

# A Teacher Wonders

## Can Grading Teachers Work?



BY MARC EPSTEIN

**R**eport Card: “a card containing a report; specifically, a card, submitted by a school, exhibiting a pupil’s record to his parents or guardian.”

This terse definition appeared in the classic 1934 edition of *Webster’s Second International Dictionary*. Until recently, if you’d ask the proverbial man on the street for a definition, I’d venture to say that the overwhelming response would closely match *Webster’s*. But all that appears to be changing since school systems throughout the country are issuing report cards to schools, and now, to their faculties. The teachers’ unions, purported by critics to be omnipotent, are doing their best to participate in the debate as school reformers insist on holding teachers accountable for improving educational performance.

A recent cartoon in the *Wall Street Journal* pretty much says it all: a young student presents his failing report card to his

teacher and opines, “Ah, Miss Brimsley, I ask you: Which one of us has truly failed?”

A series of new assumptions in the world of educational theory have become axiomatic. As the *Time* magazine cover story on February 25, 2008, “How To Make Great Teachers,” put it, “There’s no magic formula for what makes a good teacher, but there is general agreement on some of the prerequisites. One is an unshakeable belief in children’s capacity to learn. ‘Anyone without this has no business in the classroom,’ says Margaret Gayle, an expert on gifted education at Duke University.”

If every child can learn, then it follows that the reason for poor student performance must lie elsewhere. In his book *Doomed to Fail*, Paul Zoch documents the steady march of public education in this country, over the past century, toward a system of teacher-centered responsibility for learning. The latest iteration of this trend is the theory that rewarding and punishing teachers based on the extent to which their students’ test scores increase will solve the riddle of public education’s Gordian knot.

For example, several months ago, a political fight erupted between the United Federation of Teachers in New York City and the mayor over the validity of using students’ test results to determine teacher tenure. The mayor, a strong advocate of using test results to evaluate teacher performance, said, “All of us are

*Marc Epstein teaches social studies at Jamaica High School in Queens, New York. He has written articles on the New York City schools over the past five years, appearing in a variety of journals and newspapers, and was a contributor to A Consumer’s Guide to High School History Textbooks, edited by Diane Ravitch.*

judged on whether or not we do a good job. And to not judge teachers the same way, it's an insult to the teachers" (*New York Sun*, April 7, 2008). In an op-ed that appeared in the *New York Daily News* (April 8, 2008), schools Chancellor Joel Klein weighed in with his support: "Research tells us that a teacher's track record in helping students learn over a few years is a powerful indicator of whether that teacher is going to help his or her students succeed over the course of a career." At the end of the day, the New York State legislature barred the use of student test scores for making tenure decisions for a two-year period, seemingly granting another victory to the allegedly "obstructionist" teachers' union, and instead created a commission to study the issue.

The fact is, using test results to judge teacher performance is much trickier than it sounds—student test scores are influenced by all sorts of things that are beyond a teacher's control. So researchers are currently working to develop a way of isolating the teacher's impact, if indeed that is even possible. (To see how far they have come and what challenges remain, read Harvard University Professor Daniel Koretz's article that starts on page 18.)

To date, the most well-known (although not the most highly regarded) approach has been crafted by William Sanders, formerly a statistician at the University of Tennessee and now a senior research fellow with the SAS Institute Inc. It's based on the Tennessee Value-Added Assessment System, and it was developed for the Tennessee Department of Education in 1992. The Sanders model ranks teachers according to how much more, or less, growth their students have made compared with the average teacher; "effective" teachers are those whose students made above average growth (by a margin considered to be statistically significant). Sanders claims that by focusing on growth, the value-added model removes socioeconomic factors that play an important role in student achievement, such as family and home environment, and that the results can help improve teaching performance. Critics of the Sanders model argue that it is far too simple. For example, they think it does not adequately account for numerous student-background, classroom, and school factors that play a role in classroom achievement. Nonetheless, the Sanders model is just one of many. They all have their strengths and weaknesses, but none is able to fully and accurately isolate the teacher's impact on student growth.

Variations of value added have been adopted throughout the country, with New York City the nation's largest and most recent school system to sign on. The city purchased a new \$80 million computer tracking system (with so many glitches that, at best, it's a work in progress) to chart the progress of its 1 million students. The desire to apply a value-added system throughout our nation's schools prompts a critical question that has largely been ignored. Will this tracking method be useful in a school system like New York City's, where all sorts of data indicate that the students are very mobile?

Many struggling urban school districts (such as Chicago and Los Angeles) have been handed over to mayors, retired generals, a former governor, a federal prosecutor, corporate lawyers, and businessmen whose only experience with the educational system is their memories of their own education. These "reformers" argue that a new paradigm that measures teacher and school

performance the same way it's done in the "real world" will turn our schools around. But are their memories applicable to today?

When I've looked back at my class pictures beginning with kindergarten at P.S. 139 in Rego Park, New York, I can track the physical growth of my classmates year to year because everyone, with one or two exceptions, remained in my school. In fact, I can remember only one new addition from another country, a boy from Germany named Walter who entered my fifth-grade class.

Variations of value added have been adopted throughout the country, with New York City the most recent to sign on. But will it be useful in a school system where the students are very mobile? We lose plenty of students to other states, and the students coming in are often not only from other states, but from other countries.

Also, during my years in elementary school, not a single teacher was added to or subtracted from the faculty. Under these conditions, a value-added model might have provided us with useful data regarding student progress and teacher effectiveness. But those are not today's conditions.

Like other urban areas, New York is now a city of extraordinary mobility. Students move in, and students move out, changing schools and neighborhoods and cities. A recent study conducted by New York University's Institute for Education and Social Policy tracked the progress of about 86,000 children who entered the first grade in the fall of 1995.<sup>1</sup> The results are startling, even though they confirm my own observations of student turnover where I teach in Jamaica, Queens. After eight years, almost 40 percent of the students had left the New York City public schools.

Douglas Harris, a University of Wisconsin-Madison researcher who develops and studies value-added models, has noted that mobility poses a major problem for value-added models because it leads to missing data. And, although we all know that, on average, highly mobile students are not identical to their less mobile peers, these models assume that data are missing at random. As he puts it, this assumption "is especially likely to be a problem in high-poverty schools where absenteeism and mobility are high and test-taking rates are lower. It is therefore a significant question whether valid value-added estimates can be made in schools with high mobility."<sup>2</sup> Hopefully, officials in other cities will heed Harris's warning—those in New York City have not.

In New York City, the sophisticated new computer system tracks students' scores as they move around the district, and it can link to a statewide database as well. That reduces the missing data problem, but it certainly does not eliminate it. We lose

plenty of students to other states, and the students coming in are often not only from other states, but from other countries.

But as far as I can tell, neither this nor any other concern about the validity of value-added modeling bothers city officials at all. They are boldly piloting their own, highly suspect model that uses two years of student data and judges teacher performance by considering the growth of as few as three students. At best, this is irresponsible. But wait, it gets worse: since state achievement tests are given in the middle of each school year,

the growth of all students—even those who don't switch schools at all—has to be divvied up across two teachers. This model apportions the amount of growth each teacher produced according to the number of months the teacher taught that student—a tactic that is clearly a poor substitute for an exact attribution (since it's possible that, month for month, students grew more with one teacher than the other). As a teacher, this really bothers me. I don't want the credit for another teacher's good work (or the blame for another teacher's not-so-good work).

## Teacher, Mentor, Tutor, Specialist

### *Is Any One Educator Responsible for Student Learning?*

BY LINDA VALLI, ROBERT G. CRONINGER,  
AND KIRK WALTERS

A fundamental premise of much of the current research on teaching is that teaching quality is central to student learning. One result of this research, though not necessarily intended, has been the call to base individual teacher evaluations on contributions to student achievement gains. Given the potentially high stakes for teachers, these proposals almost always generate heated debate. While much of the debate revolves around methodological issues in using student achievement data to evaluate teacher performance, we raise an even more fundamental question, one that has received little attention from proponents of teacher accountability policies: just who is doing the teaching?

As part of a longitudinal study of the teaching of reading and mathematics, we sought to link fourth- and fifth-grade students to the individual teacher responsible for their instruction. While we recognized that students often interact with multiple adults around subject matter, the scope, forms, and duration of these interactions surprised

us. As we observed the flow of students and adults in and out of classrooms, we identified a range of more complex instructional designs quite different from the traditional "egg-crate" classroom, where one teacher works with a group of students in isolation from other adults.\* In our schools, instead of students having one teacher responsible for their yearly progress in a particular subject area, many students had multiple adults engaged in their instruction, especially if the students were considered part of one or more "at-risk" groups (e.g., English language learners, low-income students, or special education students).

We started asking ourselves the question, "Who (else) is the teacher?" while engaged in a multiyear study of fourth- and fifth-grade reading and mathematics classes. Our goal was to learn more about teaching practices, as well as the allocation of school resources and educational policies, that assist or hamper the acquisition of foundational skills in these two subject areas. Although the primary purpose of the study was not to examine student assignments and alternative instructional designs, we became interested in these topics at the end of the second year of data collection because it became increasingly apparent that these designs varied among schools and among classes in schools.

The schools in the study are part of one of the largest and most diverse school systems in the nation. Over 40 percent of the students are African American or Hispanic, more than 30 percent receive free or reduced-price

meals, and over 20 percent have been enrolled in English for speakers of other languages (ESOL) programs. The study design called for us to identify a group of moderate- to high-poverty schools with greater than expected achievement gains in the district, and then to follow these schools and their fourth- and fifth-grade teachers for three years.

For this study of who is doing the teaching, we drew on data collected at 18 elementary schools during the 2003-04 school year: a resource survey that asked teachers about instructional assistance; teachers' class rosters and daily logs; principal interviews about resource allocations and decision-making; and conversations with teachers about resource help and student reassignments.

As we collected data in the participating schools, we found substantial variation—some anticipated, some not—in how students and teachers were linked for instructional purposes.

Education researchers and policymakers are generally aware of some of the challenges associated with isolating teacher effects on student learning. For example, there is wide recognition that teacher absences require some sharing of instructional responsibilities among teachers. Because absences are the result of everything from attendance at individualized education program (IEP) meetings or professional development activities to personal illness or maternity leave, they may involve the sharing of instructional responsibility for a small part of a school day or a significant part of a school year. Based on the daily logs kept by teachers in the study, the average amount of time that someone other than the assigned teacher had

---

Linda Valli is professor in the Department of Curriculum and Instruction at the University of Maryland, College Park, where Robert G. Croninger is associate professor in the Department of Education Policy and Leadership. Kirk Walters is research analyst with the American Institutes for Research. This article is excerpted from "Who (Else) Is the Teacher? Cautionary Notes on Teacher Accountability Systems," an article that Valli, Croninger, and Walters originally published in the August 2007 issue of the *American Journal of Education*. The full text is available online, for \$10, at [www.journals.uchicago.edu/doi/abs/10.1086/518492](http://www.journals.uchicago.edu/doi/abs/10.1086/518492).

\* Dan C. Lortie, *Schoolteacher: A Sociological Study* (Chicago: University of Chicago Press, 1975).

When all is said and done, does this value-added model have any value at all? To me, it appears not only costly but ineffective and misleading. Astronomers have the luxury of examining the light that gets to earth and is captured by radio telescopes millions of light years after a star has exploded. Educators, unlike astronomers, must have data that can be readily acted on if the data are to be of any use. These data, I suggest, have so

many flaws and limitations that they should not be used to evaluate teachers. □

### Endnotes

1. M. Weinstein, J. Pakes, C. Donis-Keller, and A. E. Schwartz, "From One to Eight: A Longitudinal Portrait of the First Grade Class of 1995-1996" (IESP Policy Brief, 2008), <http://steinhardt.nyu.edu/iesp/briefs>.

2. D. Harris, "Would Accountability Based on Teacher Value-Added Be Smart Policy?" (paper for the National Conference on Value-Added Modeling, April 22-24, 2008).



responsibility for instruction in reading and mathematics due to absences was roughly 7 percent.

But even when teachers are present, other factors confound a clear linkage between student achievement and teacher performance. Student mobility is one factor. With the average mobility rate in these schools at 20 percent, a significant number of students in the study would have had a teacher from another school responsible for part of their instruction during the course of the school year. An additional complication arises from the public notification requirements introduced by No Child Left Behind (NCLB). In the district we studied, the testing schedule for the purposes of NCLB ran from March to March so that parents could be provided with test results prior to the beginning of the next school year. This meant that every teacher in the study shared responsibility for achievement gains with at least one teacher from the previous year. Given the district's 9.5-month school calendar, this amounts to roughly one-quarter of students' "tested" instructional time.

Even when students stayed in the classroom, someone other than the classroom teacher could have had responsibility for their instruction. We observed classrooms where the teacher of record consistently worked with one reading group while instructional assistants worked with others, where student teachers took over a substantial proportion of instructional responsibilities, and where a staff developer took over part of the lesson to demonstrate a teaching strategy.

There were also numerous instances

where students were assigned to a specific reading or mathematics class for part of the period and sent to an ESOL or resource teacher for the rest of the period, or where a student spent the entire instructional period with the classroom teacher and received an additional reading or mathematics lesson during another part of the day with a different teacher. Homeroom teachers, who were not the reading or mathematics teacher of record, gave students work during the homeroom period targeting skills or concepts presumed to be on the annual state assessment, and computer teachers pulled small groups of students from the classroom to work on writing assignments in the computer lab. In one school, literacy instruction was divided into two separate classes, with one teacher instructing students in reading and a different teacher instructing them in writing.

In addition, we observed a surprising amount of fluidity in teacher-student assignments in some of the schools. Although the principal generally made the formal assignments at the beginning of the school year, grade-level teams sometimes adjusted these assignments, with or without the principal's knowledge. For example, grade-level teachers might pair up and switch students for a particular instructional unit and then

switch students back again. In one mathematics class, two teachers were originally assigned to co-teach a large group of students, but later in the year the group was split into two separate classes.

\* \* \*

These findings raise questions about both the feasibility and desirability of teacher accountability systems based on student achievement data. In this era of

high-stakes accountability, caution must be taken to ensure that responsibility for student learning is accurately attributed. Our analysis of these 18 schools, 69 teachers, and over 1,500 students suggests that less responsibility rests with the formally assigned classroom teacher than we initially assumed or that past studies led us to anticipate. It makes little sense to have an individual accountability model when multiple actors have a role in student learning.

Furthermore, our understanding of the potential benefits of other reform efforts tempers whatever enthusiasm we might have had for the teacher accountability movement. Even if more sophisticated statistical methods eventually make possible a more accurate attribution of teaching impact for multiple actors, this may not be a desirable direction for educational policy. It can too easily derail other efforts to support high-quality teaching and learning, including the promotion of professional learning communities and the flexible, coordinated use of trained teacher resources. This does not mean that efforts to understand and improve teaching quality are ill-conceived, only that, in many instances, teaching is a collective rather than solely individual pursuit. Education policies and teacher accountability systems need to reflect this reality.