

THE CENTURY FOUNDATION

Issue Brief

MICHELLE RHEE'S SPLENDID RUN AT THE WRONG TARGET

GORDON A. MACINNES

Michelle Rhee is the face of school reform, America's first celebrity school superintendent. She burst onto the scene in June 2007 as chancellor of District of Columbia public schools, and was gone from the position and the city by Halloween 2010—but not before she adorned the cover of *Time*, appeared on the *Oprah Winfrey Show*, *Charlie Rose*, and other high-profile shows. She is energetic, focused, smart, outspoken, and combative. After stepping down as chancellor, she announced the launch of Students First, a web-based crusade she has started to raise \$1 billion to wage war on “vested interest groups and bureaucratic structures” that “put the interests of adults over the needs of children.”¹

The message that Rhee has been delivering mirrors the call for reforms pushed by the Gates Foundation, Arne Duncan, Joel Klein, and think tanks such as the American Enterprise Institute: the most important step to improving education for America's poor children is to dismiss ineffective teachers and replace them with more effective teachers. Do the same thing with principals and the game is pretty much over, they argue. Yes, to a large segment of the reform movement, it is that simple (as long as there are also plenty of charter schools).

WASHINGTON, D.C.: THE SPECIAL OPPORTUNITY

With Rhee's appointment as chancellor, Washington, D.C., became the Mecca for education reform. At least four conditions made it the perfect laboratory for making big changes quickly. Other city superintendents must have looked on with envy as the Rhee era unfolded.

First, Rhee was appointed by a newly elected mayor committed to reforming the public schools. Mayor Adrian Fenty consolidated his control of public education with legislation in 2007 that created the Office of Chancellor, which did not have to contend with a local or state school board: the mayor and chancellor run everything. A second law that passed in December 2007 removed more than nine hundred central office employees from civil service protection and made them “at-will” employees. A third law from the 1990s grants the mayor appointive power over the board that grants charters and oversees charter schools, of which there are now sixty. Throughout it all, the mayor and chancellor struck an unusually strong public partnership. No other city superintendent enjoys as much formal authority, and few have enjoyed such unqualified political support.

Second, the District of Columbia school system in 2007 was a more manageable scale as a result of charters and demographic shifts. In just the six years between 2003 and 2009, enrollment fell from 65,000

The Century Foundation conducts public policy research and analyses of economic, social, and foreign policy issues, including inequality, retirement security, election reform, media studies, homeland security, and international affairs. The foundation produces books, reports, and other publications, convenes task forces and working groups, and operates eight informational Web sites. With offices in New York City and Washington, D.C., The Century Foundation is nonprofit and nonpartisan and was founded in 1919 by Edward A. Filene.

HEADQUARTERS: 41 EAST 70TH STREET – NEW YORK, NY 10021 – 212.535.4441 – 212.535.7534 (FAX) – INFO@TCF.ORG – WWW.TCF.ORG
D.C. OFFICE: 1333 H STREET, NW – 10TH FLOOR – WASHINGTON, D.C. 20005 – 202.387.0400 – 202.483.9430 (FAX) – INFO@TCF.ORG – WWW.TCF.ORG

to 45,000, while charter school enrollment exploded. By the time Rhee resigned, charter schools enrolled 38 percent of all public school students, the highest market share in the country, except for New Orleans.

Third, the District of Columbia public schools are well financed. A 2009 federal report on the one hundred largest districts in the nation found only two that spent more than the District of Columbia the year Rhee assumed the chancellorship (New York and Boston). The District of Columbia spent a healthy \$20,029, compared to other high-poverty, majority-black cities such as Baltimore (\$14,591); Detroit (\$13,066); and Atlanta (\$14,186).²

Fourth, Rhee inherited a system that already had shifted its emphasis to improved academics. As measured by National Assessment of Education Progress (NAEP), District of Columbia public school students were making significant progress before Rhee's appointment. In the six years between 2003 and 2009, nine of the fourteen points in improved scale scores on fourth grade math and fourth grade reading tests were achieved before her arrival. These included the largest single-year gains of six points on both tests.³

In short, when Rhee arrived on the scene, the legal, political, organizational, financial, and educational conditions were aligned for significant change.

While it is premature to judge Rhee's brief tenure, it is fair to ask: Given these special circumstances, did she make the best bet for the students?

THE ASSUMPTIONS OF THE RHEE STRATEGY

Let us stipulate that the academic prospects of children are harmed when they are placed in classrooms with lazy, incompetent, or indifferent teachers, and that the prevailing system for evaluating teachers does not work well. It is also true that the system for preparing, inducting, and supporting teachers is just as broken. That said, there is a big difference between acknowledging teacher evaluation policies and practices as a problem, and making those problems the centerpiece of the effort to improve education. Furthermore, teacher performance is only part of the problem in failing schools—the socioeconomic background of their parents and their classmates are much more consequential factors that affect the academic performance of individual students.

Rather than act on the decades of indisputable findings on the factors contributing to the poor performance of District of Columbia public schools, Chancellor Rhee assumed that ineffectual teaching was the best explanation for the achievement gap between students in schools filled with children from poor families and their counterparts in wealthier districts. In fact, she spent most of her forty-one month tenure talking about bad teachers and their union, which may explain why it took three years to reach a contract with the Washington Teachers Union (WTU).

IMPACT: THE RHEE LEGACY

The heart of Rhee's effort was IMPACT (not an acronym), a highly detailed framework she implemented—without WTU support—to evaluate every teacher annually. IMPACT outlines the characteristics of effective teaching, such as clearly explaining the purpose and importance of each lesson, or quickly correcting student “misunderstandings.” Such a catalogue of good practice is to be welcomed, as many districts lack such clarity. IMPACT was implemented to spotlight two categories of teachers: those who should receive bonuses and those who should be dismissed.

IMPACT uses two main measures to sort and rate teachers. First, for teachers of reading and math in grades four through eight, 50 percent of their evaluation rides on the results on the district's standardized tests, employing the "value-added method" (VAM). VAM uses the results from the prior year's test as the baseline for predicting gains in the subsequent year. For the teachers subject to VAM, the quality of their instruction is judged by whether students do better than predicted, or noticeably worse. VAM attributes student progress or its absence entirely to the classroom teacher.

The second measure is observations of all teachers by a building administrator (thrice annually) and an independent teaching expert called a "master educator" (twice). The master educators are experienced teachers, expert in at least one content area. The observations are followed by a teacher-observer conference. For most teachers (that is, those not subject to VAM), 75 to 90 percent of their evaluation is based on these thirty-minute observations. Teachers subject to VAM also are observed five times, but it accounts for only 35 percent of their score.

IMPACT sets four gradations of teacher quality: highly effective, effective, minimally effective, ineffective. "Ineffective" teachers are subject to immediate dismissal, as are teachers scored "minimally effective" two years in a row.

WHAT IMPACT ATTEMPTS TO MEASURE

The No Child Left Behind Act (NCLB) in 2001 mandated that each state adopt core curricular standards and administer standardized tests in grades three through eight and one high school grade, and since then an assembly line has been built out of these mandates. On the surface, the idea of core standards is sensible: everyone involved in education should know what is expected of them academically. The standards advocates thought that states would strive to align their standards to those of the highest-achieving international competitors, such as Singapore or Korea. Instead, most states opted for squishy, vague, multitudinous, committee-generated mush. And many states manipulated their tests to minimize the number of schools found failing under NCLB.

The District of Columbia's "learning standards" set down what every student needs to learn in each grade and subject.⁴ As with many state curricular standards, those of the District of Columbia range from quite general (fourth grade social studies: "Explain how political, religious, and economic ideas and interests brought about the Revolution") to very specific ("List and interpret the songs that express American ideals [e.g., "America the Beautiful". . .]). Some standards include examples of student assignments and instructional materials that could be used. Many do not.

One problem with using IMPACT to judge success is that it puts the onus solely on the teacher, assuming that every teacher will know how to translate the district's learning standards into daily teachable chunks of skill and knowledge. IMPACT assumes that the central office and school administrators have equipped each teacher with the instructional materials, training, classroom support, and usable data to manage this translation. A fourth grade teacher, for example, must contend with 234 individual standards covering language arts, math, social studies, and science. There is little guidance as to the most effective sequence or any sense of how much time should be allotted to any particular content or skills grouping.

Consider just one standard in fourth grade social studies: "Students compare and contrast 15th through 18th century America and the United States of the 21st century with respect to population, settlement patterns, resource use, transportation systems, human livelihoods, and economic activity." This would be a daunting enough assignment for a collegian majoring in history, never mind a fourth grader. There are 233 other, presumably equal, standards.

This fourth grade teacher must prepare lesson plans for reading, writing, math, social studies, and science that will clearly describe the morsels of curriculum to be devoured each day and how to make it engaging to students. Moreover, since children do not learn at the same pace, the teacher is expected to specify how to deal individually or in small groups with those students who are struggling (or galloping ahead of their classmates). In each of the four subjects, the teacher also must produce interim assessments that can be used to measure progress; the model for these assessments must be approved by the principal at the beginning of the year.

The measure of how well any school, classroom, or district is doing is, of course, the NCLB-mandated standardized test, which evaluate student mastery of a tiny fraction of those 234 standards for fourth graders.

THE PROBLEM WITH IMPACT'S USE OF VAM

VAM represents a big improvement over other popular measures of student achievement and teacher effectiveness. It introduces the helpful idea of measuring *improvement*, acknowledging that some students started the year ahead of others.

The problem comes when VAM results are used to make important professional and personnel decisions about individual teachers. Here, VAM clashes with the realities of everyday classroom life in a number of ways:

- *Universal attribution of results to one teacher.* How, for example, would VAM measure the success of a sixth grade English teacher who sees an improvement in writing skills in her students in part because the sixth grade science teacher assigns frequent essays? VAM cannot reflect that contribution in scoring the sixth grade English teacher for improved writing results. Attribution is especially tricky when special education or English-learning students are in a general classroom, but receiving regular, supplementary help from specialist teachers.
- *Nonrandom distribution of students and teachers.* In the real world, parents, principals, classroom availability, tradition, and seniority all play a role in assigning students to particular teachers. In the same way, teachers are not subject to random assignment, which is why low-performing schools frequently have the most inexperienced teachers. However, VAM assumes randomness.
- *Small sample sizes, particularly in elementary grades where one teacher teaches all subjects.* The effect on a teacher's rating can be disproportionately influenced by just a few students whose performance may be affected by, say, the loss of a parent, divorce or separation, or frequent absences. The small sample of a classroom of eighteen or twenty can be further reduced if students move in from another state or country or did not take last year's test because of their English-learning status.
- *Student mobility, which tends to be highest in the lowest-performing schools.* It matters if a third or more of the class taking the May exam transferred in between October and the test.
- *Unequal opportunity since last year's test.* Some students spent the summer hanging out, while others may have attended summer school, an academic summer camp, or been tutored frequently.
- *Standardized tests are designed to measure student ability, not teacher effectiveness.* Just as SAT scores are designed to judge a student's ability to handle first-year college academics, not the effectiveness of his high school, so one should use state tests to measure student mastery of some state standards, not teacher effectiveness.

If this is not enough, consider the overwhelming consensus among scholars, psychometricians, and researchers that, while VAM deserves further testing as a promising methodology, it cannot be used fairly and reliably as the basis for decisions concerning tenure, promotion, or dismissal.

Because of small sample sizes and the nonrandom distributions of both teachers and students, the propensity for error when using VAM is very high. A Mathematica study found that, using only one year's test data, 36 percent of teachers are likely to be misclassified because of the imprecision in test data.⁵ A teacher found "inefficient," therefore, has a one-in-three chance of actually being "effective" or "highly effective." Mathematica warned that it probably understated the error rate because the nonrandom distribution of students was not included in its analysis, just the instability of tests results one year to the next. (Incidentally, Mathematica is the contractor that did the technical work to set up IMPACT.)

A series of studies from the RAND Corporation found that the rating of teachers varies greatly from year to year when using VAM.⁶ For example, only one-third of teachers placed in the top 20 percent in the first year were judged in the top 20 percent in the second year. Much can change from the content of tests, the distribution of high-need students, and classroom mobility rates.

The Board of Testing and Evaluation of the National Research Council cautioned the U.S. Department of Education that "VAM estimates of teacher effectiveness should not be used to make operational decisions because such estimates are far too unstable to be considered fair or reliable."⁷

THE PROBLEM WITH IMPACT'S CLASSROOM OBSERVATIONS

The District of Columbia employs about thirty-five experienced teachers who have been trained to observe classroom instruction and to confer with teachers about their findings. This is an important innovation. In most districts, classroom observation is reduced to a checklist affair of little utility to the teacher or school.

The problem with IMPACT is not the use of trained observers, but the conflict in their role. The requirement that the master educator confer with the teacher within fifteen days implies that the primary purpose is to help the teacher improve her or his instructional practice. However, the conference also could be to emphasize the seriousness of the situation for teachers who are observed to be "inefficient." If the latter, the master educator is not expected or equipped to assist the teacher.

Principals presumably can use their formal observations to organize assistance for a struggling teacher. The principal may be adept at working with such teachers, or be able to assign an instructional coach or assistant principal to the task. However, not all principals are well trained to evaluate instructional practice. The coursework prescribed for principal certification includes very little clinical experience applying principles of instruction to classroom reviews.

If the classroom reviews reveal significant issues with a teacher's performance, the path to improvement will require much in-class support. The principal's role as evaluator and helper collide, particularly since the principal's own evaluation includes the efficacy of her/his teacher support efforts. This potential conflict may help explain why only 4 percent of teachers were found to be "inefficient" in IMPACT's first year while 16 percent were judged "highly effective."

IMPACT DRIVES A WEDGE

Michelle Rhee is famously quoted as saying "collaboration and consensus-building are over-rated." Those sentiments are reflected in the rubrics for IMPACT—only 2 percent of a teacher's rating measures

collaboration with colleagues on instructional issues, while 75 percent is based on classroom observation against criteria that do not even mention collaboration.

Michelle Rhee is dead wrong. Collaboration and consensus-building are at the heart of every effective school and district. The notion that struggling schools and districts can be improved in an atmosphere of distrust, public accusations of incompetence, and laziness is preposterous.

Recent studies by economists and educators have documented the positive impact of peer-to-peer collaboration and learning among teachers in high-poverty schools.⁸

The school districts of Union City and West New York (in New Jersey) and Montgomery County (Maryland) may be the first where concentrations of students from poor families perform nearly as well academically as students in middle-class districts through eighth grade. These districts assume that the likelihood of a classroom teacher being able to figure out unaided how to teach effectively to every student is ridiculously low. Therefore, great weight is put on teachers collaborating to solve difficult pedagogical puzzles.

Union City is one of the poorest cities in the United States and has one of the highest proportions of foreign-born residents. Almost 90 percent of its students are eligible for free or reduced-price lunch, and 75 percent come from non-English-speaking families. Yet, Union City's eighth graders perform very close to the state average in math and literacy (in a state whose eighth graders perform second-best on the NAEP). In addition to having a very clear and teacher-friendly district curriculum, district-developed interim assessments, a robust student data system that produces timely information for teachers and principals, Union City practices teacher collaboration.

An example: the fourth grade teachers at a Union City school agree that Ms. Garcia is much more confident teaching fractions and whole numbers than any of her three colleagues, but Ms. Jones is strongest at grammatical exercises. Therefore, Ms. Garcia will receive Ms. Jones' class for ninety minutes of fractions instruction, while Ms. Garcia's fourth graders work with Ms. Jones on the semicolon and adverb. Not only is IMPACT unable to evaluate fairly such practice, its incentive structure actively discourages such powerful and effective instruction.

RHEE'S MISTAKE

It is hard to imagine a more difficult professional job than teaching in a failed school in a very poor neighborhood. Improving instruction in such places begins with recognizing this truth and acting on it.

Instead of highlighting the nature, causes, and magnitude of the educational problems of District of Columbia schools, Chancellor Rhee chose to focus only on the teachers, particularly the bad ones, and their union. This was not a tactic calculated to build the trust of rank-and-file teachers or the leadership of the WTU. Since there is no consistent and accepted research that supports the claim that the most important step that can be taken to improve American education is to dismiss bad teachers, the Rhee strategy and legacy are in doubt.

The shame of it all is that the Rhee record contains some important achievements: she performed a public service in reducing the bloat in the central office, consolidating Head Start with the District of Columbia preschool program, improving the data systems, and adopting clearer curricular and teacher evaluation standards. However, the spotlight on supposed teacher incompetence and laziness permitted little light to shine on the more productive opportunities opened up by Washington's special circumstances in what turned out to be Mayor Fenty's only term and Michelle Rhee's short term as chancellor.

JANUARY 11, 2011

THE CENTURY FOUNDATION IS PUBLISHING THE ISSUE BRIEF SERIES TO HELP EXPLAIN AND CALL ATTENTION TO PUBLIC POLICY IDEAS THAT ARE WORTHY OF DISCUSSION AND DEBATE. THE VIEWS EXPRESSED IN THIS SERIES ARE SOLELY THOSE OF THE AUTHORS OF EACH ARTICLE.

THIS AND OTHER PUBLICATIONS CAN BE FOUND ON THE CENTURY FOUNDATION WEB SITE: www.tcf.org

¹ See the Students First website, <http://www.studentsfirst.org/#>.

² Chris Plotts and Jennifer Sable, *Characteristics of the 100 Largest Public Elementary and Secondary School Districts in the United States: 2007–08*, NCES 2010-349 (Washington, D.C.: National Center for Education Statistics, 2010), A-28-9.

³ *The Nation's Report Card: Reading 2009, National Assessment of Educational Progress at Grades 4 and 8*, NCES 2010-458 (Washington, D.C.: National Center for Education Statistics, 2009), 16.

⁴ “Learning Standards for Grades Pre-K–8,” District of Columbia Public Schools, <http://dcps.dc.gov/DCPS/In+the+Classroom/What+Students+Are+Learning/Learning+Standards+for+Grades+Pre-K-8>.

⁵ Peter Z. Schochet and Hanley S. Chiang, *Error Rates in Measuring Teacher and School Performance Based on Student Test Score Gains*, NCEE 2010-4004 (Washington, D.C.: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education, 2010), <http://ies.ed.gov/ncee/pubs/20104004/pdf/20104004.pdf>.

⁶ Cory Koedel and Julian R. Betts, *Re-Examining the Role of Teacher Quality in the Educational Production Function*, Working Paper #2007-03 (Nashville, Tenn.: National Center on Performance Initiatives, 2007), http://economics.missouri.edu/working-papers/2007/wp0708_koedel.pdf; Daniel F. McCaffrey, Daniel Koretz, J. R. Lockwood, and Laura S. Hamilton, *Evaluating Value-Added Models for Teacher Accountability* (Santa Monica, Calif.: RAND Corporation, 2003), http://www.rand.org/pubs/monographs/2004/RAND_MG158.pdf.

⁷ Board on Testing and Assessment, “Letter Report to the U.S. Department of Education on Race to the Top Fund,” National Research Council, 2009, 10, http://www.nap.edu/catalog.php?record_id=12780.

⁸ If interested, see Clement Kirabo Jackson and Elias Bruegmann, “Teaching Students and Teaching Each Other: The Importance of Peer Learning for Teachers,” *American Economic Journal: Applied Economics* 1, no. 4 (2009): 85–108; Yvonne L. Goddard, Roger D. Goddard, and Megan Tschannen-Moran, “A Theoretical and Empirical Investigation of Teacher Collaboration for School Improvement and Student Achievement in Public Elementary Schools,” *Teachers College Record* 109, no. 4 (2007): 877–896.