



District  
Management  
Group

## SPOTLIGHT

# From Teacher Quality to Effectiveness: Developing a Systemic Approach

*by Nicholas P. Morgan*

Originally published in *District Management Journal*, v.6 Winter 2011

*Districts have a significant opportunity to rethink how professional development is delivered, shifting toward models where professional development is aligned more closely with the district's vision for effective teaching.*

# From Teacher Quality to Effectiveness: *Developing a Systemic Approach*

“Our challenge is to make sure every child in America is learning from an effective teacher—no matter what it takes. So today, I ask you to join President Obama and me in a new commitment to results that recognizes and rewards success in the classroom and is rooted in our common obligation to children.”

– Secretary Arne Duncan in a speech to the NEA, July 2, 2009

| NICHOLAS P. MORGAN

As Secretary Duncan’s articulation of the national challenge illustrates, our pursuit of improved teaching and learning has shifted in response to research findings and outcomes data: the pursuit of “highly qualified teachers” has transitioned to a pursuit of “teacher effectiveness.” Race to the Top and the ESEA Blueprint ask states and local districts to establish definitions of teacher effectiveness “that are based in significant part on student growth and also include other measures, such as classroom observations of practice.” No single measure of student learning, standardized test or otherwise, is a complete or fair measure of what students learn or how teachers teach.

This discussion is not a new one. Since *A Nation at Risk* was released in 1983, captivating the dialogue of education reform, a focus on teacher **quality** has been at the forefront of discussions by practitioners and researchers alike. However, a second paradigm—that of teacher **effectiveness**—has existed nearly as long. Released by the Task Force on Teaching as a Profession in 1986, *A Nation Prepared* focused on teaching professionalism and the notion of the highly *effective* teacher. A focus on improving teacher effectiveness forces us to look at student performance outcomes instead of teacher characteristics or qualifications.

Over the last fifty years, education research has focused significant resources and attention on understanding which teacher characteristics and qualifications drive classroom success.

Unfortunately, research has not been able to paint a picture where certain qualifications lead to student success. Broad categories of qualifications, like teacher preparation (including coursework, selectivity of sending institution, certification type, test scores and more), have been sliced in many ways to find nuggets of predictive information, mostly with little success. Further, since the No Child Left Behind Act’s stipulations have resulted in a Highly Qualified Teacher in the vast majority of classrooms, where does the conversation progress from here?

The recent Race to the Top competitive grant process has provided a catalyst for new thinking and practical approaches to teacher effectiveness, and has helped shed light on current opportunities nationwide. Teacher effectiveness is a central tenet in each of the four “core assurances” of the program, which include college and career-ready standards and assessments, teacher effectiveness and equitable distribution of effective teachers, data systems to support student growth and evaluation, and support and interventions for turning around the nation’s lowest performing schools.

Of the 41 states responding to the first round of the Race to the Top competition, only about half of the states had quantitative growth models in place or in progress. Growth models are data models to track student achievement growth over time, allowing deeper analysis of teacher effectiveness trends. The status demonstrates that nationally, the country is still in the early stages of incorporating quantitative

student achievement outcomes in myriad applications for improved district management. Round one of Race to the Top also demonstrated a shift in the way qualitative teacher effectiveness data are being gathered. Of the 41 first-round respondents, only ten states differentiated teacher effectiveness using multiple rating categories. The current pass/fail rating categories that are commonplace do not allow for nuanced information to help improve teacher support and other related district services. Nationally, the country is still in the early stages of incorporating improved methods of collecting qualitative student achievement outcomes for human capital management. This article and its corresponding toolkit address approaches for combining qualitative and quantitative teacher effectiveness measures to improve a district's evaluation systems.

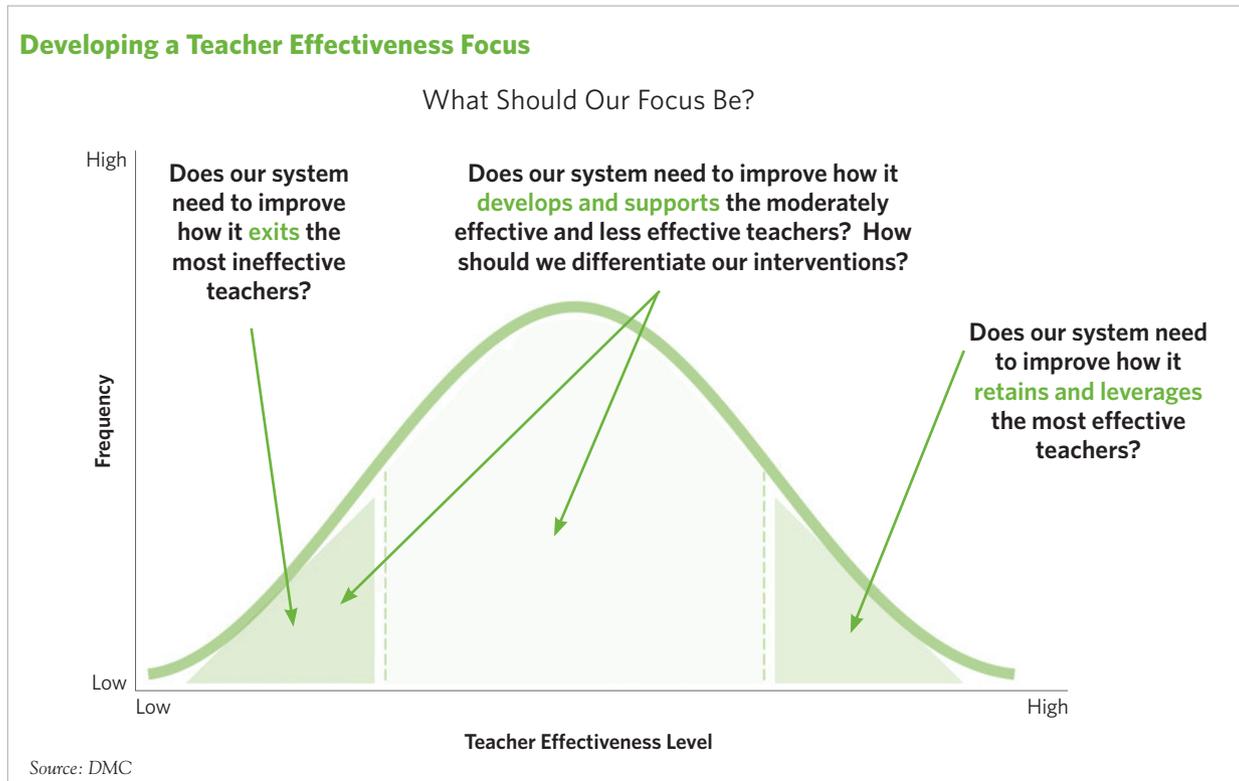
Evidence from teacher effectiveness research continues to underscore the importance of great teaching: an effective teacher can quickly change the academic trajectory of a student and an ineffective teacher risks derailing progress. Research also continues to show that teacher effectiveness is highly variable. Rather than continuing to search for evidence of what does work, a growing emphasis in the national dialogue is on

policy reform based on evidence of practices that don't work. For instance, should districts continue to invest in qualifications that have no demonstrable effect on student outcomes?

Research conducted by Robert Gordon, Doug Staiger, and Tom Kane has been instrumental in drawing out some key insights regarding teacher effectiveness.<sup>1</sup> First, we should recognize that while small variances may exist between different categories of teacher qualifications, the bigger issue is the broad distribution of effectiveness within each category. For instance, whereas teacher certification predicts little about a teacher's overall effectiveness, variance within categories of certification is broad. This research underscores a few key conclusions, all demonstrating how much effective teaching matters for student achievement outcomes.

The research conclusion that past performance is predictive of future performance serves as one of the underpinnings for the increased use of value-added data in teacher evaluations. Breaking out performance categories allows us to focus on what we should do to 1) shift the curves and 2) manage the shape of the performance distribution. Importantly, as shown in Figure 1 below, districts need to ask some hard >

FIGURE 1



questions regarding overall district strategy as they tackle systemic pursuits of teacher effectiveness. If we consider a generic performance distribution, at least three broad categories with corresponding strategic questions emerge: 1) the most effective teachers, 2) the broad middle, and 3) the least effective teachers. Once identified, what should the district’s role be in managing these distinct groups? While most districts are likely to answer “We need to do all three,” the more difficult questions arise when thinking about prioritization and sequencing. What should the district tackle first? When resources are limited, what should our priority be?

This begs the question: how do we manage the effectiveness curve? Is our goal to shift the entire curve to the right, or is it to fundamentally alter the shape of the distribution? How should we allocate our resources accordingly?

A widely-read report entitled *The Widget Effect* from the New Teacher Project has been an additional national catalyst in helping define concrete issues to improve teacher effectiveness.<sup>2</sup>

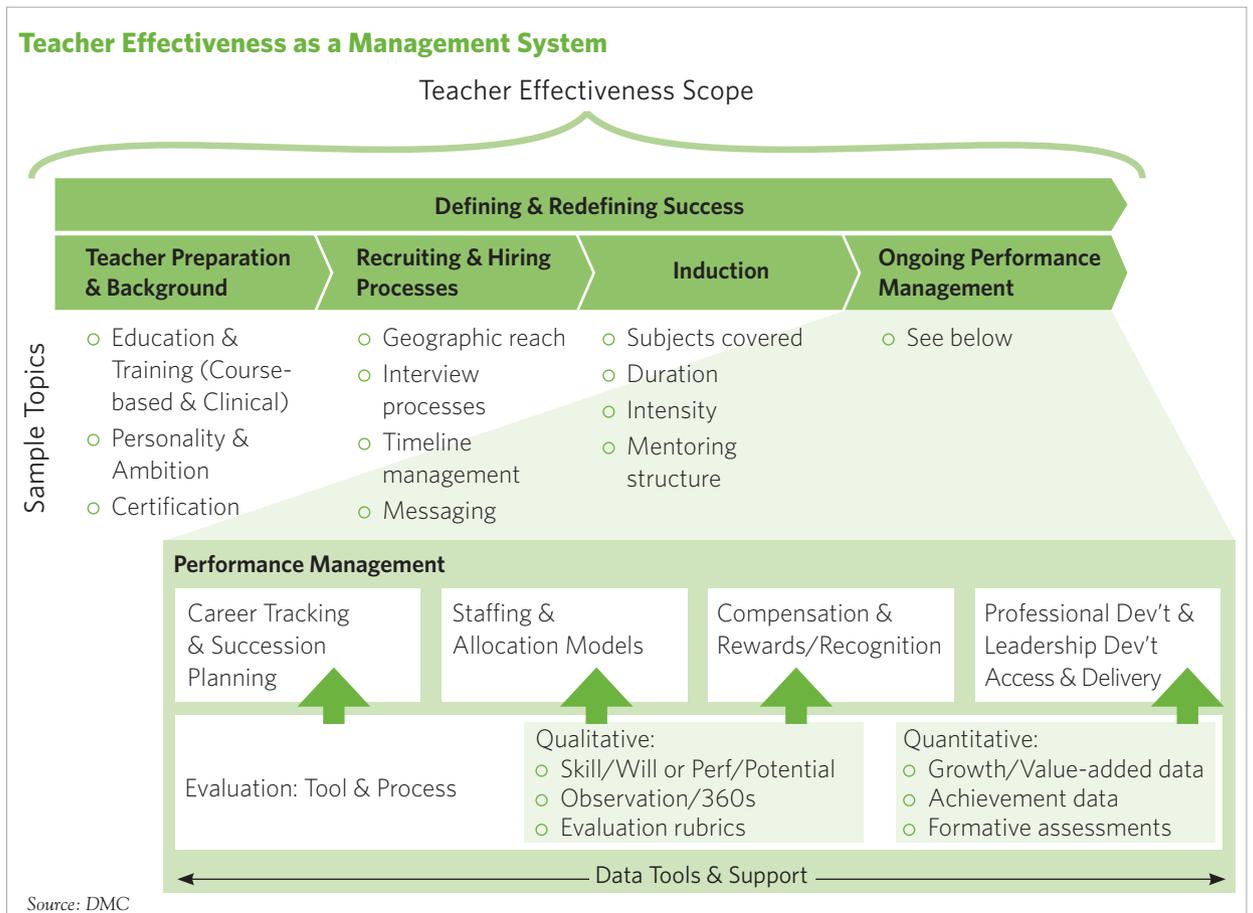
First, the overall notion that many district functions are often run without taking performance into account signals the scope of the opportunity that improved performance data might yield. Such functions as recruitment, placement, professional development, compensation, retention, and layoffs are rarely managed using teacher performance, if at all. In practice, the “widget effect” is characterized by institutional indifference to variations in teacher performance.

Teacher evaluation systems reflect and reinforce this indifference in several ways. *The Widget Effect* has helped a dialogue to emerge around five systemic deficiencies. They are:

- All teachers are rated good or great
- Excellence goes unrecognized
- Inadequate professional development
- No special attention to novices
- Poor performance goes unaddressed

Each of these issues helps to inform an opportunity for district leadership to rethink how teacher effective-

FIGURE 2



ness is managed. How good is our teacher evaluation system? How well is it implemented? Does the performance data from our evaluation system inform the delivery of professional development? How do we allocate our development resources to those most in need, including novice teachers or those struggling? How do we reward our most effective talent?

### Teacher Effectiveness as a System

The importance of a robust evaluation system has been a recent focus of major district reform agendas, often with significant national attention. How does the evaluation process play into the larger question of teacher effectiveness? How do we actually define and measure what an effective teacher looks like? In order to further define and structure the pursuit of teacher effectiveness reforms, DMC's framework, shown in Figure 2, lays out key human capital components, with an emphasis on the role played by good evaluation tools and processes.

As the framework shows, teacher effectiveness in its entirety is a far-reaching topic that addresses most human capital elements that involve teachers. From the moment a prospective teacher begins his/her training to the day that teacher exits the profession, each stage offers an opportunity for the district to rethink how it can improve and support teacher performance and student outcomes.

Overall, districts should begin by defining what success looks like for the system overall, as well as more specifically for key stages in a teacher's career progression. For example, how do you really measure how successful your induction program is? How closely tied is that measurement to student outcomes? DMC focuses significant attention on the power of outcomes-based management, and teacher effectiveness is no exception. Articulating specific performance goals and managing toward real results allows the district to focus its resources and measure progress. For DMC, these performance goals should span student achievement as well as operational and financial effectiveness. Ultimately, for teacher effectiveness to be managed cohesively, the district will need a common vision and articulation of what effective teaching looks like, allowing teachers and administrators alike to pursue a common goal.

As introduced above, improving teacher effectiveness can no longer be a conversation simply about pursuing

different teacher qualifications. Districts must manage performance more actively on an ongoing basis, and allocate limited resources in a more strategic manner. "Performance management" is a buzz phrase used in a variety of ways in education. DMC thinks of performance management as a collection of activities that seek to increase the overall performance of the district toward clear systemic goals. For teacher effectiveness, performance management must be driven by the teacher evaluation "engine", which in turn drives a variety of related human capital processes. Performance management processes need to be supported by data systems that support putting key information and insights in decision-makers' hands.

“Districts have a significant opportunity to rethink how professional development is delivered, shifting toward models where professional development is aligned more closely with the district's vision for effective teaching.”

From an ongoing process perspective, performance management includes a focus on the delivery of targeted professional development to support teachers' growth and development. Districts have a significant opportunity to rethink how professional development is delivered, shifting toward models where professional development is aligned more closely with the district's vision for effective teaching. A vision of managing teacher effectiveness as a support system versus one that has a punitive focus is critically dependent on its orientation to professional development.

Performance management also includes managing incentives, such as rethinking compensation and recognition. While not the focus of this *Spotlight*, compensation reform continues to evolve nationally, with a variety of models being evaluated. Districts are exploring and implementing innovative combinations ▶

of performance-based financial bonuses, as well as considering using performance to modify base pay models such as traditional “steps and lanes” schedules.

Additionally, as the Race to the Top core assurance states, teacher effectiveness is also about the *equitable distribution* of effective teachers. Performance management processes should also enable a discussion to address the intentional distribution of the most effective teachers across the district. Staffing models should also be well informed by the data generated from robust teacher evaluation systems. Finally, additional policy-related decisions such as tenure policy should also be based on rigorous analysis of teacher effectiveness data.

“For a fuller picture of teacher effectiveness, it is important to combine qualitative measures of teacher practice with quantitative measures of student achievement outcomes.”

### A Focus on the Evaluation System

Successful evaluation systems need two main components: 1) a set of tools, including rubrics and scoring mechanisms, and 2) a good process to drive timely use of the tool with fidelity. Districts may reflect on their current practices and find that they have one but not the other. A poor tool will not yield effectiveness insights for the teacher or district, and a poor process may yield low compliance, low fidelity, problems with inter-rater reliability, high confusion, and more. A good set of tools should provide useful information to support key decisions in the overall scope of teacher effectiveness. It should be evaluative, but it should also be developmental. Further, it should be simple enough to be understood and used without misinterpretation. Perhaps the most important hallmark of a good process is that it is actually followed with fidelity. How would you rate your district?

Many districts may find that both the tools and the process of teacher evaluation need a second look.

While there is no single established “best practice” for teacher evaluation, guidance for structural considerations is becoming more commonplace. For instance, the New Teacher Project recommends six design principles:<sup>3</sup> 1) an annual process, 2) clear, rigorous expectations, 3) multiple measures, 4) multiple ratings, 5) regular feedback, and 6) significance.

Regarding process, frequency of evaluation has been identified as a significant issue. Data from the National Center for Teacher Quality on collective bargaining agreements has highlighted the variability and infrequency of teacher evaluations in the country’s fifty largest school districts.<sup>4</sup> For tenured teachers, only 17 districts require evaluations at least once a year. For untenured teachers, the number of districts receiving annual evaluations only rises to 26. How the process is conducted is also of great interest to the broader teacher effectiveness pursuit.

Do the measurement approaches (typically classroom observations) have sufficient rigor or frequency to conduct a truly useful evaluation?

Regarding the tools, the structure of the evaluation rubric itself is the source of significant attention in districts nationwide. Many evaluation models exist that have received significant attention for their structures, but little consensus exists regarding a single preferred approach. More comprehensive models may have a higher number of indicators, but may sacrifice ease of use. Conversely, a district may prioritize simplicity to encourage frequency of implementation or a low time burden on staff. Districts may want to modify an existing structure to adapt to local context and needs. Finally, is there a possibility to broaden the use of other qualitative measurement approaches such as peer or student surveys? Further detail on these structural design elements are addressed in the accompanying DMC toolkit item: *How to Design a Multidimensional Teacher Evaluation System*.

Many typical teacher evaluation forms are constructed on a “pass/fail” basis, which does not allow for a more granular understanding of performance, and does not allow for more nuanced teacher support. What the tool itself measures should be scrutinized to match the district’s vision for effective teaching.

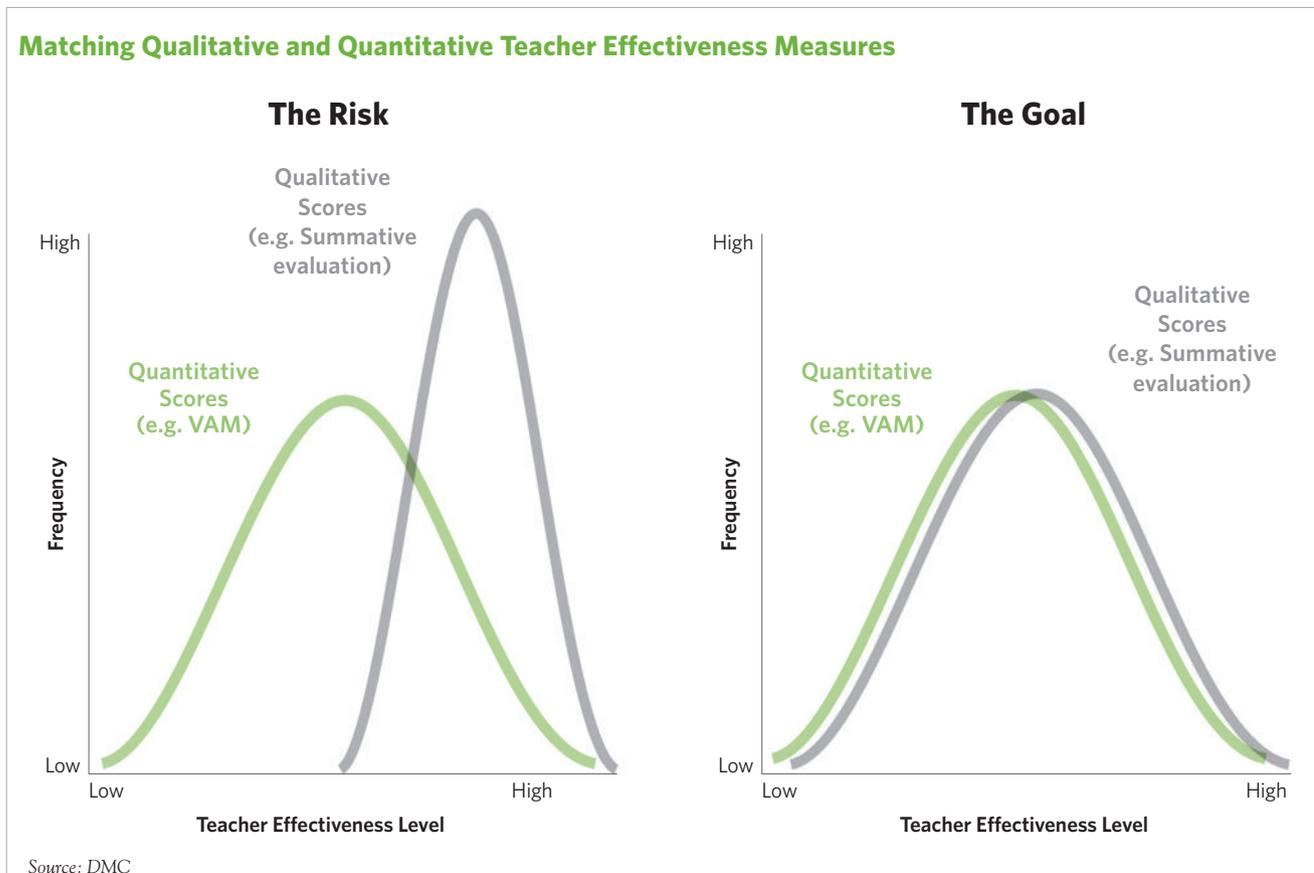
Finally, as districts look to redesign performance measurement, it is important to consider what the core stakeholder group involved—the teachers—think about the topic. Survey data shows that what teachers

consider accurate performance measures may surprise you. The Bill & Melinda Gates Foundation, in conjunction with Scholastic and Harris Interactive, surveyed forty thousand teachers in 2009, asking them their views on the accuracy of various performance measures.<sup>5</sup> Interestingly, traditional principal observation and review ranked roughly mid-pack for accuracy as a performance measurement approach, with about 20% of teachers reporting it as “very accurate.” Regarding quantitative data, teachers appeared reluctant to embrace achievement status as a measure, ranking it lowest of all measures in terms of accuracy. However, using student achievement growth over the course of an academic year was the second-most accurate measure listed, with 55% of respondents calling the measure “very accurate.” The most accurate measure, reported by 60% of survey respondents to be “very accurate,” was student engagement, although it was not clear which specific measurement protocols the teachers had in mind.

### Quantitative & Qualitative Measures

Historically, most teacher evaluations have focused on teacher practice only, using a set of qualitative measurement approaches such as classroom evaluations and artifact analysis. For a fuller picture of teacher effectiveness, it is important to combine qualitative measures of teacher practice with quantitative measures of student achievement outcomes. The national dialogue has recently centered significantly on quantitative metrics to track teacher effectiveness. With high-profile stories capturing national headlines – such as workforce reductions in Washington, D.C. based largely on quantitative teacher effectiveness data, or the *LA Times*’ report on the Los Angeles Unified School District’s teacher performance using value-added data, the debate over appropriate use of student achievement data in teacher evaluations has been raging. At the heart of this debate are value-added statistical models, which attribute a portion of student achievement growth over time to a specific factor such as a school, teacher, or program. This article addresses these models in further detail below. ▷

FIGURE 3



The federal Race to the Top program has served as a catalyst for this conversation, as states devoted significant effort to include quantitative measures for academic growth in their plans. Importantly, despite apparent agreement that multiple measures are a good thing, there is little consensus about the weight that qualitative or quantitative measures should be given in a teacher's summative evaluation. Many states, such as Tennessee, Rhode Island, and Florida set requirements that at least half of the total be student achievement-based. The Washington D.C. public school system, with its innovative IMPACT teacher evaluation program, has no less than twenty different categories of evaluations for different groups of employees, each with different weighting of quantitative and qualitative components. Almost by definition, district evaluation systems that combine teacher practice and student achievement evaluation components are new, and a knowledge-base about the how these systems actually works in practice is still nascent.

The challenge that districts face in combining qualitative data regarding teacher practice and quantitative data on student achievement outcomes is that the stories these data sets tell regarding teacher effectiveness might be very different at the outset. In a world such as the one described in *The Widget Effect*, where teachers predominantly get high marks for performance, how will the system react to a conflicting picture based on student achievement data? This potential conflict is demonstrated in Figure 3. With fundamentally different performance distributions possible from quantitative and qualitative measurement approaches, districts must focus on developing systems to bring the distributions in line with one another. Essentially, this illustrative example underscores a significant research agenda for the coming years: which qualitative measurements and systems actually correlate with student achievement outcomes? While selective studies have been conducted on specific evaluation rubrics, this field of study has little widespread practical research to guide new rubric design for qualitative measures to correlate with student achievement outcomes. For example, in Hamilton County, Tennessee, where a new evaluation system is being implemented, the district will investigate correlation to the state's value-added system (TVAAS) over the coming year and adjust the system based on its findings.

Another good example of how this challenge needs to be managed is in New Haven, Connecticut, where the district received significant media praise for its collaborative effort with the union to redesign the teacher evaluation system. The new system gives teachers a summative ranking on a one-to-five scale on two dimensions: 1) Instructional Practice and Professional Values (weighted 80%/20%) and 2) Student Learning Growth. In this type of structure, it is possible that a teacher be given a low rating for Instructional Practice and Professional Values, but demonstrate high growth for their students. Conversely, it is possible for a teacher to demonstrate high scores for Instructional Practice and Professional Values, but have little student growth. In either situation, focused policy review needs to be conducted outside the context of the specific teacher's evaluation, to determine why such a mismatch is occurring and what, if anything, needs to be corrected. In New Haven, the individual ratings will also be reviewed to ensure that the given rating in these situations is fair and accurate based on evidence shared by the instructional manager and teacher. Individual ratings may be adjusted for unfairness or inconsistency.

### Using Quantitative Student Achievement Data in Evaluations

Many districts have acknowledged the need to have a greater connection between their evaluation systems and student outcomes in order to measure and improve overall teacher effectiveness. However, *how* this should take place and with which data has been the subject of intense scrutiny and debate recently. First, the usual caveats apply. Only certain subjects are tested, and even then, the overall quality of the insights is dependent on appropriate test design.

“Achievement” or “status” data compares results of a snapshot in time to a benchmark, and therefore offers few real insights for individual teacher effectiveness. Likewise, “improvement” data, which looks at the change in these snapshots over time, measures the academic performance over time, but for different cohorts of students, so again offers few real insights for the effectiveness of any individual teacher. Beginning significantly with changes to Adequate Yearly Progress measurement under the No Child Left Behind Act, *growth* has increasingly become the focus of the national dialogue. The use of the term “growth model”

can encompass different analytical approaches, however, and much recent attention has been focused on “value-added” models which some consider a subset of growth methodologies.

Generally speaking, growth models track the test scores of the same students from one period to the next. For example, gain scores can be computed to compare the performance of the current year’s fourth graders with that of the same group of students last year, when they were in third grade. However, growth models usually do not control for student or school background factors, and therefore they do not attempt to address which specific factors are responsible for student growth. This is the goal of value-added models, which are complex statistical models that isolate the effect of a teacher, program, or school. Further adding to the complexity is the fact that many different value-added models are being used across the country.

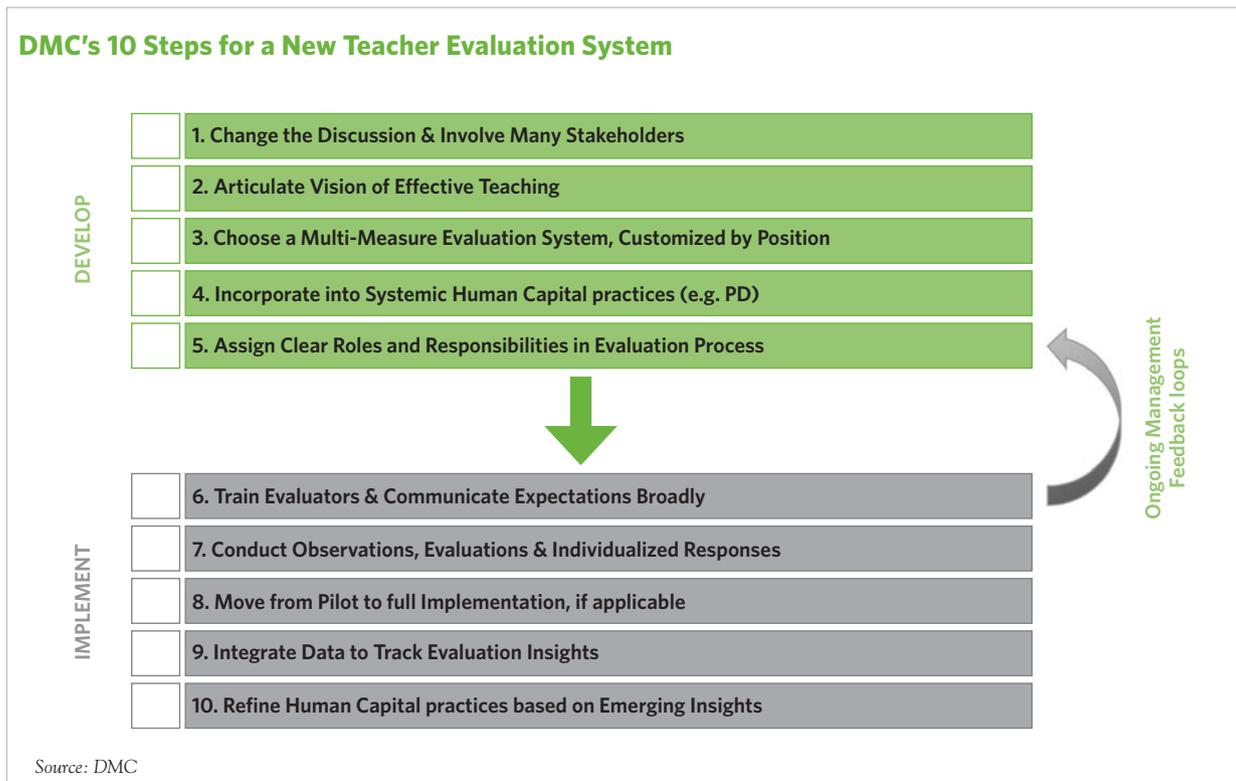
Value-added models have been the source of most of the scrutiny and debate surrounding such prominent endeavors at the Washington D.C. public schools, the *LA Times* exposé on LAUSD, and similar freedom-

of-information act requests by various media outlets for data from the New York City Department of Education. While beyond the scope of this *Spotlight*, understanding value-added models is a complex endeavor. The criticism few would disagree with is that these approaches are best debated by professional statisticians.

So, what exactly is a value-added model? Value-added is the difference between two “possible outcomes”: 1) the observed outcome the teacher (and school) actually experienced and, 2) the expected outcome given an “average teacher experience” or given the student’s latent academic growth rate. In other words, value-added is equal to the actual growth minus the predicted growth. We want to see value-added scores that are above the prediction.

Many strengths and weaknesses have been addressed with specific value-added modeling approaches, but DMC remains optimistic that this type of quantitative approach to measuring teacher effectiveness is a valuable addition to a suite of measurement components. Pursuing a new, more informed way of measuring teacher effectiveness, and using that data to improve >

FIGURE 4



overall systemic management is an innovation in the world of K-12 leadership and management.

As DMC has highlighted in past work, innovation requires testing and refinement, and the use of value-added methods to improve teacher effectiveness should be pursued with an innovation mindset.

Further, the recent decision by the Department of Education's What Works Clearinghouse to modify the definition of "gold standard" research will allow greater use and application of quasi-experimental research methods like value-added modeling. Value-added modeling can be used to pursue policy insights for human capital dimensions such as tenure, recruiting, staffing, and compensation policies, all of which are critical components of the performance management system outlined above.

## Developing and Implementing a Teacher Effectiveness System

As shown in Figure 4, DMC has designed a process for districts to pursue a more robust teacher effectiveness program that emphasizes the systemic nature of teacher evaluations. The first two steps are critical for laying the foundation of a new system in the district. First, districts need to change the nature of the discussion about teacher effectiveness in the district and involve a broad range of stakeholders in that discussion. Second, districts should articulate a clear vision of what great teaching should represent.

The third step in the process is to create a multi-measure evaluation system that assesses how well the vision for great teaching is being met. The brief toolkit exercise included in this magazine is intended to stir critical thinking for this process. Decisions need to be made as to which qualitative and quantitative measures to use, who will gather this information, how and how often it will be gathered, and how much weight to ascribe to the various measures chosen. DMC offers more in-depth tools for supporting this process as well. The evaluation system should form the basis upon which developmental teacher support can be provided. The key design elements should support broader conversations to address other positions, in particular those where a closer articulation to student outcome responsibilities should be defined.

Great teacher evaluations should support teacher development, and should also provide critical information to help improve a variety of district-wide human capital systems, including recruiting and staffing models. Once the revised evaluation system has been agreed upon, the district can proceed to the final two steps in the development process: aligning human capital practices, such as professional development, to the vision for effective teaching created in the previous steps, and assigning clear roles and responsibilities for execution.

The final five steps in DMC's process address implementation stages. Perhaps most important is step six, where training and communications occur broadly. Rollout and change management issues are addressed in steps seven and eight, which should be designed with focus on the local context. Step nine addresses the necessary data infrastructure and analysis to support the system. Finally, step ten highlights a need for ongoing alignment of human capital activities based on the insights generated from the new system. This system is designed to be a continuous improvement cycle, and should assist the district in becoming a true learning organization. DMC offers a variety of tools and services to assist districts with this process.

As districts pursue teacher effectiveness, direct and indirect costs should also be part of the discussion. DMC believes that pursuing key steps in teacher effectiveness reform can mean a lot work, but that it can be done quickly and relatively inexpensively. Further, DMC encourages districts to consider whether systemic improvements may actually result in greater fiscal efficiency for the district.

<sup>1</sup> Gordon, R., Kane, T. and Staiger, D. Kane "Identifying Effective Teachers Using Performance on the Job," *The Hamilton Project* white paper 2006-01, The Hamilton Project, Washington, DC, 2006.

<sup>2</sup> Weisberg D., Sexton S., Mulhern J., and Keeling D. "The Widget Effect: Our National Failure to Acknowledge and Act on Teacher Differences," The New Teacher Project, 2009.

<sup>3</sup> "Teacher Evaluation 2.0: Six design Standards," The New Teacher Project, 2010.

<sup>4</sup> National Center for Teacher Quality TR3 Database.

<sup>5</sup> "Primary Sources: America's Teachers on America's Schools" A Project of Scholastic and the Bill and Melinda Gates Foundation, 2010.



**NICHOLAS P. MORGAN** IS MANAGING DIRECTOR AT THE DISTRICT MANAGEMENT COUNCIL. HE CAN BE REACHED AT [NMORGAN@DMCOUNCIL.ORG](mailto:NMORGAN@DMCOUNCIL.ORG).