

Looking Back on DC Education Reform 10 Years After, Part 1: The Grand Tour

Richard P Phelps

Ten years ago, I worked as the Director of Assessments for the District of Columbia Public Schools (DCPS). My tenure coincided with Michelle Rhee's last nine months as Chancellor. I departed shortly after Vincent Gray defeated Adrian Fenty in [the September 2010 DC mayoral primary](#).

My primary task was to design an expansion of that testing program that served the [IMPACT teacher evaluation system](#) to include all core subjects and all grade levels. Despite its fame (or infamy), the test score aspect of the IMPACT program affected only 13% of teachers, those teaching either reading or math in grades four through eight. Only those subjects and grade levels included the requisite pre- and post-tests required for teacher "value added" measurements (VAM). Not included were most subjects (e.g., science, social studies, art, music, physical education), grades kindergarten to two, and high school.

Chancellor Rhee wanted many more teachers included. So, I designed a system that would cover more than half the DCPS teacher force, from kindergarten through high school. You haven't heard about it because it never happened. The newly elected Vincent Gray had promised during his mayoral campaign to *reduce* the amount of testing; the proposed expansion would have increased it fourfold.

VAM affected teachers' jobs. A low value-added score could lead to termination; a high score, to promotion and a cash bonus. VAM as it was then structured was obviously, glaringly flawed,¹ as anyone with a strong background in educational testing could have seen. Unfortunately, among the many new central office hires from the elite of ed reform circles, none had such a background.

Before posting a request for proposals from commercial test developers for the testing expansion plan, I was instructed to survey two groups of stakeholders—central office managers and school-level teachers and administrators.

¹ Even a primary grades teacher with the same group of students the entire school day had those students for less than six hours a day, five days a week, for less than half the year. All told, even in the highest exposure circumstances, a teacher interacted with the same group of students for less than a tenth of each student's waking hours in a year, and for less than a twentieth in the tested subjects of English and math. In the lowest exposure circumstance, a high school teacher might interact with a class of English or math students for less than three percent of a student's annual hours.

Not surprisingly, some of the central office managers consulted requested additions or changes to the proposed testing program where they thought it would benefit their domain of responsibility. The net effect on school-level personnel would have been to add to their administrative burden. Nonetheless, all requests from central office managers would be honored.

The Grand Tour

At about the same time, over several weeks of the late Spring and early Summer of 2010, along with a bright summer intern, I visited a dozen DCPS schools. The alleged purpose was to collect feedback on the design of the expanded testing program. I enjoyed these meetings. They were informative, animated, and very well attended. School staff appreciated the apparent opportunity to contribute to policy decisions and tried to make the most of it.

Each school greeted us with a full complement of faculty and staff on their days off, numbering a several dozen educators at some venues. They believed what we had told them: that we were in the process of redesigning the DCPS assessment program and were genuinely interested in their suggestions for how best to do it.

At no venue did we encounter stand-pat knee-jerk rejection of education reform efforts. Some educators were avowed advocates for the Rhee administration's reform policies, but most were basically dedicated educators determined to do what was best for their community within the current context.

The Grand Tour was insightful, too. I learned for the first time of certain aspects of DCPS's assessment system that were essential to consider in its proper design, aspects of which the higher-ups in the DCPS Central Office either were not aware or did not consider relevant.

The group of visited schools represented DCPS as a whole in appropriate proportions geographically, ethnically, and by education level (i.e., primary, middle, and high). Within those parameters, however, only schools with "friendly" administrations were chosen. That is, we only visited schools with principals and staff openly supportive of the Rhee-Henderson agenda.

But even they desired changes to the testing program, whether or not it was expanded. Their suggestions covered both the annual districtwide DC-CAS (or "comprehensive" assessment system), on which the teacher evaluation system was based, and the DC-BAS (or "benchmarking" assessment system), a series of four annual "no-stakes" interim tests unique to DCPS, ostensibly offered to help prepare students and teachers for the consequential-for-some-school-staff DC-CAS.²

² Though officially "no stakes," some principals analyzed results from the DC-BAS to identify students whose scores lay just under the next higher benchmark and encouraged teachers to focus their instructional efforts on them. Moreover, at the high school level, where testing occurred only in grade 10, students who

At each staff meeting I asked for a show of hands on several issues of interest that I thought were actionable. Some suggestions for program changes received close to unanimous support. Allow me to describe several.

1. Move DC-CAS test administration later in the school year. Many citizens may have logically assumed that the IMPACT teacher evaluation numbers were calculated from a standard pre-post test schedule, testing a teacher's students at the beginning of their academic year together and then again at the end. In 2010, however, the DC-CAS was administered in March, three months before school year end. Moreover, that single administration of the test served as both pre- and post-test, posttest for the current school year and pretest for the following school year. Thus, before a teacher even met their new students in late August or early September, almost half of the year for which teachers were judged had already transpired—the three months in the Spring spent with the previous year's teacher and almost three months of summer vacation.

School staff recommended pushing DC-CAS administration to later in the school year. Furthermore, they advocated a genuine pre-post-test administration schedule—pre-test the students in late August–early September and post-test them in late-May–early June—to cover a teacher's actual span of time with the students.

This suggestion was rejected because the test development firm with the DC-CAS contract required three months to score some portions of the test in time for the IMPACT teacher ratings scheduled for early July delivery, before the start of the new school year. Some small number of teachers would be terminated based on their IMPACT scores, so management demanded those scores be available before preparations for the new school year began.³ The tail wagged the dog.

2. Add some stakes to the DC-CAS in the upper grades. Because DC-CAS test scores portended consequences for teachers but none for students, some students expended little effort on the test. Indeed, [extensive research](#) on “no-stakes” (for students) tests reveal that motivation and effort vary by a range of factors including gender, ethnicity, socioeconomic class, the weather, and age. Generally, the older the student, the lower the test-taking effort. This disadvantaged some teachers in the IMPACT ratings for circumstances beyond their control: unlucky student demographics.

performed poorly on the DC-BAS might be artificially re-classified as held-back 9th graders or advanced prematurely to 11th grade in order to avoid the DC-CAS.

³ Even a primary grades teacher with the same group of students the entire school day had those students for less than six hours a day, five days a week, for less than half the year. All told, even in the highest exposure circumstances, a teacher interacted with the same group of students for less than a tenth of each student's waking hours in a year, and for less than a twentieth in the tested subjects of English and math. In the lowest exposure circumstance, a high school teacher might interact with a class of English or math students for less than three percent of a student's annual hours.

Central office management rejected this suggestion to add even modest stakes to the upper grades' DC-CAS; no reason given.

3. *Move one of the DC-BAS tests to year end.* If management rejected the suggestion to move DC-CAS test administration to the end of the school year, school staff suggested scheduling one of the no-stakes DC-BAS benchmarking tests for late May–early June. As it was, the schedule squeezed all four benchmarking test administrations between early September and mid-February. Moving just one of them to the end of the year would give the following year's teachers a more recent reading (by more than three months) of their new students' academic levels and needs.

Central Office management rejected this suggestion probably because the real purpose of the DC-BAS was not to help teachers understand their students' academic levels and needs, as the following will explain.

4. *Change DC-BAS tests so they cover recently taught content.* Many DC citizens probably assumed that, like most tests, the DC-BAS interim tests covered recently taught content, such as that covered since the previous test administration. Not so in 2010. The first annual DC-BAS was administered in early September, just after the year's courses commenced. Moreover, it covered the same content domain—that for the entirety of the school year—as each of the next three DC-BAS tests.

School staff proposed changing the full-year “comprehensive” content coverage of each DC-BAS test to partial-year “cumulative” coverage, so students would only be tested on what they had been taught prior to each test administration.

This suggestion, too, was rejected. Testing the same full-year comprehensive content domain produced a predictable, flattering score rise. With each DC-BAS test administration, students recognized more of the content, because they had just been exposed to more of it, so average scores predictably rose. With test scores always rising, it looked like student achievement improved steadily each year. Achieving this contrived score increase required testing students on some material to which they had not yet been exposed, both a violation of professional testing standards and a poor method for instilling student confidence. (Of course, it was also less expensive to administer essentially the same test four times a year than to develop four genuinely different tests.)

5. *Synchronize the sequencing of curricular content across the District.* DCPS management rhetoric circa 2010 attributed classroom-level benefits to the testing program. Teachers would know more about their students' levels and needs and could also learn from each other. Yet, the only student test results teachers received at the beginning of each school year was half-a-year old, and most of the information they received over the course of four DC-BAS test administrations was based on not-yet-taught content.

As for cross-district teacher cooperation, unfortunately there was no cross-District coordination of common curricular sequences. Each teacher paced their subject matter however they wished and varied topical emphases according to their own personal preference.

It took DCPS's Chief Academic Officer, Carey Wright, and her chief of staff, Dan Gordon, less than a minute to reject the suggestion to standardize topical sequencing across schools so that teachers could consult with one another in real time. Tallying up the votes: several hundred school-level District educators favored the proposal, two of Rhee's trusted lieutenants opposed it. It lost.

6. Offer and require a keyboarding course in the early grades. DCPS was planning to convert all its testing from paper-and-pencil mode to computer delivery within a few years. Yet, keyboarding courses were rare in the early grades. Obviously, without systemwide keyboarding training in computer use [some students](#) would be at a disadvantage in computer testing.

Suggestion rejected.

In all, I had polled over 500 DCPS school staff. Not only were all of their suggestions reasonable, some were essential in order to comply with professional assessment standards and ethics.

Nonetheless, back at DCPS' Central Office, each suggestion was rejected without, to my observation, any serious consideration. The rejecters included Chancellor Rhee, the head of the office of Data and Accountability—the self-titled "Data Lady," Erin McGoldrick—and the head of the curriculum and instruction division, Carey Wright, and her chief deputy, Dan Gordon.

Four central office staff outvoted several-hundred school staff (and my recommendations as assessment director). In each case, the changes recommended would have meant some additional work on their parts, but in return for substantial improvements in the testing program. Their rhetoric was all about helping teachers and students; but the facts were that the testing program wasn't structured to help them.

What was the purpose of my several weeks of school visits and staff polling? To solicit "buy in" from school level staff, not feedback.

Ultimately, the new testing program proposal would incorporate all the new features requested by senior Central Office staff, no matter how burdensome, and not a single feature requested by several hundred supportive school-level staff, no matter how helpful. Like many others, I had hoped that the education reform intention of the Rhee-Henderson years was genuine. DCPS could certainly have benefitted from some genuine reform.

Alas, much of the activity labelled "reform" was just for show, and for padding resumes. Numerous central office managers would later work for the Bill and Melinda

Gates Foundation. Numerous others would work for entities supported by the Gates or aligned foundations, or in jurisdictions such as Louisiana, where ed reformers held political power. Most would be well paid.

Their genuine accomplishments, or lack thereof, while at DCPS seemed to matter little. What mattered was [the appearance of accomplishment](#) and, above all, loyalty to the group. That loyalty required going along to get along: complicity in maintaining the façade of success while withholding any public criticism of or disagreement with other in-group members.

Unfortunately, in the United States what is commonly showcased as education reform is neither a civic enterprise nor a popular movement. Neither parents, the public, nor school-level educators have any direct influence. Rather, at the national level, US education reform is an elite, private club—[a small group of tightly-connected politicians and academics](#)—[a mutual admiration society](#) dedicated to the career advancement, political influence, and financial benefit of its members, supported by a gaggle of wealthy foundations (e.g., Gates, Walton, Broad, Wallace, Hewlett, Smith-Richardson).

For over a decade, The Ed Reform Club exploited DC for its own benefit. Local elite formed the [DC Public Education Fund](#) (DCPEF) [to sponsor education projects](#), such as IMPACT, which they deemed worthy. In the negotiations between the Washington Teachers’ Union and DCPS concluded in 2010, DCPEF arranged a 3 year grant of \$64.5M from the Arnold, Broad, Robertson and Walton Foundations to fund a 5-year retroactive teacher pay raise in return for contract language allowing teacher excessing tied to IMPACT, which Rhee promised would lead to annual student test score increases by 2012. Projected goals [were not met](#); [foundation support continued nonetheless](#).

Michelle Johnson (nee Rhee) now [chairs the board](#) of a charter school chain in California and occasionally collects [\\$30,000+ in speaker fees](#) but, otherwise, seems to have deliberately withdrawn from the limelight. Despite contributing [her own additional scandals](#) after she assumed the DCPS Chancellorship, Kaya Henderson ascended to [great fame and glory](#) with a “distinguished professorship” at Georgetown; honorary degrees from Georgetown and Catholic Universities; gigs with the Chan Zuckerberg Initiative, Broad Leadership Academy, and Teach for All; and board memberships with The Aspen Institute, The College Board, Robin Hood NYC, and Teach For America. Carey Wright is now [state superintendent in Mississippi](#). Dan Gordon runs a 30-person consulting firm, [Education Counsel](#) that [strategically partners](#) with [major players](#) in US education policy. The manager of the IMPACT teacher evaluation program, Jason Kamras, now works as [Superintendent](#) of the Richmond, VA public schools.

Arguably the person most directly responsible for the recurring assessment system fiascos of the Rhee-Henderson years, then Chief of Data and Accountability Erin McGoldrick, now specializes in “data innovation” as [partner and chief operating officer](#) at an education management consulting firm. Her firm, Kitamba, strategically partners with its [own panoply](#) of major players in US education policy. Its list of recent clients includes the DC Public Charter School Board and DCPS.

If the ambitious DC central office folk who gaudily declared themselves leading education reformers were not really, who were the genuine education reformers during the Rhee-Henderson decade of massive upheaval and per-student expenditures three times those in the state of Utah? They were the school principals and staff whose practical suggestions were ignored by central office glitterati. They were whistleblowers like history teacher [Erich Martel](#) who had [documented](#) DCPS' student records' manipulation and phony graduation rates years before the [Washington Post's celebrated investigation of Ballou High School](#), and was demoted and then "excessed" by Henderson. Or, school principal [Adell Cothorne](#), who spilled the beans on test answer sheet "erasure parties" at Noyes Education Campus and lost her job under Rhee.

Real reformers with "skin in the game" can't play it safe.

The author appreciates the helpful comments of Mary Levy and Erich Martel in researching this article.