

Posted at 01:25 PM ET, 01/21/2012

Dear Michelle Rhee: About that teacher evaluation study

By [Valerie Strauss](#)

Dear Michelle Rhee, former D.C. schools chancellor and current leader of StudentsFirst:

I just wanted to dash off a quick note about that [commentary](#) you wrote in Education Week about the big value-added teacher evaluation study that made headlines this month.

The study, titled "[The Long-Term Impacts of Teachers: Teacher Value-Added and Student Outcomes in Adulthood](#)," was conducted by two Harvard University researchers and one from Columbia. (Not shabby credentials; it's no wonder the [New York Times](#) made such a big deal of the study as an exclusive and then you decided to write about it.) The researchers claimed in their study that teachers with a high value-added score make a huge difference in the adult lives of their former students. You lauded the conclusions and said they support your own [belief in test-based school reform](#). What I wanted you to know is that they actually don't.

I read the nearly 100 pages of the report (it's taken me a few weeks) and looked at all the graphs and charts. I confess that I didn't really understand all of it; a lot of that technical stuff is over my head. But I took in enough to realize that your interpretation of the study doesn't square with the facts, and I wonder if you were misled by the authors' own confusing [executive summary](#) (which I went back to read after finishing the study just to see what the researchers themselves thought was most important).

What Raj Chetty and John N. Friedman of Harvard and Jonah E. Rockoff of Columbia did was to study the school records of 2.5 million students in a major urban district over 20 years and also got income tax records from the Internal Revenue Service to inspect. (Who knew the IRS was so friendly? Incidentally, [Diane Ravitch guessed](#) the district was New York City, but I am digressing.)

The researchers were trying to draw conclusions about the worth of value-added scores, which have become very popular with reformers such as yourself but have been [savaged by critics](#) (disclosure: I'm one of them) because a number of studies have shown them be unreliable, invalid and unfair. A value-added scores is derived from a formula (there are many different ones) that uses a student standardized- test score to determine how much value a teacher had in student learning. But of course, I'm telling you what you already know, since this method is part of the troubled IMPACT teacher evaluation system you instituted in D.C. schools when you were chancellor.

(Really, Ms. Rhee, how can a formula ever accurately factor in the impact of a sleepless night in a homeless shelter on a hungry student's performance on a [high-stakes test](#)? Did you [know that 22 percent](#) of

American children live in poverty and that low test scores are always correlated with family income? But again, I digress.)

From the mountain of data they collected, the authors concluded in the first of two parts of the study that a teacher with a high value-added score produces students who in the future will get high test scores. That's the same thing, essentially, as saying that teachers who help kids get high test scores will keep doing that in the future. That hardly seems worth a front-page story in the New York Times.

The big news was in the second part of the study. The authors concluded that teachers with high value-added scores will have a sustained effect on students that lasts through the students' adult lives.

In other words, as a friend of mine said, "High value-added teachers make high-scoring kids make successful adults."

How did the researchers measure success? They created proxies for success. Students with high-value-added teachers are supposedly:

- 1) less likely to have children as teenagers
- 2) more likely to be enrolled in a good college by the age of 20
- 3) by 28, have higher lifetime income than students who didn't have high-valued teachers.

Wow. All that from one, admittedly big, study.

The authors say the data come from the years 1989 to 2009, which to me implies that the standardized test scores used are recent. Actually, the scores are from the 1990s, well before the [No Child Left Behind era](#) ushered in those high-stakes standardized tests you like to use to hold students and and teachers and schools accountable.

The only data that come from 2009 concern the incomes of students who turned 28 in that year. Now, students who turned 28 in 2009 were born in 1981. Since the researchers used test data from grades four to eight, the students in question would have been 10-year-old fourth graders in 1991 and 14-year-old eighth graders in 1995.

So the authors used value-added data from 1991 to 1995 and then followed specific students from that period until they were 28 in 2009 and measured their income and other factors.

The executive summary, incidentally, never says the test data are so old. And on Page 5 of the actual report, the authors wrote: "*An important limitation of our analysis is that teachers were not incentivized based on test scores in the school district and time period we study.*"

"*Not incentivized*" is a euphemism in this case for old data.

The following sentence on Page 5 says, *“The signal content of value-added might be lower when it is used to evaluate teachers because of behavioral responses such as cheating or teaching to the test.”*

Signal content? That means, in this context, “a less reliable predictor.”

This all leaves me wondering exactly how reliable any of this actually is.

Now you might be wondering, why am I telling you all of this? Why does it matter?

Well, it matters a lot. It shows that students going to school in the early 1990s — in the days before No Child Left Behind gave high-stakes tests such dangerous importance in accountability systems starting in 2002 — wound up doing quite well. They earned more, lived in better places, got into better colleges and didn’t wind up as teenage parents.

So I ask you: Why exactly do we need high-stakes testing when the old tests seemed to be working just fine in helping students succeed? High-stakes tests have had many awful consequences: narrowed curriculum, cheating scandals, etc. This study seems to me to show we don’t need them!

I should note that Footnote 9, which starts on Page 5, and Footnote 64 on Page 50 say that even in the low-stakes tests that were the basis of this study, there’s a tendency for the top 2 percent of teachers ranked by value-added to have patterns of test-score gains that are consistent with cheating — and this percentage is, of course, much higher in the high-stakes era. You surely know all about the cheating scandal [in Atlanta](#) that pushed out the superintendent and others in a bunch of cities. In fact, there are [investigations now](#) into cheating when you were chancellor!

Cheating distorts the outcome, which leaves one to wonder why the authors put this important factor in a few small footnotes. Hmm.

Back to your Education Week commentary. You wrote that the study proves that the test-based reform program you started in D.C. schools when you were chancellor from 2007 to 2010 actually works. But it doesn’t prove anything of the sort.

Using the authors’ own markers of success, we can’t know until 2016 whether D.C. public school students who were in eighth grade in 2010 and had high-value-added teachers will get into good colleges. And we can’t know about the fourth graders until 2020.

Other things we won’t know:

* Not until 2020 will we know if D.C. students who were in eighth grade in 2010 and had teachers with high value-added scores will live in a high-income zip code.

* Not until 2024 will we know if D.C. students who were in fourth grade in 2010 will live in a high-income zip code.

* Not until 2024 will know if D.C. students who were in eighth grade in 2010 have higher incomes.

* Not until 2028 will we know if D.C. students who were in fourth grade in 2010 have higher incomes.

Besides, other issues have been raised about the study that give a reader pause as to what real conclusions we can draw from it.

According to [economist Bruce Baker of Rutgers University](#), writing on the School Financeblog, the income gains the authors talk about aren't really so big as it seems from reading the executive summary.

He wrote:

“One of the big quotes in The New York Times article is: ‘Replacing a poor teacher with an average one would raise a single classroom’s lifetime earnings by about \$266,000, the economists estimate.’ This comes straight from the research paper. BUT ... let’s break that down. It’s a whole classroom of kids. Let’s say ... for rounding purposes, 26.6 kids if this is a large urban district like NYC. Let’s say we’re talking about earning careers from age 25 to 65 or about 40 years. So, $\$266,000/26.6 = \$10,000$ lifetime additional earnings per individual. Hmm ... no longer catchy headline stuff. Now, per year? $\$10,000/40 = \250 . Yep, about \$250 per year.”

He also wrote that it is difficult to figure out which teachers supposedly produced these great outcomes for kids. He wrote:

“Just because teacher [value-added] scores in a massive data set show variance does not mean that we can identify with any level of precision or accuracy which individual teachers (plucking single points from a massive scatter plot) are ‘good’ and which are ‘bad.’ Therein exists one of the major fallacies of moving from large scale econometric analysis to micro level human resource management.”

Baker also concluded that while it’s a “really cool academic study,” the findings “cannot be immediately translated into what the headlines have suggested – that immediate use of value-added metrics to reshape the teacher workforce can lift the economy, and increase wages across the board!”

“ The headlines and media spin have been dreadfully overstated and deceptive,” he wrote.

The misleading commentary includes your commentary, Ms. Rhee.

If you really penned it, you might want to reconsider some corrections. If you relied on staff to write it as many organizational leaders do — I don't suppose Education Secretary Arne Duncan, for example, writes all of his great speeches — you might want to consider firing them.

Best regards, and I apologize that what I thought would be a quick note turned into a long one. These issues are really complex and can't be solved with bromides about how great teachers get great results.

Sincerely, The Answer Sheet