



# LESSONS *from* D.C.'S EVALUATION SYSTEM



**TEACHERS GIVE IMPACT LOW MARKS  
ON SUPPORT AND PROFESSIONAL  
DEVELOPMENT**

**By Susan Headden and Elena Silva**

**A**s a nation, we are changing the way we evaluate teachers, moving from a patchwork of weak and haphazard approaches to whole data-driven systems with dramatically high stakes. From Memphis to Chicago to Baltimore, districts and states are working to develop these systems, acknowledging the crucial role played by teachers and pushed toward greater accountability by competitive federal grant programs. At least 40 states have applied to the U.S. Department of Education for waivers that, in exchange for more flexibility on No Child Left Behind provisions, require comprehensive teacher evaluation systems (U.S. Department of Education, n.d.).

As they embark on this challenging task, districts can take some lessons from the successes and shortcomings

of an evaluation system that is often held up as a model: the IMPACT system in Washington, D.C. The controversial legacy of former schools chancellor Michelle Rhee, IMPACT sets clear expectations for instruction and holds teachers to well-defined standards of performance. Now into its third year, the program appears to be meeting its goals of rewarding effective teachers and eliminating educators it considers incompetent. And it has given the public reason to have more faith in its school system. But IMPACT, which ranks teachers on several measures, has earned plenty of criticism from teachers who say it is rigid and punitive and forces them to teach in an overly prescriptive way. More important, teachers say that in its rush to strengthen accountability, IMPACT misses what they say they need most — greater support and more meaningful professional development.

Like most public school systems, the District of Columbia Public Schools was badly in need of a new way to



ensure that it was putting a good teacher in every classroom. In 2007, when then-mayor Adrian Fenty took control of the city schools, the district's scores on the National Assessment of Educational Progress were among the lowest in the nation, and its black-white achievement gap was the largest of 11 urban districts that reported their results (National Center for Education Statistics, n.d.). And yet the district's evaluation system, which called for observations just once a year and graded teachers on a short checklist, rated 95% of teachers satisfactory or above.

### A NEW SYSTEM

The architects of the new system started with the basics: defining good teaching. The Teaching and Learning Framework, as the resulting document is called, was a way for principals, teachers, and administrators to work together to improve instruction. Instead of focusing on what to teach, they concentrated on how to teach, with specific directions that spanned subject areas. (See "Elements of good teaching" in box at right.) "We focused first on pedagogy, whereas most other reforms focused on curriculum," said Scott Thompson, director of teacher effectiveness strategy for DCPS. "You could have the greatest curriculum in the world, but if the teachers are ineffective in conveying it, then it's not going to matter."

Defining good teaching is one thing. Implementing an

### ELEMENTS OF GOOD TEACHING

The nine elements of the Teaching and Learning Framework form the essential rubric on which classroom performance is judged through IMPACT. They are:

1. Lead well-organized, objective-driven lessons.
2. Explain content clearly.
3. Engage students at all learning levels in rigorous work.
4. Provide students with multiple ways to engage with content.
5. Check for student understanding.
6. Respond to student misunderstandings.
7. Develop higher-level understanding through effective questioning.
8. Maximize instructional time.
9. Build a supportive, learning-focused classroom community.

Learn more about the development of the framework and lessons learned through that process in *District of Columbia Public Schools: Defining Instructional Expectations and Aligning Accountability and Support* (Curtis, 2011).

evaluation system around it, as D.C. and other school systems have found, is a far more complicated task. With input from teachers, administrators, and policy experts, D.C. produced a system that rates teachers on a combination of factors, some weighted more heavily than others. Classroom performance, as judged by the teaching and learning rubric, counts for 3%; student test scores (value-added data) for teachers in grades that take standardized tests count for 50%; commitment to the school community gets 10%;

and school value-added data — a measure of the school’s overall impact on student learning — is worth another 5%.

Because teachers in nontesting grades do not receive value-added data, their classroom performance counts for 75% of their score. A component called “teacher-assessed student achievement data” counts for 10%, and the other factors count the same as they do for the other teachers. For both categories of teachers, the final score is then adjusted up or down based on a factor called “core professionalism,” which covers things like coming to work on time.

The value-added measure has been as polarizing in D.C. as it has been elsewhere because it ties teacher performance to factors they say they often can’t control. And the scores have been undermined by reports of cheating by teachers and administrators on the tests on which it is substantially based (Gillum & Bello, 2011). But the classroom observations are just as controversial. Under IMPACT, every teacher in the district is observed five times a year: three times by a school administrator (usually the principal) and twice by a “master educator,” an outside teacher trained in the same discipline. The observations take 30 minutes, and all but one of the administrator visits are unannounced. Based on them, teachers are ranked from 1 to 4. Critics say that 30 minutes is too short a time for an

evaluator to assess performance, and that the assessment of that performance is subjective. The evaluations allow for virtually no input from teachers and provide no way for the instructor to put the lesson or her students in context.

Combined with other factors, all these ratings produce an overall IMPACT score that translates into highly effective, effective, minimally effective, or ineffective. A rating of ineffective means the teacher is immediately subject to dismissal; a rating of minimally effective gives the teacher one year to improve or be fired; effective wins the teacher a standard contract raise; a highly effective rating qualifies the teacher for a bonus of up to \$25,000.

At the end of IMPACT’s second year, roughly 17% of teachers were eligible for bonuses ranging from \$3,000 to \$25,000. With a second consecutive year of highly effective ratings, 7% were eligible to have a base salary increase of up to \$27,000. Six percent of teachers were fired — 2% who were rated ineffective and 4% who received minimally effective ratings for the second year in a row. Thus it can be said that IMPACT has served a purpose as a sorter, separating the good from the bad.

But could those highly effective teachers have gotten even better? Could the ineffective ones have been turned around? Most important, what about the teachers in the vast middle of the pack? How can the district help them become highly effective? And is the teacher evaluation system the right place to do it?

## TEACHER RESPONSE

IMPACT has three stated purposes: to outline clear performance expectations, to provide clear feedback, and to ensure that every teacher has a plan for getting better and receives guidance on how to do so. It is on this third component — professional development — that many teachers give IMPACT its own low grade.

In faculty lounges and in chat rooms, D.C. teachers trade IMPACT complaints — about ratings that vary from one evaluator to the next, or about master educators who didn’t seem to understand what they were doing or appreciate the challenges presented by their students. Bill Rope, who teaches 3rd grade at Hearst Elementary School, is one of many who argue that IMPACT reduces teaching to a formula. He has also experienced what he calls the system’s inconsistency. Rope was rated highly effective last year, but in a subsequent evaluation, a different master educator gave him an overall score of 2.78 — toward the low end of effective. Although she gave Rope 3s and 4s on “higher-level understanding” and “correcting student misunderstanding,” she rated him only minimally effective at “maximizing instructional time.” Rope was also downgraded for giving students only two ways to engage in content “when more would have been appropriate.” The evaluator also rated Rope only minimally effective at “engaging students at all learning levels in rigorous work.”

After Rope complained that several of these observations were misplaced, the master educator took the unusual step of adjusting two of his scores, giving him a higher overall rating. But Rope echoes the complaints of many of his fellow educators when he says the system narrows the curriculum. Last year, Rope says, he knocked himself out satisfying all of IMPACT’s demands. “I did everything you were supposed to do,” he said, “and I hated it.” In short, teachers say IMPACT demands that they essentially teach to their own test.

IMPACT’s architects argue that good teachers routinely demonstrate every element on the Teaching and Learning Framework without thinking. “It’s not as if this is a new way of teaching,” insists Thompson. “Good teachers get high marks for doing what they are already doing.” Administrators have also checked scores and found significant differences only in less than 1% of teacher observations. Likewise, the district has found that the scores given by principals and master educators have been remarkably similar: In only five out of 3,500 evaluations was there a gap of larger than two points between master educator and principal scores. To make sure that that everyone considers the same performance to be worth the same grade, the master educators and the principals “norm” the scores; they watch videos of teachers in action, role playing, and discussing what constitutes a 2, a 3, and so on.

The experiences of Rope and other teachers aside, master educators say they have been pleasantly surprised with how willing teachers have been to engage with the evaluators even when

**In faculty lounges and in chat rooms, D.C. teachers trade IMPACT complaints.**



they don't like what they have to say. Cynthia Robinson-Rivers, a master educator specializing in early childhood education, says, "We expected more hostility [to the feedback sessions] but usually they go just fine. I evaluated 230 teachers last year, and I can only name four or five who were hostile." Teachers who disagree with their observation scores can appeal, although they rarely do and only 15% of appeals last year were successful. With rare exceptions, the IMPACT team reports, teachers generally assess themselves the way the evaluators do. "When the class didn't go well, teachers know it didn't go well," Robinson-Rivers says.

### PROFESSIONAL DEVELOPMENT

IMPACT may be largely evaluative, but it would be incorrect to say that it is without a developmental function entirely. In particular, guidance on improving teaching comes through the post-observation feedback reports, which are thorough and specific. In a conference soon after the observation, the master educator explains the scores, then offers the teacher concrete advice on how he or she might improve. This is a distinct departure from past practice. Eric Bethel, a former elementary teacher who is now a master educator, told Education Sector he had never received instructional advice under the previous system, only a rating of "exceeds expectations." The master educator showed him, among other things, how he could use positive reinforcement to better control student behavior. "The observations allowed me to grow in very specific areas," he said.

The master educator can also serve to reinforce what the teacher is already doing, making a strong teacher even better. That was the case when master educator Matt Radigan informally observed Susan Haese, a 1st-grade teacher at Key Elementary School whom Radigan considered a 4. Afterward, he told her, "I want to celebrate what you did and repeat it." He then gave her very specific tips for building reading fluency, including having the students first read to themselves to build meaning, then read aloud as if they were on the radio. Radigan, a former instructional coach, says master educators often work with teachers after hours when they request it.

But as much as teachers appreciate this sort of feedback, it is not the same as formal development. And IMPACT very deliberately puts the development burden on the teacher — it is now up to the teachers to get themselves the help they need instead of making the principal responsible for providing it. "There is a shift," Thompson says. "Now we see the teacher as

taking a more active role."

Some teachers welcome this change. "For the first time in a very long time," one teacher is quoted as saying, "I finally felt like I was being respected as a professional, and that a process was put in place to hold me accountable for my performance. As a result, I began to become more interested in further developing my leadership skills as an educator" (Martinez, 2011).

### A BETTER WAY?

But others think a better system is in place just next door to D.C. in the highly regarded school system of Montgomery County, Md. The Teacher Professional Growth System, as it is intentionally named, specifically integrates evaluation, support, and development. Growth is the primary focus, evaluation the second.

A district with as many as 1,000 new teachers each year, Montgomery includes six standards for teacher performance, based on the National Board for Professional Teaching Standards, with criteria for how the standards are to be met (Montgomery County Public Schools, 2011). It provides training for evaluators and teachers; a professional growth cycle that integrates the formal evaluation year into a multiyear process of professional growth, with continual reflection on goals and progress meeting those goals; and formal evaluation with narrative assessments that provide qualitative feedback to teachers about their work. Under a Peer Assistance and Review program, consulting teachers — experienced educators who leave the classroom for three years — provide instructional support to new teachers and those not performing to standard. After these teachers have been given every opportunity and still don't improve, they can be dismissed. The district doesn't use test scores to judge teachers, but it uses them to inform discussions about instruction.

One big difference between the Montgomery County and DCPS systems is the size of the caseloads. Each of Montgomery's consulting teachers work with about 15 to 17 teachers, whereas D.C.'s master educators can have caseloads of up to 100. In and of itself, this sort of workload would seem to keep the master educators from focusing intensively on teachers who really need it — something Mark Simon believes is by design. Simon is a policy analyst at the Economic Policy Institute; he helped design Montgomery County's system when he was president of the Montgomery County Teachers Union. "It is not development or support work," he says. "It is simply rank-

### ABOUT EDUCATION SECTOR

Education Sector is a nonprofit, nonpartisan organization committed to achieving measurable impact in education policy by improving existing reform initiatives and by developing innovative solutions to the nation's most pressing education problems.

Learn more about the organization at [www.educationsector.org](http://www.educationsector.org) and access a range of reports and resources, including *Inside IMPACT: D.C.'s Model Teacher Evaluation System*, by Susan Headden (2011), which is closely related to this article.

ing and rating, with a narrative hastily written according to a rubric . . . . The goal (of IMPACT) was to ID the worst teachers for firing and the best for bonuses. It was not intended to improve teaching” (Education Sector, 2011).

The D.C. system is often described as one that was “done to” teachers instead of “done with” them. Teachers say they had little input in its design and not enough time to get used to it before it took effect. The union was not a partner in developing the system the way it was in Montgomery County (D.C. law precludes union involvement in negotiating evaluations). In the first year of IMPACT, 92 appeals were submitted to the chancellor’s office (10 were granted). Last year, the chancellor’s office received 260 appeals, all of which are still pending. According to Simon, in Montgomery, nearly 500 teachers have been removed for performance reasons over 10 years, but that there are very few appeals or challenges. “There is a sense of legitimacy in the process,” he says.

### IMPACT NOW

In important respects, comparing D.C.’s school system to Montgomery County’s is an apples-and-oranges exercise, and an unfair one at that. Montgomery is a far more affluent county, and while it struggles with racial and socioeconomic achievement gaps, its students overall have consistently met a high standard of academic performance. D.C., by contrast, is urban, largely low-income, and predominantly black, with a history of dismal academic performance. IMPACT was a dramatic and necessary response to what could only be called a desperate situation for D.C.’s public education system and the children and families it serves. “If teachers are anxious because they have low scores, I empathize,” says Jason Kamras, chief of DCPS’s Office of Human Capital Management, “but at the end of the day, we have to hold the line on quality. I believe with every fiber of my being that we can’t have different standards for other people’s children than we have for our own.”

Yet Kamras and other district officials concede that IMPACT isn’t perfect, and they have listened to teachers, principals, and coaches and made changes in response. D.C.’s big push this year is connecting evaluation to development, as well as providing teachers with better academic and curricular support. Among other tools, the district is producing an online video library it calls “Reality P.D.” — clips of DCPS teachers demonstrating various aspects of the rubric and sharing tips. The district is also using data generated by IMPACT to improve instruction. In the first year, teachers consistently scored lowest on measures of rigor and probing for higher-level understanding. That finding led the district to further clarify and emphasize these skills in the revised framework and in professional development. The information leads to improvements at individual schools as well.

Thompson says the district is planning to take what he calls a “long top-to-bottom look” at IMPACT to determine

what significant revisions to the system, if any, are necessary. The review will include taking stock of evaluation designs and implementation lessons in other districts and states. So far, district officials say they are seeing a lot of similarities between IMPACT and other emerging evaluation systems: Versions of the teaching and learning rubric are being adapted and adopted in Chicago, Memphis, Baltimore, and Indiana.

As IMPACT passes through its third year, Kamras says he’s sensitive to the anxiety that teachers feel about the system but remains steadfast in his focus on accountability. “There is still a perception that IMPACT is a ‘gotcha,’ ” he says. “But I think the big thing has been getting over the hump. We went from zero accountability right to 100% accountability. So without changing the fundamentals, I want to reduce the anxiety level.” To hear teachers talk, he will have a better chance of succeeding if IMPACT’s focus on accountability is matched by professional support and development that is truly integral to the system.

### REFERENCES

**Curtis, R. (2011, March).** *District of Columbia Public Schools: Defining instructional expectations and aligning accountability and support.* Washington, DC: The Aspen Institute.

**Education Sector. (2011, August).** Excerpt from *Teacher evaluation: How observation fits in.* Available at [www.educationsector.org/events/online-discussion-teacher-evaluation-how-observation-fits](http://www.educationsector.org/events/online-discussion-teacher-evaluation-how-observation-fits).

**Gillum, J. & Bello, M. (2011, March 30).** When standardized test scores soared in D.C., were the gains real? *USA Today.* Available at [www.usatoday.com/news/education/2011-03-28-1Aschooltesting28\\_CV\\_N.htm](http://www.usatoday.com/news/education/2011-03-28-1Aschooltesting28_CV_N.htm).

**Martinez, B. (2011, October 17).** *IMPACT in Washington D.C.: Lessons from the first years.* Washington, DC: Democrats for Education Reform.

**Montgomery County Public Schools. (2011, August 20).** *Teacher professional growth system handbook.* Rockville, MD: Author. Available at [www.montgomeryschoolsmd.org/departments/development/documents/TeacherPGS\\_handbook.pdf](http://www.montgomeryschoolsmd.org/departments/development/documents/TeacherPGS_handbook.pdf).

**National Center for Education. (n.d.).** *National Assessment of Educational Progress (NAEP) state profiles.* Washington, DC: Author. Available at <http://nces.ed.gov/nationsreportcard/states>.

**U.S. Department of Education. (n.d.).** *ESEA flexibility.* Washington, DC: Author. Available at [www.ed.gov/esea/flexibility](http://www.ed.gov/esea/flexibility).

•

**Susan Headden (sheadden@educationsector.org) is senior writer/editor and Elena Silva (esilva@educationsector.org) is senior policy analyst at Education Sector, an education policy think tank. ■**