenters included **Tony Bagshaw**, managing director for human capital at Battelle for Kids; **Mary Brownell, Ph.D.**, Irving and Rose Fien Endowed Professor of Education and director of the National Center to Inform Policy and Practice in Special Education Professional Development; and **Lynn Holdheide**, research associate at Vanderbilt University and the TQ Center.

The presenters described findings from survey research on state and district evaluation systems, the use of value-added models in teacher evaluation, and empirical and quasi-empirical research on value-added models and the effect of teachers on student learning. They drew attention to the challenges related to the evaluation of teachers of at-risk populations—specifically special educators and ELL specialists—that must be considered when creating and implementing an evaluation system intended for all teachers. States and districts have experienced these challenges, and several workshop participants noted that they are aware of teacher concerns about this issue. Although the special skills and training required of these teachers should be considered when developing an evaluation system, there are few available examples of how these teachers have been accommodated.

A key challenge noted is that special educators often operate in coteaching situations, which makes evaluation difficult. Moreover, when it comes to using value-added models to evaluate teachers, the following challenges arise:

- Student achievement may differ and may need to be assessed differently for students with disabilities and ELLs, and this factor affects their teachers’ evaluations.
- Students with disabilities spend the majority of their day in regular education classrooms, so it is difficult to determine which teacher is responsible for what portion of a student’s learning.
- Use of value-added models may have a negative effect on the teaching profession by making it a less attractive profession to potential teachers.

Presenters further noted that data transparency is important to teachers when it comes to value-added models, and teachers will not change their behavior if they do not believe that the data are accurate. For all of these reasons, presenters concluded that value-added models may not be the most “defensible” foundation for the evaluation of special educators and that more defensible frameworks would include the use of classroom observation data and student achievement data aggregated to the school level rather than the individual student level. In the cases of coteaching, a conversation between coteachers is required to apportion the credit, or blame, for student learning. Involving special education and ELL administrators in conversations about reforming an evaluation framework or system also was recommended. The presenters noted the importance of considering growth in language acquisition, as well as academic content acquisition, in growth scores for English language specialists.

However, making any modifications to a district’s system for evaluating special educators or ELLs is hampered by contractual agreements that often prohibit such change.

Concurrent Session III: “Using Teacher Evaluation Results to Improve Teaching and Learning”

Laura Goe, Ph.D., research scientist at ETS and principal investigator for research and dissemination at the TQ Center moderated this session. Dr. Goe began by describing the following elements that affect how teacher effectiveness is measured:

- What is valued in the education system
- Technology available
- Types of evidence available
- Available rubrics
- Understanding of what a rigorous evaluation requires
- Resources (e.g., time, money, people, and policies)
- Cooperation of teachers
- Level of motivation for measuring teacher effectiveness

Dr. Goe described from a national perspective the spectrum of evaluation systems, from poor to very high quality.

Then, **Mary Gable, Jean Satterfield, and Colleen Seremet, Ed.D.**, from the Maryland Department of Education described the initiative in their state to improve teacher evaluation based on a systems approach. This approach ties in the state evaluation framework to the state teaching standards and systems of professional development, induction, and preparation. The challenge was that Maryland’s 24 districts historically have owned their own teacher evaluation systems, so in this context of local control, bringing in a state system was no simple feat.

To begin, staff from the Maryland Department of Education conducted 24 focus groups with teachers throughout the state, engaging in conversations about how to ensure that they are being fair and reasonable with their evaluation system as they move from a teacher input model to a teacher output model. Specifically, staff asked teachers whether the four elements of the Danielson *Framework for Teaching* (which was the most commonly used framework among the state’s districts) were reasonable, and teachers agreed that they were. Staff spoke with teachers about who should conduct evaluations, how often, and for which decision points (e.g., tenure, certification, professional development, compensation, and career advancement) the information should be used. Currently, an educator effectiveness workgroup is gearing up to make policy decisions based on these focus groups.

Gable, Satterfield, and Seremet shared the state’s quality control tools:

- Standards for professional development
- Planning guide
- Evaluation tools
- Spending analysis tool

Finally, turning to the local level, **Randy Keillor**, coordinator of the Student Performance Improvement Program (SPIP) in Independent School District 15 in St. Francis, Minnesota, described the development, elements, and outcomes of this program. Keillor emphasized the joint nature of the reforms between the union and the district from their beginnings, in 1995, to the present day and demonstrated how changes at the state level spark action at the district level.

Although teacher professional development reforms have been under way in St. Francis for 15 years, the current program was initially funded through Minnesota's Q Comp program in 2005 with the support of 85 percent of teachers in the district. Of the many interlinked components of SPIP, the performance review teams are of particular relevance. These teams meet in the fall to discuss what student growth will look like for a particular teacher. They then meet again in the spring to assess whether that growth was achieved. Keillor also detailed how these evaluations are tied to a system of professional development led by peer leaders, to the district's career ladder, and to teachers' compensation. In conclusion, Keillor noted, "A system that is indifferent to the performance of its employees and rewards them alike regardless of effort or effectiveness is based on an assumption that what those employees do really isn't very important or difficult."

Some participants asked why some teachers in the district did not support the program and why other districts in the state have not adopted this progressive model. Keillor responded that there were concerns about the state funds that supported districtwide salary increases disappearing, some general opposition to merit pay, and a dislike among some teachers for having other adults in their classrooms. Other districts have avoided adopting such a model largely due to poor relationships between union leadership and the governor, frustration with the Q Comp application process, and a lack of prior experience with this type of work. Other questions from the audience centered on whether SPIP was voluntary for teachers, who the peer leaders are, whether state tests are used as evidence of teacher effectiveness, and how salary bonuses are paid to teachers.

Dinner Panel: "Connecting Evaluation to Teacher Tenure and Dismissal: De-electrifying the Third Rail"

Tricia Coulter, Ph.D., deputy director of the TQ Center, moderated this session, which aimed to shed light on the important, and often contentious, third rail issues of teacher tenure and dismissal.

Rick Hess, Ph.D., director of education policy studies at the American Enterprise Institute for Public Policy Research, began by noting that tenure historically was needed because teachers were treated unfairly and did not have equal opportunity and other protections that now are in place. To create a tenure system that works today, Dr. Hess noted the following points:

- Tenure should not be automatic.
- Granting tenure should rely on evidence.
- Teacher effectiveness should be the main criteria for achieving tenure.
- A minimum number of years of service should be required before tenure is granted.

Given the changing labor market, Dr. Hess noted, that tenure systems must be developed to help strengthen the profession. He then compared and contrasted some recent initiatives to reform teacher tenure, including Florida Senate Bill 6, Colorado Bill 10-1-91, Louisiana House Bill 10-33, and the proposed teacher contract in Washington, D.C.

Lori Snyder-Lowe, superintendent of Morgan Local School District, described her experiences with teacher dismissal and developing a new evaluation system for teachers to achieve tenure. She contended that tenure should be earned and that the perception of tenure as a right needs to change. She noted that most teachers do want to be evaluated but that reviews must be based on multiple measures so that they are well rounded.

Doug Prouty, president of the Montgomery County Education Association, contended that the question about tenure is the wrong question to ask; eliminating tenure alone will not help teachers improve or lead to the removal of those who do not improve. Prouty described his district's Peer Assistance and Review program as one that allows for feedback, is not biased, and creates trust in the system. He noted that developing more meaningful, high-quality teacher assessments should be a priority over revamping teacher tenure. He also noted that there is a fear that people will be treated unfairly or arbitrarily by a principal and that a system of checks and balances must be in place to guarantee due process.

Closing Keynote Panel: “Engaging Stakeholders in Systemic Reform”

Jo Anderson, senior advisor to U.S. Secretary of Education Arne Duncan, U.S. Department of Education, moderated this session, which included the following three panelists from diverse stakeholder groups.

Dennis Van Roekel, president of the National Education Association, emphasized that the key teacher evaluation issue is that the process should be about improving practice. He contested the “false assumption” that if veteran teachers are not being dismissed, it means that ineffectiveness is not being addressed. Poor performers often leave voluntarily or are removed before they get to the due process stage. Due process was designed to protect effective teachers from bad practice, not to protect bad teachers.

Daniel Domenech, Ph.D., the executive director of the American Association of School Administrators, contrasted his experiences in Virginia and New York, where state law differs in terms of the due process practices. He suggested that many principals do not bother evaluating teachers because they know they have no power to remove teachers from the classroom even if they are found to be ineffective.

Susan Traiman, director of public policy at Business Roundtable, discussed how evaluation has become a central issue—in her opinion, due to anger sparked by The New Teacher Project's report titled *The Widget Effect: Our National Failure to Acknowledge and Act on Differences in Teacher Effectiveness* and *The New Yorker* article about New York City's “rubber room.” She noted the importance of considering the role of summative and formative evaluation. In the private sector, supervisors and employees constantly “check in” with one another, particularly Gen Y workers who desire more feedback and do not want to wait for a performance review.

Teacher evaluation tends to de-emphasize the need for ongoing feedback, which is helpful in cultivating and nurturing employees. In addition, Traiman noted that we must decide what to do about the teachers who are “just okay.” Specifically, we must reach a consensus that being “just okay” is not good enough.

The panelists also pointed to international comparisons on teacher evaluation and professional development and weighed in on their views of the Obama administration’s *Blueprint for Reform*. **Van Roekel** viewed it as a good place to start, framing a promising direction, but he noted that some teacher evaluation policy areas are too prescriptive and that a second generation of assessment tools is needed. He further stated that, although the status quo is not acceptable, a system that is imposed on teachers will not work. Rather, evaluation systems must be locally implemented, and the role of the state should be to create frameworks for local entities to use in developing their own meaningful evaluation system.

Dr. Domenech suggested that the average state education agency (SEA) lacks the capacity to implement the needed reforms. The reality is that state employees are being paid through federal grants, and he does not see how the SEAs can deliver on the *Blueprint for Reform* given their current capacity. **Traiman** took a poll of the audience to see whether states had a good model for evaluating teachers, and only one state representative raised a tentative hand. Traiman made clear that the private sector also often does not handle staff evaluation well.

An audience member asked for examples of states and districts that have strong evaluation systems, and the presenters offered these examples:

- Kansas
- Connecticut
- Denver
- Syracuse

Gretchen Weber, senior program associate at Learning Point Associates and director of field services at the TQ Center, closed out the workshop by polling state participants on their views about developing a teacher evaluation system. About half of the states considered themselves at the stage of “just dipping a toe in the water”; others were more advanced with their work. States viewed their teacher evaluation system as being tied more to their professional development system than their licensure, tenure, or compensation systems, although data-based links were made between all of these systems for many states. Despite the depth and breadth of knowledge gained during the workshop, the majority of state participants felt that there were still holes in their expertise, particularly in determining multiple measures for their teacher evaluation systems.