



November 2012 | Volume 70 | Number 3

Teacher Evaluation: What's Fair? What's Effective? Pages 58-63

## A Tale of Two Districts

Mark Simon

**There's no need for risky experimentation; we know what works in teacher evaluation.**

These days, everyone seems to be wringing their hands about how to construct new evaluation systems that will make teachers better. This unnecessary angst has led to crazy experiments in reform that have embraced churn for the sake of churn, put school districts at risk, and demoralized many of our most talented teachers.

A few school districts, however, have resisted panic, pressures, and fads. Instead, they have invested in models that work.

### Two Districts from a Bird's Eye View

My daughter just graduated from high school in the District of Columbia Public Schools. This school system has been at the forefront of a risky, six-year adventure in "bold" reform. The centerpiece of the reform is IMPACT, a teacher evaluation system that has been both lauded and criticized. It's been a rocky ride. The new teacher evaluation approach was implemented in 2009 under mayoral control. The mayor under whose watch it was initiated was defeated in his reelection bid, and schools chancellor Michelle Rhee resigned in 2010. But the reform agenda has continued largely unchanged until this year, although its track record is, at best, mixed.

Right next door in Maryland, Montgomery County Public Schools' teacher evaluation model is in its second generation after 12 years. I taught in the county and then represented teachers when the district administration and the teachers union collaborated in crafting the new teacher evaluation system. Over the years, Montgomery County's system has produced rich evidence of success. Most important, the collaborative relationships established between the administration and the teachers union have enabled the district to continue to refine what works.

Far from being an objective observer, I have been intimately involved in the efforts in both districts, with a close-up view of how these teacher evaluation reforms were developed and how they are perceived by the workforce. I believe that a comparison of their track records shows that one approach is actually better. Such a comparison enables us to distill a set of principles that are crucial to the success of teacher evaluation systems. (See "[Effective Teacher Evaluation: Lessons from Experience](#)" on p. 61.)

### The Montgomery County Case

In the 1990s, Montgomery County was a good school district, but not a groundbreaking one. Collective bargaining, which had begun in 1968, had never produced much substantial change. But in 1997, the teachers union made an audacious proposal. The union wanted to include in the bargaining process issues of quality control, professional roles, school culture, accountability, and other areas normally reserved as management prerogatives. Amid controversy, and after a divided school board vote, the district accepted.

It was a risk on both sides. As the preamble to the 1998 contract stated,

*For the union, taking responsibility for the improvement of the quality of teaching and learning represents an expanded role, and for the administration, forging a partnership with the union over ways the system and schools can improve is also new.*

This contract began a collaborative, 18-month process of redesigning the teacher evaluation system, which was renamed the Professional Growth System.

### A New Philosophy

In a nutshell, the new philosophy on teacher evaluation viewed teaching as incredibly complex work. No teacher comes to the work entirely competent. The focus of teacher evaluation in Montgomery County is professional growth—the nurturing of good

teaching—not the sorting and ranking of the teacher workforce.

Although an evaluation system must be able to weed out people who never should have entered teaching, that objective only applies to a tiny percentage of the workforce and must not be the system's main purpose. Good teachers are not found through some magical recruitment pipeline. They are made, over time.

## Elements of the Montgomery Model

*District-union collaboration.* Beginning in 1997, there has been consistent collaboration between labor and management in the design, implementation, and evaluation of the Professional Growth System. That collaboration has spread to other issues related to budgeting, instruction, curriculum, and assessment. Since 1999, the three staff unions have sat at the superintendent's leadership team meetings.

*Quality control through Peer Assistance and Review.* The Peer Assistance and Review (PAR) program is overseen by a panel of 16 full-time teachers and principals who are recommended by the teachers and principals unions. The PAR panel interviews, hires, and then oversees the work of approximately 40 consulting teachers, who work full time with teachers new to teaching and veteran teachers who are having difficulty. Consulting teachers are the "best of the best," and they are selected for their ability to work well with their adult peers. They must agree to return to classroom teaching positions after their three years in their consulting role.

*Less frequent evaluations; more support for improvement.* Rather than devote scarce school district resources to routine, bureaucratic annual evaluations, Montgomery County decided to devote the resources where they were most needed. Teachers who are doing a good job are formally evaluated just once every three, four, or five years depending on their length of service. The focus is on teachers new to the district and on more experienced teachers who are having difficulty.

*Principals' role in triggering a formal evaluation.* Principals have the discretion to identify teachers who seem to be struggling and who may need an evaluation outside their regular evaluation cycle. But the principal's identification also triggers an independent second opinion from a PAR consulting teacher, who is accountable to the independent PAR panel. This quality-control mechanism ensures that personality or style conflicts between teachers and principals don't get elevated into unfair judgments of any teacher's practice.

*Intensive support for new teachers.* Teachers new to teaching are each assigned a consulting teacher from the PAR program. After observing the new teacher and determining how much support he or she needs, consulting teachers have the discretion to leave the teacher on his or her own or to spend significant amounts of time observing, helping with planning, demonstrating model lessons, and so on. Each consulting teacher has a caseload of about 18 clients and develops a stake in the success of each one.

*Knowledgeable evaluators.* Principals, PAR consulting teachers, department chairs, and others with a role in teacher evaluation attend 12 full-day classes that provide intensive training in what constitutes good teaching. This deep training may be the most important part of the evaluation system. Evaluators develop skill in how to observe teachers and how to work with them to improve their practice. They are trained, first and foremost, to be respectful coaches for a craft that can be done well in many different ways.

Evaluators don't come in with a checklist that they use to write up a generic observation or final evaluation. Each evaluator uses his or her expert judgment. The evaluator begins by considering the teacher's intention and how well the teacher is achieving it, not by applying a standard rubric that prescribes what the teacher should be doing with students. In other words, evaluators are teacher leaders. Their knowledge base is their credibility.

*A culture focused on teaching and learning.* Teachers, too, are encouraged to take a series of courses called Studying Skillful Teaching (up to 12 full-day classes, or three 3-credit courses), which help them develop a common language to talk about the knowledge base and skills involved in the complex craft they practice. In this way, the teacher evaluation system has elevated the culture of the whole district and focused it on teaching and learning. It has changed the nature of conversation in the faculty lounge.

## The District of Columbia Case

Montgomery County's neighbor to the south, the District of Columbia Public Schools, was ready for a change in 2007. Its evaluation system, a bureaucratic obstacle to quality control and teacher improvement, needed to be replaced. The school board had been abolished, and Michelle Rhee was hired as the district's first chancellor in a move to centralize authority under mayoral control. Rhee, who had never run a school or managed a school system, set about remaking the teacher evaluation system with a free hand.

### Design Behind Closed Doors

Through contentious bargaining in 2007–08, Rhee cleared the path for a new evaluation system. A 1996 act of the U.S. Congress, which still exerts much control over the District of Columbia, gave the administration the right to unilaterally impose

changes in the teacher evaluation system.

Rhee deputized former national teacher of the year Jason Kamras and a consultant, Michael Moody, to develop the system behind closed doors. Kamras and Moody used language and ideas from many other systems to construct a method for conferring a judgment on each teacher's worth. But there was little evidence in Rhee's first year that improving teaching and learning was a priority (Davis, Sylvia, & Simon, 2008). According to Kamras, the initial goal was simply to "identify the highest performing teachers and ... to identify and focus support on the lowest performing teachers and be able to release them if they don't improve" (Curtis, 2011, p. 12).

The school district rolled out its IMPACT system in 2009. Although IMPACT was a giant step forward, bringing a focus on the components of good teaching for the first time, the system has remained controversial—a point of conflict between the administration and the teaching workforce.

## Elements of the D.C. Model

*One-size-fits-all process.* Under IMPACT, every teacher is formally observed five times a year—twice by a master educator and three times by the school principal. Each observer writes a formal post-observation report that is based on a rubric. The rubric translates the observation into a numeric score that ranks the teacher as *highly effective*, *effective*, *minimally effective*, or *ineffective*. Teachers who are rated *highly effective* get substantial bonuses. Those who are rated *minimally effective* are given one year to improve; if they don't, they are then subject to dismissal. Teachers who receive an *ineffective* rating can be fired.

*Minimal training for evaluators.* Each master educator has a caseload of 80 to 100 teachers. That means master educators only see each teacher when they show up (usually unannounced) for an observation and once a few days later for a conference to explain the observation write-up and score. Master educators are trained to faithfully apply the rubric, with careful attention given to making sure that different evaluators would reach the same conclusion on the basis of the rubric. Almost all (87 percent) of the master educators were initially recruited from outside the school district, where they had experienced different frameworks and training. Principals receive no training in how to recognize or discuss good teaching, only in faithful application of the rubric.

Schools are also provided with instructional coaches, whose job is to work with all the teachers at the school on their skills, with an emphasis on literacy and numeracy. There is little integration, however, between IMPACT and these professional development coaches except that the coaches see their role as helping teachers score high on IMPACT.

*Lack of collaboration and union opposition.* IMPACT is run unilaterally by the school district administration. After polling more than 500 of its members at a meeting in 2011, the Washington Teachers Union issued a scathing vote of no confidence in IMPACT. The union continues to file numerous grievances on behalf of teachers about IMPACT's implementation and results. Last year, Jason Kamras started to meet with the union president in problem-solving sessions every two weeks.

*Churn as the order of the day.* In tested grades and subjects, student scores on the district's standardized test (DC-CAS) account for 40 percent of a teacher's evaluation score. (This percentage was originally 55 percent, but it was reduced in 2012–13 as part of the first significant changes to IMPACT.) Teachers and principals may get bonuses or summarily lose their jobs on the basis of test performance. Examples have started to accumulate of teachers being fired even though students and colleagues considered them excellent teachers.

The school district seems to lowball enrollment projections and increase class size when budgeting, so schools often have to let "excess" teachers go in the spring and then hire other teachers back in the fall. According to Mary Levy (2012), an education finance expert who has studied D.C. Public Schools data for more than 30 years, in the low-income schools that constitute the system's great majority, it's not unusual for 40 percent of an individual school's staff to turn over in a single year and for more than 60 percent to turn over in two years. Staff churn has become the cultural norm, by design.

New research has documented the negative effect of teacher turnover and the importance of a collegial professional learning culture. High rates of turnover are bad for the professional culture in schools, and they particularly harm high-poverty students (Johnson, Kraft, & Papay, 2012; Ronfeldt, Loeb, & Wyckoff, 2011).

## Different Approaches, Different Results

On the surface, the Montgomery County and District of Columbia teacher evaluation systems seem to have similar components, including a similar-looking rubric to describe, in shorthand, what good teaching looks like. Both have expert teachers doing observations and rendering judgments. The District of Columbia's "Framework for Teaching and Learning" and its "Nine Commandments of Good Teaching" resemble Montgomery County's "Studying Skillful Teaching."

But in the D.C. Public Schools, IMPACT has suffered resistance and pushback. It has contributed to firing many teachers, but it seems to have done little to win over the workforce or to create a culture focused on a deep understanding of good teaching and learning. According to an analysis of the school district's payroll and personnel data conducted by Levy (2012), 21 percent of all teachers had left the system by the time IMPACT had been in place for two years. Half of newly hired teachers

left within their first two years, and 80 percent were gone by the end of their fifth year—far above the national average (Ingersoll, 2003).

Put bluntly, teachers are fleeing the D.C. Public Schools. Good teachers are leaving voluntarily at all-time high rates, citing the lack of a professional culture, collegial trust, and respect. Principals are also leaving the system at a rate of about 20 percent each year. Michelle Rhee brought in 90 principals, a majority of whom are now gone.

Montgomery County, on the other hand, has a new teacher turnover rate well below the national average. According to the most recent Montgomery County data (2012), 6.1 percent of new teachers leave after one year, 12.6 percent leave within their first two years, and just 29.9 percent leave within their first five years.

It is unclear what the constant churn, the ranking and rating, has accomplished in the D.C. Public Schools. Student achievement gains, as measured by standardized test scores, have been unimpressive. There was a slight bump up in 2009, tainted now by alleged widespread cheating in 103 schools that year (Strauss, 2012). In 2010 and 2011, scores dropped overall. Scores in 2012 were flat in reading; they were slightly up overall in math, but with no big gains. National Assessment of Educational Progress (NAEP) scores have improved from their previous abysmal levels, but the rate of improvement has actually slowed compared with gains that the district was making under previous superintendents Clifford Janey and Paul Vance, before Michelle Rhee came to town.

As Mary Levy and others have documented, the achievement gap by income and race has widened in the last five years by anywhere between 29 percent and 72 percent depending on the grade level and subject. There is little evidence that much has improved in the city's high-poverty, all-black schools. With hindsight, D.C.'s risky reform strategies feel like a missed opportunity to get it right.

Montgomery County, in contrast, has experienced consistent student achievement gains, particularly in high-poverty schools, and has narrowed the achievement gap by race and class, according to district data. It has created a professional culture focused on a sophisticated understanding of the craft of teaching. Throughout its 12 years, Montgomery's Professional Growth System has maintained the enthusiastic support of its workforce.

## A Model to Emulate

In spite of the vast differences in their approaches, Montgomery County and the District of Columbia are both part of a national focus on improving teacher quality through greater emphasis on teacher evaluation. Like Montgomery County, other districts—including Minneapolis, Minnesota, and Cincinnati, Columbus, and Toledo, Ohio—have also had decades of experience creating collaborative professional development systems. Hillsborough County, Florida, has recently jumped in with a promising, highly collaborative, multifaceted approach.

Many other school systems across the United States have taken an approach similar to that of the D.C. Public Schools, although perhaps not as extreme. Some, like Memphis, Tennessee, have modeled their system directly on IMPACT.

Ironically, education "reformers" more often tout the District of Columbia than Montgomery County as a model of teacher evaluation reform. Perhaps that's because they perceive IMPACT as disrupting the status quo, which is a good thing in its own right in the reformers' eyes. Over the past five years, more school systems in the United States have moved in the direction of fashioning teacher evaluation as a giant sorting mechanism whose purpose is to rank and rate teachers, bestow bonuses and other extrinsic benefits on the high flyers, and target the low scorers for remediation or dismissal.

That's a shame, because Montgomery County's approach and the principles that underlie it offer a proven alternative that can change the culture of education overall. Constructing a teacher evaluation system is an opportunity to signal many things to the workforce. Montgomery County's evaluation system is rigorous and holds teachers to high standards, but teachers feel they are part of a learning organization and their craft is respected. The teachers union is a partner with the district in improving schools. Teacher evaluation is done *with* teachers, not *to* them. Michael Winerip (2011) titled his glowing *New York Times* account of Montgomery County's PAR program "Helping Teachers Help Themselves." This is the kind of professional culture that America's best teachers seek.

Everything is not perfect in Montgomery County. Maintaining the right balances in the Peer Assistance and Review program is not easy. And solving problems related to the complexities of individual schools' cultures is a never-ending process.

But at least Montgomery County leaders have built a teacher evaluation system with the right purpose, based on sound principles that teachers respect. They have avoided the ranking and rating fixation aimed at firing or conferring bonus rewards on a few at the margins. Instead, their approach of nurturing good teaching skills and a learning culture among the entire workforce has reaped benefits worth recognizing and emulating.

## Effective Teacher Evaluation: Lessons from Experience

A comparison of Montgomery County Public Schools and the District of Columbia Public Schools yields the following guidelines for effective teacher evaluation:

**1 Collaboration.** The cornerstone of effective teacher evaluation is a collaborative environment in which the teachers union and the district codesign, coimplement, and coevaluate the evaluation system. If the evaluation process does not have credibility, if teachers don't value what they learn from it, or if they perceive it as unfair, it will fail.

**2 Professional culture.** Strong teacher voice helps establish a culture focused on good teaching. Teacher leadership and a professional, knowledge-based culture are preferable to a hierarchical system.

**3 Deep knowledge base in teaching.** Deep training of all evaluators and teachers communicates respect for the complexity of the craft, establishes a common language around good teaching, and creates credibility for decisions that affect individual teachers' careers. If the credibility of the system is not earned, it will foster cynicism.

**4 Integration with professional development and school culture.** Evaluation is most effective when it is integrated with other processes that support professional growth. The goal should not simply be to rank and rate teachers, but to create a healthy professional culture.

**5 Responsiveness to differentiated needs.** The evaluation process should be differentiated on the basis of what each teacher needs. One-size-fits-all processes waste time and add unnecessary expense.

**6 Reliance on intrinsic rewards.** In a professional culture, evaluation doesn't have to result in wasteful extrinsic rewards. The process can be its own reward. Daniel Pink<sup>1</sup> cautions that people who do work involving complex decision making and professional skill are motivated not by monetary rewards, but by autonomy, mastery, and purpose.

## References

- Curtis, R. (2011). *District of Columbia Public Schools: Defining instructional expectations and aligning accountability and support*. Washington, DC: Aspen Institute.
- Davis, L., Sylvia, K., & Simon, M., (2008, September 28). Bargaining for better teaching. *Washington Post*. Retrieved from [www.washingtonpost.com/wp-dyn/content/article/2008/09/26/AR2008092603350.html](http://www.washingtonpost.com/wp-dyn/content/article/2008/09/26/AR2008092603350.html)
- Ingersoll, R. (2003). [The wrong solution to the teacher shortage](#). *Educational Leadership*, 60(8), 30–33.
- Johnson, S. M., Kraft, M., & Papay, J. (2012). How context matters in high-need schools: The effects of teachers' working conditions on their professional satisfaction and their students' achievement. *Teachers College Record*, 114(10). Retrieved from [www.tcrecord.org/content.asp?contentid=16685](http://www.tcrecord.org/content.asp?contentid=16685)
- Levy, M. (2012, July 13). *Testimony before the Committee of the Whole*, D.C. Council.
- Montgomery County Public Schools, Employee and Retiree Services Center. (2012). *Staff statistical profile*. Rockville, MD: Author.
- Ronfeldt, M., Loeb, S., & Wyckoff, J. (2011). *How teacher turnover harms student achievement* (Working Paper No. 17176). Stanford, CA: Center for Education Policy Analysis.
- Strauss, V. (2012, August 11). The deafening silence on test cheating [blog post]. Retrieved from *The Answer Sheet* at [www.washingtonpost.com/blogs/answer-sheet/post/the-deafening-silence-on-test-cheating/2012/08/11/d4d565e2-e2fb-11e1-a25e-15067bb31849\\_blog.html](http://www.washingtonpost.com/blogs/answer-sheet/post/the-deafening-silence-on-test-cheating/2012/08/11/d4d565e2-e2fb-11e1-a25e-15067bb31849_blog.html)
- Winerip, M. (2011, June 6). Helping teachers help themselves. *New York Times*, p. A10.

## Endnote

<sup>1</sup> Pink, D. H. (2011). *Drive: The surprising truth about what motivates us*. New York: Riverhead.

[Mark Simon](#) is an education policy analyst at the Economic Policy Institute and national coordinator of the Mooney Institute for Teacher and Union Leadership in Washington, D.C.

Copyright © 2012 by ASCD