Can Evaluating Teachers Improve Teacher Effectiveness?

IMPACT Evaluations in Washington, DC

By Tamara Ortega Goodspeed and T'Nia Crutchfield

Research shows that effective teachers can have a profound impact on student success both inside and outside of the classroom. At the same time, poor teaching holds students back, with deficits that compound over time. So it is vital that schools and schools systems be able to distinguish between effective and ineffective teachers and help all teachers improve.

However, efforts to assess teacher performance based on student learning often meet with strong resistance, especially when the assessments are tied to decisions about pay or job security. Critics argue that factors beyond teachers' control affect student learning, that high-stakes teacher evaluations don't measure what matters most and are prone to errors (or to cheating), and that in the end they do not help teachers improve. Proponents argue that without teacher evaluations linked to learning, it is impossible to identify problems or help and hold teachers accountable for improving instruction.

Can evaluating teachers based on student performance improve teacher effectiveness? What are the risks and rewards of such a system?

Since 2009, public school teachers in Washington, DC, have been evaluated using a high-stakes, incentive-based evaluation system called IMPACT that rates teachers on the basis of multiple classroom observations, their contribution to students' learning as measured by achievement tests, contributions to the school community, and core professionalism. IMPACT lays out clear expectations, measures performance toward those expectations, and provides feedback to teachers along with ideas for growth. Highly effective teachers are rewarded, and ineffective teachers are removed from the classroom. Initial results suggest that the system has encouraged lower performing teachers to either improve or resign, and helped retain higher performing teachers and improve their performance. The teacher evaluations are part of

PREAL Policy Briefs provide nontechnical overviews of key education policy topics. In this issue, Inter-American Dialogue Senior Education Associate Tamara Ortega Goodspeed and Dialogue intern T'Nia Crutchfield focus on Washington, DC's IMPACT teacher evaluation system, its key features, successes and challenges, and potential implications for teacher policy in Latin America.

Additional information on IMPACT can be found on the **DCPS website** and at the web links referenced throughout the document.

This brief was made possible by support from the Inter-American Development Bank's Institutional Capacity Strengthening Thematic Fund (ICSF), established with contributions from the Government of the People's Republic of China. However, the contents are the sole responsibility of the authors and the Inter-American Dialogue.

a broader set of DC education reforms that have contributed to student achievement gains.

This brief highlights key features of the DC teacher evaluation system, summarizes its successes and challenges, and draws potential implications for teacher policy in Latin America.

> "Initial results suggest that the system has encouraged lower performing teachers to either improve or resign, and helped retain higher performing teachers and improve their performance."

A School System Confronting Low Levels of Student Learning

The Washington, DC school district (DCPS) has long been among the lowest performing US districts on national achievement tests. DCPS serves over 45,000 students in more than 100 schools. Only 12% of students are white, and more than three-quarters come from families with income levels low enough to qualify for free or reduced-cost school meals.[1] Although test scores have improved over the last several years, less than 30% of students in the fourth grade scored proficient or better in math and less than a quarter scored at that level in reading on the 2013 National Assessment of Educational Progress (NAEP).[2] By contrast, more than 40% of public school students nationally scored proficient or higher in 4th grade math, while 34% did so in reading. The percentage of DC eighth graders scoring proficient or better was even lower. While these scores are alarming, they represent an improvement over 2007, when only 14% of DC fourth graders scored proficient or better in reading and math.

In 2007, then-DCPS Chancellor Michelle Rhee began a major overhaul of the school system designed to address chronic low achievement by closing low-performing schools and hiring new principals. She also rolled out a plan to improve the quality of instruction based on a new Teaching and Learning Framework (TLF), which defines the behaviors expected of all teachers in three areas—planning, teaching, and increasing effectiveness. Teachers are currently evaluated only on the "Teach" standards, which address nine key pedagogical practices:[3]

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

In line with the TLF, Rhee instituted teacher evaluations that linked performance (measured primarily through classroom observations and student test scores) to pay and job security. Instituted in 2009, the annual evaluations, called IMPACT, continue, with some adjustments, today. As one of the first evaluation systems in the United States to use this approach, results are being closely watched. How does it work?

The nuts and bolts of IMPACT

IMPACT is a high-stakes, incentive-based evaluation system that

rates teachers on their performance in key areas including classroom practice and student learning. Teachers are categorized into seven different groups, with Groups 1 and 2 covering the majority of general education teachers in grades 1 through 12.[4] Group 1 includes English language arts teachers in grades four through ten and math teachers in grades four through eight whose students take the DC Comprehensive Assessment System (DC CAS)—DCPS' annual assessment of student knowledge based on the Common Core Standards[5]—each year and therefore generate value-added student achievement data that can be associated with a particular teacher. Currently, about 15% of DC public school teachers fall into this category. Group 2 includes all general education teachers, grades one through twelve, whose students do not take a DC CAS test in their subject and therefore cannot generate value-added student achievement data.

Group 1 and 2 teachers are evaluated in four areas:

- Observations of classroom practice based on the TLF "Teach" behaviors;
- Value-added student achievement scores based on DC CAS (Group 1) and/or improvements in student achievement based on other rigorous tests (Groups 1 & 2);
- Commitment to the school community (CSC); and
- Core professionalism (CP).

Each component has a different weight, with classroom practice receiving the highest weight and CSC the lowest. Core professionalism is scored only as a 10 to 20 point deduction from the overall IMPACT score and does not receive a specific weight. A brief description of each component follows:

Classroom Practice

The "Teach" standards from the Teaching and Learning Framework provide a guide for measuring instructional expertise via multiple classroom observations. TLF is DCPS' definition of what an effective teacher does; and scores based on the TLF "Teach" standards count for 40 percent of the overall IMPACT rating for teachers in Group 1, and 75 percent of the overall impact rating for teachers in Group 2 (whose IMPACT score does not include individual value-added data). The TLF score is based on four formal classroom observations a year: two by an administrator (principal or assistant principal) and two by a "master educator," an independent expert practitioner who conducts observations and provides feedback and guidance for growth in post-observation conferences and written reports.[6] Beginning teachers and those who are lower on the career ladder also receive additional support through informal administrator observations, which do not count toward IMPACT scores. Teachers at higher levels of the career ladder who demonstrate consistently high IMPACT scores may be observed less than four times.[7]

Student Achievement Data

Teachers in both Groups 1 and 2 are measured on gains in student learning during the school year. For Group 1 teachers, this component comprises 50 percent of their IMPACT score and includes individual value-added scores (IVA) based on students' performance on DC CAS math and reading tests (35 percent), and teacher-assessed student achievement data (TAS), measured by assessments other than the DC CAS (15 percent). Group 2 teachers are assessed only on TAS (15 percent) for this component.

To calculate IVA, the district estimates the average impact a teacher will have on student learning, based on changes in individual student scores from one year to the next, and compares it to actual learning. An independent national research firm isolates teachers' contributions from other factors that can affect student learning by taking into account factors such as students' DC CAS score from the previous year; eligibility for free or reduced-cost lunch; special education status; limited English proficiency status; attendance from the previous year; whether the student transferred schools midyear; average test scores from the students' class the previous year; and the extent of the variation in scores from the students' class the previous year. The firm then aggregates individual student scores for eligible teachers according to the subject they teach, which students they taught, and how much time they

spent with each student to determine an "average likely score" for each teacher's class.[8]

The average likely score is then compared with the average actual score of the students in a teacher's class at the end of the year to calculate value-added, using the following formula:

Actual DC CAS Score
\- **Likely** DC CAS Score
Value-Added

> "DCPS remains one of the few school districts in the United States where teachers judged ineffective are quickly removed from the classroom."

Teachers must have at least 15 students with DC CAS scores from both the previous and current year in order to receive an IVA score. IVA scores are calculated separately for reading and math, and teachers who teach both subjects received an average of the two scores.

TAS measures student learning based on rigorous assessments other than DC CAS that are aligned with DCPS content standards. Teachers and administrators agree on which assessment(s) to use, how to weight them, and achievement targets at the beginning of the school year and compare them to actual student achievement scores at the end of the year. Teachers are then assigned an IMPACT score on a scale of 1 to 4, based on the defined achievement targets. Although the scale varies according the assessment used and specific school circumstances, DCPS guidebooks for Group 1 and 2 teachers recommend that in general, level 1 represent average learning growth among students of less than one year or less than 70% mastery of content standards, and that level four represent exceptional learning (1.5 years of growth or more or at least 90% mastery of content standards).[9] Scores for TAS must be approved and validated by the administration, and progress toward goals is monitored throughout the school year to allow for data-based instructional adjustments.

Commitment to the School Community (CSC) & Core Professionalism (CP)

CSC counts for 10 percent of the IMPACT score for all teacher groups, and is based on administrator assessments that measure support of school initiatives; collaboration with students, families, and colleagues; and efforts to promote high expectations. Core professionalism is assessed by principals, and measures teachers' adherence to four tenets of professionalism: no unexcused absences; no unexcused late arrivals; the degree to which they follow school and school system policies and procedures; and respectful interaction with colleagues, students, families, and community members. There are only three rankings for Core Professionalism: meets standard, slightly below standard, and significantly below standard. CP does not receive a separate weight and can only be subtracted from a teacher's overall IMPACT score, reducing it by ten points for teachers receiving a "slightly below standard" rating on any of the four criteria or by 20 points for those receiving a "significantly below standard" rating on any criteria.

Rating Teacher Effectiveness and Assigning Consequences

When scores from all components are compiled, the weighted average, less any deductions for core professionalism, establishes a teacher's overall IMPACT score. Scores range from 100 to 400, and teachers are placed into one of five performance categories: Ineffective (scores below 200); Minimally Effective (scores from 200–249); Developing (scores from 250-299); Effective (scores from 300–349); and Highly Effective (scores of 350 or above).

Highly effective teachers receive increased recognition and are eligible for career advancement, expanded leadership opportunities, and pay increases of up to $25,000 for Group 1 teachers working in high poverty and traditionally low performing schools.[10] Group 1 teachers receive larger bonuses than Group 2 teachers in recognition of the increased challenges and rigor of having 50% of their IMPACT scores based on student achievement data.

Ineffective teachers are subject to immediate dismissal, as are teachers ranked minimally effective two years in a row and teachers ranked as developing three years in a row. Teachers ranked as developing or minimally effective are also ineligible for salary increases or career ladder advancement until they receive an effective or highly effective ranking.

The district provides numerous opportunities and resources for teachers at all levels to improve their practice, including induction and coaching for new teachers, non-evaluative instructional coaches, consulting teachers (strong teachers who agree to let other teachers observe their classes and to answer questions), and a library of online resources that allow teachers to observe good teaching in action.[11]

For the 2012-13 school year, about 30 percent of DCPS teachers were rated highly effective and another 45 percent were rated effective. Only 5 percent were rated minimally effective and 1 percent ineffective. [12]

DCPS remains one of the few school districts in the United States where teachers judged ineffective are quickly removed from the classroom. Since its launch in 2009, around 400 educators (in a system that employs more than 4,000 teachers) have been dismissed based on their IMPACT results. At the same time, a substantial number of highly effective teachers have been awarded financial bonuses—ranging from 500 teachers in the first year of the program to some 1,000 teachers in 2013—at a cost of around $6-8 million per year.[13] Bonuses were initially funded by outside donors, but the school system assumed that cost starting in 2012-2013.[14]

> "One of IMPACT's strengths is the system's use of more than one measure of teacher effectiveness."

Promising, but not without Potential Pitfalls

Multiple Measures of Teacher Effectiveness

Even critics agree that one of IMPACT's strengths is the system's use of more than one measure of teacher effectiveness. Such an approach is aligned with recent research in the United States showing that using multiple measures (including classroom observation and student achievement measures) provides a more fair and reliable assessment of effective teaching than student achievement measures alone.[15] Classroom observations conducted by master teachers and administrators provide objective, constructive feedback that many DCPS teachers reported they welcomed and appreciated.[16] At the same time, the IMPACT system also uses more than one measure of student achievement, and takes into account non-academic aspects of a teachers' job that are equally important to building a healthy school environment.

No measure is perfect, however. While perhaps the strongest criticisms of the system have focused on the value-added component of the evaluations (see next section), it is worth noting that some DC teachers have also complained that inconsistencies in the way

observations are conducted, and the results given, from one evaluator to another, made it hard for them to understand how to improve. Others argued that 30 -minute observations were not long enough to incorporate all "Teach" standards and that observers were not always aware of special circumstances (for example if there were high numbers of special needs students) in the classroom. In response, DCPS officials point out that teachers have a "context box" to share special circumstances with evaluators and note that all standards can be observed in the allotted time. Under the current union contract, evaluators are not allowed to share their observations with instructional coaches working to help teachers (although teachers may share them if they so choose), further complicating using the evaluation to improve practice.[17]

> "One of the dangers of a high stakes accountability system is that people have an incentive to cheat in order to make themselves look better or avoid negative consequences."

DCPS has been receptive to feedback, and the system has undergone several adjustments and improvements designed to address concerns that arose in the course of implementation, including giving less weight to value-added test scores (from 50 to 35% of the total score for group 1 teachers), adding a mechanism that allows principals to ask for a one year waiver of consequences for teachers subject to separation that they believe are capable of improvement, and making some classroom observations optional for teachers with sustained, demonstrated instructional excellence.

Value-Added Data

A key feature of the DC IMPACT system is the use of value-added measures of student achievement to assess teacher performance. Such measures have become one of the most contentious areas of debate in teacher evaluation in the United States. Proponents argue that value-added is a more meaningful way to assess teacher effectiveness than average test scores that do not take into account student background or prior academic performance.

Opponents, on the other hand, make at least four claims against the use of value-added measures to improve teacher quality.

- First, IVA undervalues teachers because it does not, and cannot, measure a complex process like teaching effectively. A creative teacher, for example, who uses innovative teaching strategies in the classroom may receive a low IVA score if his/her students do not achieve the predicted growth on the DC CAS, regardless of their sound teaching practices.
- Second, IVA calculations are complex, and consequently are difficult for parents, teachers, and administrators to understand or use to evaluate and improve performance. Such complex calculations are also more susceptible to error, with important consequences.
- Third, value-added scores can fluctuate widely from year to year, due to measurement errors in the achievement tests, students who test poorly on a given day, or in some cases, cheating (see below).
- Fourth, only those teachers whose students take standardized tests in the subject they teach in two consecutive years generate value-added data. In other words, most value added measures apply to only a limited numbers of teachers in certain tested subjects and grades. The remaining teachers are evaluated on other criteria. This raises questions of teacher equity, since value-added criteria may be harder to meet.

The DC system works to minimize these risks by adjusting the

weight of IVA in overall evaluations and balancing it with other factors; providing clear information on the IMPACT system and coaching teachers to help them improve; working with nationally respected researchers to design, compute, and review IVA calculations; and giving larger bonuses to highly effective Group 1 teachers. IMPACT also gives teachers a chance to correct errors in their student rosters to ensure they are only evaluated on students they teach.[18] At the same time, system administrators maintain a firm commitment to the idea that the ultimate measure of effective teaching is whether or not children are learning. However, the controversy over value-added continues to fuel substantial debate and research on how best to measure teachers' impact on student learning, and DC experiences will be closely watched.

Cheating

One of the dangers of a high stakes accountability system is that people have an incentive to cheat in order to make themselves look better or avoid negative consequences. In 2011, allegations of widespread cheating on the DC CAS led to an investigation by the U.S. Education Department's office of the inspector general. At issue were an unusually large number of "wrong to right" erasures on exams between 2008 and 2010, suggesting that teachers and administrators provided inappropriate help to students on the exam.[19]

DCPS released an independent evaluation of the 2009 DC CAS in response to the allegations that found no evidence of wrongdoing and plausible explanations for the high wrong-to-right erasure marks.[20] However, critics still argue that district leaders did not act quickly or broadly enough to investigate suspicions of cheating.[21]

To guard against cheating in the future, DCPS has redoubled its test security procedures, has an independent firm conduct annual test-integrity reviews, and test scores that have been identified as having been compromised for security reasons are invalidated and not used in value-added analysis.[22] Subsequent evaluations found evidence of cheating by only three teachers in 2011 and by teachers in 18 classrooms in 2012.[23] In 2013, evaluators found evidence of serious falsifications in four schools, and moderate or minor violations of test security at 13 other schools.[24] These figures do not indicate a systemic culture of cheating, but they do suggest that systems seeking to develop teacher accountability measures must think critically about unintended consequences of high performance incentives and how to prevent cheating once accountability measures are implemented.

Teacher Retention and Performance

Evidence from a 2013 study of the DC IMPACT program suggests that it is meeting at least one of its intended goals—to retain good teachers and remove ineffective teachers from the classroom. According to the study, which covered the first two years of IMPACT's implementation, DCPS retained nearly 90% of teachers rated "highly effective" over the two years. All teachers rated "ineffective" were dismissed, and 30% of first-time "minimally effective" teachers voluntarily resigned from the school system. At the same time, the study found that high-performing teachers close to the threshold for receiving bonuses, and low-performing teachers who stayed in the system improved their performance.[25] The net result is an increase in overall teacher quality.

However, because the study focuses on those teachers just above the threshold for getting fired and just below the threshold for getting bonuses, it did not provide information on how the system affects "average" teachers (those who fall in the middle of the distribution). Nor is it clear yet how recent changes to the IMPACT system may affect findings over time. Still, initial results are promising.

In any high-stakes evaluation system, there is a danger that good teachers will be mistakenly dismissed or that bad teachers will be retained. For example, a programming error by the firm contracted to calculate teachers' value-added scores resulted in erroneous IMPACT ratings for 44 teachers in 2012-2013 (around 1% of the DCPS teacher workforce). Half of those teachers received evaluations that were too high, and half too low. One teacher (in a system of 4,000) was fired as a

result of the error. Both the contracting firm and DCPS publicly acknowledged and corrected the problem immediately when it was discovered. DCPS decided to leave the ratings that were too high unchanged and raised those that were too low. The school system offered to reinstate and compensate the teacher that was fired as a result of the error.[26]

To be sure, correlation does not equal causation, and it is not possible to say to what extent achievement gains in DC are due to changes in teacher evaluation or to other changes in the system or to some combination of factors. However, they do provide a promising indication of progress, and further study of the links between specific reforms and student learning is needed to take full advantage of potential lessons.

"Countries considering reforms similar to IMPACT will need to plan ahead to make sure they have human, institutional and financial resources available."

Student Test Scores

One interesting development has been a rise in DCPS students' test scores on national assessments since 2007. Results from the 2013 NAEP show statistically significant gains in math and reading for both fourth and eighth grade students.[27] On average, the district saw double-digit gains in both 4th and 8th grade math scores, and close to 10 point gains in reading in both grades compared with 2007, with the bulk of the point gain occurring since 2009. Average scores nationally improved by 5 points or less over the same time period.[28] And unlike the DC CAS, scores from the lower-stakes NAEP have not been subject to allegations of cheating.

Implications for Policy

Several aspects of the DC experience are worth highlighting in the Latin American context:

1. ***A clear focus on learning and transparency.*** DC administrators have insisted from Day One that the ultimate measure of good teaching is how well students learn, and have steadfastly maintained that focus in the face of obstacles and pushback from different actors. They established a clear teaching framework and have worked hard to explain to teachers what they are expected to do and how to do it. Guidebooks on how evaluations are conducted, how scores are calculated and what constitutes performance at any given rating level are easily accessible on the DCPS website. The district also has developed a virtual library that shows what good practice looks like. Few Latin American countries have defined effective teaching so clearly or made learning the central measurement of success.

2. ***A serious commitment to creating the best quality evaluation system possible, based on cutting edge research, customized to DC, and maintained over time.*** DCPS has gone out of its way to ensure that the design, review, and implementation of IMPACT conform to the best available research, and has invested considerable human and financial resources to do so.[29] It has recruited high-quality individuals to run the system, and reforms enacted under one administration have continued in the next. It is particularly telling that when the mayor who was responsible for appointing Rhee was voted out of office, at least in part because of Rhee's unpopular reform style, his rival hired Rhee's second in command, Kaya Henderson, to take her place. Henderson has made adjustments to IMPACT, but left the essentials untouched. Such policy continuity across rival administrations is rare in Latin America. Moreover, finding highly qualified

personnel and substantial resources to run a system like IMPACT well may be challenging in poorer countries. Countries considering reforms similar to IMPACT will need to plan ahead to make sure they have human, institutional and financial resources available.

3. ***Strong leadership and collaboration.*** IMPACT was possible in part because the chancellor of DC schools has the power to evaluate teachers, without having to negotiate every detail with unions.[30] However, Rhee's aggressive approach created numerous public relations and implementation conflicts. The new chancellor and mayor have continued the program, but with more input from and cooperation with teachers. The head of the Washington Teachers Union noted in 2012 that although he still considers the system to be flawed, revisions have made it better, and the union has been coaching members on how to work within the system.[31] Bonuses under IMPACT were collaboratively developed by DCPS and the Washington Teachers Union and are part of the union-negotiated contract.[32] In other words, DCPS leaders were able to push through initial resistance and hold firm on consequences, while at the same time being "open to making smart adjustments to the system" that included teachers' input.[33] The combination may suggest strategies for managing changes Latin America, where resistance to teacher evaluations is likely to be even greater.

4. ***Quick and responsible response to errors.*** No system will be perfect from the outset, particularly systems that combine complex value-added formulas with high stakes. This means systems need to be willing and able to adjust in response to problems that arise during implementation. It also means recognizing that errors are inevitable and developing plans for dealing with them when they arise. DCPS' strategy of acknowledging and addressing the mistaken performance ratings quickly in a way that appeared fair to those teachers adversely affected helped maintain confidence in the reform. In Latin American countries, where concerns about corruption are high and public trust in governments is low, building and maintaining a reputation for dealing with mistakes transparently and fairly will be an important component of any reform.

ENDNOTES

[1]For more data on the DCPS school system, see District of Columbia Public Schools, "DCPS at a Glance."

[2]NAEP is a national student achievement test that measures US students' learning in fourth, eighth, and twelfth grades. Although it measures performance at the national level for all three grades in a variety of subjects, state and district level information for large urban areas is only available yearly for 4th and 8th grade reading and math. Science and writing data for states and large urban districts is also collected every four years. In addition to scale scores, NAEP also reports the percentage of students that score at below basic, basic, proficient, and advanced levels. For additional information on NAEP and NAEP results, see the National Center for Education Statistics website.

[3]For more on the Teaching and Learning Framework see District of Columbia Public Schools, "Teaching and Learning Framework (TLF)." 2013-2014.

[4]There are also separate guidelines for calculating IMPACT scores for early childhood education teachers, special education teachers, teachers who teach students who are English Language Learners, student support professionals, visiting teachers, shared special subject teachers, and media specialists. Detailed IMPACT Guidebooks for Group 1, Group 2, and each of the other categories are available here.

[5]The Common Core Standards were developed by state education chiefs and governors in 48 states and define clear college- and career-ready standards for kindergarten through 12th grade in English language arts and mathematics. Today, 43 states have voluntarily adopted and are working to implement the standards. For more on Common Core Standards see the Common Core webpage.

[6]For more information, see District of Columbia Public Schools, "IMPACT - The DCPS Effectiveness Assessment System for School-Based Personnel."

[7]For more on the DCPS career ladder, see District of Columbia Public Schools, "Leadership Initiative For Teachers (LIFT)."

[8]For more on how DC value added is calculated, see District of Columbia Public Schools, "Individual Value-Added Student Achievement Data;" and Isenberg, E., and Walsh, E., "Measuring Teacher Value Added in DC, 2012-2013 School Year." Mathematica Policy Research, September 11, 2013.

[9]District of Columbia Public Schools, "General Education Teachers, Group 1." 2013 -2014.

[10]Highly effective teachers receive a $10,000 bonus if they work in schools where 60% or more of students qualify for free or reduced price meals, an additional $10,000 if their school is one of the 40% lowest performing schools, and an additional $5,000 if they are a Group 1 teacher. Highly effective teachers working in schools with lower levels of students qualifying for free or reduced price lunch receive smaller bonuses ($2,000), although they may earn an additional $1,000 bonus if they are a Group 1 teacher. For more information see this DCPS overview.

[11]For more on these opportunities for support and professional growth see District of Columbia Public Schools, "Ensuring Teacher Success."

[12]Anderson, N. "D.C. Schools Gave 44 Teachers Mistaken Job Evaluations." Washington Post, December 23, 2013.

[13]Costs were substantially less in the first year. According to DCPS, around $25 million in bonuses have been awarded over the last four years.

[14]See Brown, E. "98 D.C. Teachers Fired for Poor Performance." The Washington Post, August 1, 2012; and Hanushek, E., "Evidence-based Debates on Teacher Quality." Education Next, March 24, 2014.

[15]Bill and Melinda Gates Foundation, "Ensuring Fair and Reliable Measures of Effective Teaching." 2013.

[16]Headden, S., "Inside IMPACT: D.C.'s Model Teacher Evaluation System." Education Sector, 2013.

[17]Ibid.

[18] This is one of the suggestions for guarding against errors identified by the Gates teacher evaluation study in the U.S. See Bill and Melinda Gates Foundation, "Ensuring Fair and Reliable Measures of Effective Teaching." 2013, p. 13.

[19]See Toppo, G, "Memo warns of rampant cheating in D.C. public schools." USA Today, April 11, 2013; and Gillum, J. and M. Bello, "When standardized test scores soared in D.C., where the gains real?" USA Today, March 30, 2011.

[20]District of Columbia Public Schools, "DC Public Schools Releases All 2009 DC CAS Security Investigation Reports." March 31, 2011.

[21]See, for example, Strauss, V., "Why not subpoena everyone in D.C. cheating scandal – Rhee included? (update)." The Washington Post, April 12, 2013; and Sanchez, L., "Long-Lost Memo Stirs Allegation Of Cheating In D.C. Schools." NPR, May 6, 2013.

[22]EdCORE, "Evaluation of the DC Public Education Reform Amendment Act (PERAA)." July 15, 2013.

[23]Brown, E. "Teachers in 18 D.C. Classrooms Cheated on Tests Last Year, Probe Finds." The Washington Post, April 12, 2013.

[24] Brown, E. "Review Finds Serious Test-taking Violations in Four D.C. Schools." The Washington Post, March 19, 2014.

[25] Dee, T. & Wyckoff, J., "Incentives, Selection, and Teacher Performance: Evidence from IMPACT." National Bureau of Economic Research, 2013.

[26]Anderson, N. "D.C. Schools Gave 44 Teachers Mistaken Job Evaluations." Washington Post, December 23, 2013. See also National Council on Teacher Quality, "The value-add of value-add." PDQ: Pretty Darn Quick Blog, December 24, 2013.

[27] National Center for Education Statistics, "2013 Mathematics Assessment Report Card;" and "2013 Reading Assessment Report Card."

[28] District of Columbia Public Schools, "2013 DC NAEP Student Achievement Results."

[29]For more on these and other issues that make the DC experience unique, see Hess, R., "Making Sense of the Dee-Wyckoff IMPACT Study." Education Week, October 18, 2013.

[30]Hess, R., "Making Sense of the Dee-Wyckoff IMPACT Study." Education Week, October 18, 2013.

[31]Brown, E. "98 D.C. Teachers Fired for Poor Performance". The Washington Post, August 1, 2012.

[32]District of Columbia Public Schools, "IMPACTplus."

[33]Hess, R., "Making Sense of the Dee-Wyckoff IMPACT Study." Education Week, October 18, 2013.

1211 Connecticut Ave., NW, Suite 510

Washington, DC, 20036

Phone: 202-822-9002 • Fax: 202-822-9553

Email: education@thedialogue.org

Web: thedialogue.org & prealblog.org