



Evaluating Teacher Effectiveness: A Workshop Connecting Research to Policy and Practice

**April 28–29, 2009
Hyatt Regency Denver at
Colorado Convention Center**

Executive Summary

Day 1

Luncheon Welcome and Keynote Address: “Teacher Effectiveness: The Research Base and a Policy Perspective”

This series of presentations was moderated by **Sabrina Laine, Ph.D.**, director of the National Comprehensive Center for Teacher Quality (TQ Center) and chief program officer at Learning Point Associates. Dr. Laine welcomed participants to the workshop and provided an overview of the current research on teacher effectiveness, as well as a federal perspective on the Obama administration’s increased interest in the topic and funding opportunities through the American Recovery and Reinvestment Act (ARRA).

Laura Goe, Ph.D., research scientist at ETS, reviewed research and practice on teacher evaluation methods along with issues to consider in designing and implementing such systems. According to Dr. Goe, the measurement of teacher effectiveness is affected by what is valued; technological advances and limitations; available data, evidence, and information; resources (e.g., staff, money, time, policy levers); and the cooperation of teachers themselves.

Dr. Goe reviewed two methods of teacher evaluation: value-added and classroom observation. Although value-added methods are used to track student academic progress, they are of little use in helping teachers improve their practice, because value-added scores reveal nothing about what goes on in teachers’ classrooms. Classroom observations, on the other hand, are great for formative evaluation, but are more expensive to conduct (i.e., in terms of personnel time, training, and calibrating) and are only as good as the instruments and the observers.

Dr. Goe outlined suggestions for designing a system to measure teacher effectiveness:

- Involve teachers and stakeholders in developing the evaluation system.
 - Increases teacher and stakeholder buy-in.
 - Increases validity of the system.
- Use multiple indicators, not just an observation score or a value-added rank.
 - There are many other important things that can be measured economically.

- Differentiate among teachers by grade level and experience.
 - Standards may be the same, but what constitutes effectiveness may be different.
- Use appropriate weights to give more importance to components of the system that are most important, i.e., some aspects of teacher effectiveness may be more important to the state education agency (SEA) or local education agency (LEA) than others.
- Measure what is most important to you, administrators, teachers, and other education stakeholders.
 - The system will drive improvement.
 - Teachers will strive to improve in areas that they know will count.
 - Doing so will ensure that what teachers are striving for is truly important in your definition of successful teaching.
- Give teachers opportunities to improve in areas where they score poorly.
 - Provide assistance in determining problem areas and planning strategies to address them.
 - Provide resources (e.g., time, dollars, people) to help them improve:
 - Mentors
 - Targeted professional development
 - Subsidized college courses
- For value-added:
 - Examine vertical alignment.
 - Look for “ceiling effects.”
 - Consider what to do about evaluating teacher effectiveness for the 50 percent or more of teachers who cannot be accurately evaluated with value-added models.
- For observations:
 - Set aside funds to support training and calibrating of observers.
 - Devise a system that involves multiple observations—not just one—during the year.
 - Use at least two different evaluators for each teacher and average the scores.
- Triangulate data:
 - Examine whether value-added scores correlate with observation scores and other measures.

Following Dr. Goe’s presentation, **Paul L. Kimmelman, Ed.D.**, senior advisor with the TQ Center and Learning Point Associates, shared a “Washington insider’s” perspective on teacher effectiveness and evaluation through the lens of ARRA. According to Kimmelman, the challenge from Secretary of Education Arne Duncan to all states is for bold reform through quick spending and thorough planning. His recommendation to workshop participants was to think about devoting stimulus dollars to the development of new data systems, since one of the four

assurances of President Obama's education plan focuses on building up longitudinal data systems nationwide.

After Dr. Goe's and Dr. Kimmelman's presentations, workshop participants were given the opportunity to ask questions of the speakers. Insightful and relevant questions included:

- Are funds available within ARRA to supplement the creation and implementation of teacher performance plans?
 - There is some flexibility as long as a program is compliant with the Individuals with Disabilities Education Act (IDEA), and \$200 million has been slated for the next round of Teacher Incentive Fund (TIF) grants.
- Has any research been done on validating evaluation processes in districts using Quality Compensation for Teachers (Q Comp)?
 - TQ Center does not conduct original research; Regional Educational Laboratory (REL) Midwest is open to pursuing such a study, but the Minnesota Department of Education has not shown interest yet. There are, however, examples of other places, such as Denver Public Schools, that have multiple measures of evaluation and are using those measures in conjunction with performance pay. The Center for Educator Compensation Reform (CECR) website includes a national map showing all states, cities, and districts where there are alternative compensation initiatives (see <http://www.cecr.ed.gov/initiatives/maps/>).
- What is the connection between the stimulus funds and Elementary and Secondary Education Act (ESEA) reauthorization?
 - It is safe to say that the four assurances of President Obama's education plan will mirror the core principles of ESEA, when reauthorized; these principles are built into the foundations of ESEA.
- What is the timeline for applications for Teacher Incentive Fund (TIF) grants?
 - TIF applications are slated for fall 2009, but it is crucial to engage stakeholders now so that program plans are in place when the request for proposal (RFP) is issued.

Working Session I: "Teacher Effectiveness Definitions and Measures for States"

The workshop's first working session featured a guided discussion facilitated by **Tricia Coulter, Ph.D.**, from TQ Center and Learning Point Associates. Dr. Coulter asked participants to think in detail about the specific variables to be measured within a hypothetical teacher evaluation system. During the report-out, members of the various working groups noted several competencies that all teachers should have (and on which they should therefore be assessed):

- Promoting student engagement with curriculum
- Applying knowledge of content and its relationship with other facets of learning
- Promoting a classroom environment that is conducive to learning
- Having a deep knowledge of pedagogy

Many of the group echoed these sentiments and raised issues concerning methods of and artifacts for assessment. According to one participant, “If you’re going to hold teachers accountable, you need to have an alignment with state curriculum and assessments, materials, and preparation programs.”

Upon completion of the roundtable discussions and report out, Dr. Coulter shared the TQ Center’s five-point definition of teacher effectiveness. In this definition, effective teachers:

1. Have high expectations for all students and help students learn, as measured by value-added or alternative measures.
2. Contribute to positive academic, attitudinal, and social outcomes for students, such as regular attendance, on-time promotion to the next grade, on-time graduation, self-efficacy, and cooperative behavior.
3. Use diverse resources to plan and structure engaging learning opportunities; monitor student progress formatively, adapt instruction as needed; and evaluate learning using multiple sources of evidence.
4. Contribute to the development of classrooms and schools that value diversity and civic-mindedness.
5. Collaborate with other teachers, administrators, parents, and education professionals to ensure student success, particularly the success of students with special needs and those at high risk for failure.

Working Session II: “Using Classroom Observation to Gauge Teacher Effectiveness”

Bridget Hamre, Ph.D., research scientist with The Center for Advanced Study of Teaching and Learning (CASTL) at the University of Virginia, facilitated this session with the overall goal of providing a rationale for qualitative methods of teacher evaluation and, more specifically, introducing workshop participants to the Classroom Assessment Scoring System (CLASS). The evaluation system, designed by Dr. Robert Pianta and colleagues at the University of Virginia, uses in-class and video observations to evaluate teacher-student interactions. Though this rapport between teachers and students often stems from teacher’s noncognitive skills such as sensitivity and flexibility, Dr. Hamre argued that with the proper professional development, even these teacher attributes can be shaped and honed.

After discussing the objectives of and strategies within CLASS, Dr. Hamre gave participants a chance to apply this learning to practice. In small groups, participants watched a short video of a high school mathematics lesson and then used the CLASS rubric to score the mathematics teacher regarding student perspectives and the quality of feedback given to students. Group members shared their observations of the mathematics teacher and of the overall experience of CLASS scoring. This report-out generated discussion about note taking and the appropriate avenue to teach note-taking methodology. Over time, Dr. Hamre reported, as observers become more comfortable with the CLASS tool, they learn to “chunk behavior” and exercise a kind of note-taking shorthand. Dr. Hamre also shared best practices with regard to observation and evaluation and described several essential components:

- Communication
 - Provide opportunities to learn the tool in a meaningful way.
 - Be thoughtful about how and with whom the results are shared.
 - Do not make inferences that go beyond the scope of the tool.
 - Concentrate on the feedback, not the score.
- Modification
 - Do not modify standardized tools, as this jeopardizes validity and reliability; supplement the tools instead.
- Timing, length, and time of year
 - Acknowledge that there is, on average, a slight decline in emotional support during the first 30 minutes of the day as well as toward the end of the day; support also will be lower in September, December, May, and June
- Rater effects
 - Minimize rater effects through training.
 - Remember that having more raters is better than having longer observation times.
 - Use video data so that many raters can code, which dramatically increases the reliability of codes.

Dr. Hamre shared with the whole group some current CLASS implementation projects. In the Office of Head Start pilot program, CLASS is being implemented in 50 Head Start grantee locations this year; feedback will be given at the grantee level, as program heads are interested in assessing low-quality programs in order to make decisions about eliminating funding.

In addition, Dr. Hamre and her CASTL colleagues are involved in a randomized controlled trial (RCT) funded by the National Institute of Health called MyTeachingPartner. In this experiment, which is targeted to rural areas where observers are scarce, observed teachers participate in a four-stage reflection cycle:

1. The teacher video-records him- or herself teaching and sends the video in to a CLASS consultant.
2. The consultant reviews the video and edits it based on predetermined areas of interest (delineated by the observed teacher).
3. The teacher reviews the video and feedback and completes a teaching self-reflection.
4. The teacher and consultant discuss teaching practices.

Dinner Keynote: “Using Multiple Measures: The Denver Experience”

This presentation, facilitated by **Sabrina Laine, Ph.D.**, featured Brad Jupp of Denver Public Schools. At the time of the meeting, Jupp was serving as senior academic policy advisor for

Denver Public Schools; as of this writing, he had been named senior education program specialist of the U.S. Department of Education.

Jupp enlightened workshop participants to one school district's history with and implementation of using multiple measures to get at the issue of evaluating teacher effectiveness. He touched on the issues of performance compensation, measurement tools, and the process of systems-reform while keeping an eye on the ultimate mission of education, which is the positive transformation of children's lives. Jupp said that, "I think of [teacher] effectiveness as evidence that you've increased the knowledge in the student body you've taught ... evidence that your kids have learned ... evidence that you've driven knowledge."

According to Jupp, there are four main tenets to the successful implementation of any evaluation system:

1. Rethink standards.
 - Standards must be value-driven and capture the attitude that all teachers believe that all kids can learn. It is important to be clear about these values at all times.
2. Keep it simple.
 - Complexity "generates no light and no heat" in the classroom. A simple evaluation system is always better.
3. Trust your evaluators.
 - Communicate with principals to match teachers with high value-added scores and take the practices of these teachers to scale.
4. Empower teachers and principals to observe these differences.
 - Strive for transparency and singleness of purpose.

After presenting his "big ideas," Jupp took some time to answer audience questions. Highlights of this session are below:

- *How would you measure the effectiveness of teams of teachers?*
 - As long as teachers have been working together all year, then both teachers get the incentive, but the issue of response to intervention (RTI) teachers who are only in the classroom one-third of the time remains unanswered. In the ProComp system, a teacher must teach his or her students for at least a semester and a day, and those students must be in class 85 percent of the time, in order to be eligible for the bonus.
- *How do you handle the push-back from teachers who are not teachers of record and will not be eligible for incentives?*
 - First, there is less push-back than imagined, because teachers migrate away from high-stakes grades and subject areas regardless of performance pay. The way to ensure fairness in a system such as ProComp is to provide a menu of incentive options and to commit to using multiple measures, such as a value-added score and an observation record.

- *In terms of teachers who do not qualify for the incentives, what support systems are in place to help them improve practice and dissuade jealousy?*
 - The menu of options has created a sense of fairness, as has the “trade-off” nature of high-stakes teaching and incentives. For instance, if you teach social studies, you are not eligible for the performance incentive, but on the other hand, if you teach social studies, you do not have to worry about your students being tested.
- *How do you adapt for special education teachers?*
 - In terms of individual student growth, teachers of students with mild or moderate disabilities are just as likely to succeed as general education teachers. Teachers with students with profound disabilities must write rigorous student learning plans; the utility of the creation of these plans is the conversation that takes place between the teachers, principals, and paraprofessionals regarding what should be included in the plans.

Day 2

Breakfast Keynote: “Teacher Effectiveness and Quantitative Methodology”

After having learned the previous day about qualitative methods of measuring teacher effectiveness from Dr. Hamre’s explanation of CLASS, **Daniel Goldhaber, Ph.D.**, a quantitative methodologist from University of Washington Bothell, The Urban Institute, and Center for Analysis of Longitudinal Data in Education Research (CALDER), discussed the utility of using value-added measures to estimate teacher effectiveness. Dr. Goldhaber noted that value-added measures are helpful in the realms of practice and policy because the numbers can be used to influence decisions regarding tenure and performance pay.

Dr. Goldhaber’s presentation generated many questions from participants, including:

- *It would seem from the numbers presented that value-added measurement only estimates 10 percent of a teacher’s contribution to a student’s learning. Is this accurate?*
 - Fairly, but this does not undercut the impact of that measure.
- *Is professional development an appropriate follow-up to value-added measurements of teacher performance?*
 - Yes, if the professional development is authentic and targeted. It’s important to remember that there are perverse incentives for much of the professional development we have now; we need to keep [professional development] ongoing and continuous, so that there is a loop of reflective feedback.

Christopher Brandt, PhD., of REL Midwest at Learning Point Associates introduced Dr. Goldhaber. Dr. Brandt informed participants that the sponsorship of Dr. Hamre and Dr. Goldhaber is representative of the new focus of the Institute of Education Sciences on experts bringing evidence to practitioners (EEPs). Dr. Brandt also reported that the regional educational laboratories are engaged in pursuing the following work in states:

- Regional needs assessment

- Fast-response studies
- Long-term rigorous research
- Technical assistance

Working Session III: “Determining Processes That Build Sustainable Teacher Accountability Systems”

During this session **Lucy Steiner**, a senior consultant from Public Impact, reviewed the results of surveys that were completed to explore the change processes in the creation and implementation of teacher evaluation systems. Steiner defined the objectives of the Public Impact study:

For this project, the PI study team conducted semistructured interviews with district personnel and state department of education officials who had participated in efforts to improve teacher evaluation or, in the case of New York City, teacher tenure systems. The study team conducted phone interviews with representatives from the following districts and states: (1) Denver Public Schools, (2) Chicago Public Schools, (3) New York City Department of Education, (4) South Carolina Department of Education, (5) Ohio State Board of Education, and (6) Minnesota Department of Education.

She used the findings to engage workshop participants in robust conversation about the change process taking place in their respective states and regions and informed them that she would be using their feedback as data for the next stage of the study.

Of the many goals that Steiner mentioned school officials identifying, several have received the most traction:

- Increasing student achievement by improving teacher quality
- Improving the quality of teacher-evaluation tools
- Linking teacher evaluation with professional development
- Building a comprehensive performance management system
- Linking evaluation results to alternative compensation systems
- Developing guidelines for districts looking to improve their teacher evaluation systems

After discussing these objectives, which resonated with audience members, Steiner highlighted challenges that also were identified in interviews:

- Principals’ concern about time and effort
- Teachers’ concern about the high-stakes nature of evaluation systems
- Human capital and capacity concerns
- Measuring the impact of a new evaluation system on student achievement
- Changing collective bargaining agreements
- Building trust among faculty members

Luncheon Panel Discussion: “A State Perspective on Teacher Evaluation”

During the final component of the workshop, state school officials bridged the gap between evaluation theory and research and the practice of implementing such programs. Representatives from three different states—**Kaneal Alexander** of Tennessee, **Diane Bradford, Ed.D.**, of Georgia, and **Larry Shumway** of Utah—gave perspectives on promising practices and everyday challenges from the field. Moderated by **Tricia Coulter, Ph.D.**, the panelists reflected on their own work and helped workshop participants translate their experiences into the beginnings of viable options elsewhere.

Alexander provided an overview of Tennessee’s comprehensive assessment system and framework for professional growth. Tennessee’s system uses a number methods by which teachers are evaluated, including self-assessment, review of lesson plans, and direct observation. These aspects are evaluated against a six-point rubric, and the results of this evaluation are used to direct professional development opportunities. The six points of the rubric are:

- | | |
|------------------------------|-------------------------|
| 1. Planning | 4. Learning environment |
| 2. Teaching strategies | 5. Professional growth |
| 3. Assessment and evaluation | 6. Communication |

Dr. Bradford reviewed the system Georgia has in place, which includes a comprehensive system called Classroom Analysis of State Standards (CLASS). The CLASS Keys Teacher Evaluation System is a set of standards against which teachers are evaluated. These standards are organized into five strands to teacher quality:

1. Curriculum and planning
2. Standards-based instruction
3. Assessment of student learning
4. Student achievement
5. Professionalism

The CLASS Keys system intersects with the School Keys system, Georgia’s data-driven system of school improvement and support, and informs professional learning, school improvement initiatives, and professional growth plans of individual teachers.

Shumway began his presentation by delineating the major essential components of high-quality teaching. Shumway noted that effective instruction is always:

- Student-centered
- Focused on right curriculum
- Explicit
- Assessed
- Adapted