When The Legend Becomes Fact, Print The Fact Sheet

Matthew Di Carlo
November 11, 2011

The New Teacher Project (TNTP) just released a "fact sheet" on value-added (VA) analysis. I'm all for efforts to clarify complex topics such as VA, and, without question, there is a great deal of misinformation floating around on this subject, both "pro-" and "anti-."

The fact sheet presents five sets of "myths and facts." Three of the "myths" seem somewhat unnecessary: that there's no research behind VA; that teachers will be evaluated based solely on test scores; and that VA is useless because it's not perfect. Almost nobody believes or makes these arguments (at least in my experience). But I guess it never hurts to clarify.

In contrast, the other two are very common arguments, but they are not myths. They are serious issues with concrete policy implications. If there are any myths, they're in the "facts" column.

The first objection – that the models aren't "fair to teachers who work in high-needs schools, where students tend to lag far behind academically" – is a little confusing. In one sense, it's correct to point out that value-added models focus on growth, not absolute scores, and teachers aren't necessarily penalized just because their students "start out" low.

But most of the response to this "myth" addresses a rather different question – whether or not the models can fully account for the many factors out of teachers' hands. TNTP's take is that VA models "control for students' past academic performance and demographic factors," which, they say, means that teachers "aren't penalized for the effects of factors beyond their control." Even under ideal circumstances, that's just not accurate.

The evidence they cite is a frequently-misinterpreted paper by researchers at Vanderbilt University and the SAS Institute, published in 2004. What the analysis finds is that the results of a specific type of VA model (TVAAS) – one with very extensive data requirements, spanning multiple (in this analysis, five) years and subjects, in one specific location (Tennessee) – are not substantially different when variables measuring student characteristics (i.e., free/reduced lunch eligibility and race) are added to the models.

This does not, however, mean that the TVAAS model – or any other – can account for all the factors that teachers can't control. For one thing, the free/reduced-price lunch variable is not a very good income proxy. Eligible students vary widely in family circumstances, which is a particular problem in high-poverty areas where virtually all the students qualify.

That paper aside, it's true that students' prior achievement scores account for much of the income-based variation in achievement gains (ironically, prior test scores are probably better at this than free/reduced-priced lunch). But not all of poverty's impacts are measurable/observed, and, perhaps more importantly, there are several other potential sources of bias, including the fact that students are not randomly assigned to classrooms (also here). VA scores are also affected by the choice of model, data quality and the test used. And, of course, even if there is no bias at all, many teachers will be "treated unfairly" by simple random error.

These are the important issues, the ones that need discussion. If we're going to use these VA estimates in education policy, we need to at least do it correctly and minimize mistakes. In many places around the nation, this isn't happening (also see Bruce Baker's discussion of growth models). As a result, the number of teachers "penalized" unfairly – whether because they have high-needs students or for other reasons beyond their control – may actually be destructively high. TNTP calls this a "myth." It's not.

The second "myth" they look at is the very common argument that VA scores are too volatile between years to be useful. This too is not a "myth," but it is indeed an issue that could use some clarifying discussion. TNTP points out that all performance measures fluctuate between years, and that they all entail uncertainty. These are valid points. However, their strongest rebuttal is that "teachers who earn very high value-added scores early in their career rarely go on to earn low scores later, and vice-versa."

Their "evidence" is an influential paper by researchers from Florida State University and the RAND Corporation (it was published in 2009). The analysis focuses on the stability of VA estimates over time. While everyone might have a different definition of "rarely," it's safe to say that the word doesn't quite apply in this case. Across all teachers, for instance, only about 25-40 percent of the top quintile (top 20%) teachers in one year were in the top quintile the next year, while between 20-30 percent of them ended up in the bottom 40%. Some of this volatility appears to have been a result of "true" improvement or degradation (within-teacher variation), but a very large proportion was due to nothing more than random error.

The accurate interpretation of this paper is that value-added estimates are, on average, moderately stable from year-to-year, but that stability improves with multiple years of data and better models (also see here and here for papers reaching similar conclusions). This does not mean that teachers' scores "rarely" change over time, nor does it disprove TNTP's "myth." In fact, the paper's results show that VA estimates from poorly-specified

models with smaller samples are indeed very unstable, probably to the point of being useless. And, again, since many states and districts are making these poor choices, the instability "myth" is to some degree very much a reality.

Value-added models are sophisticated and have a lot of potential, but we have <u>no idea</u> how they are best used or whether they will work. It is, however, likely that poor models implemented in the wrong way would "penalize" critically large numbers for reasons beyond their control, as well as generate estimates that are too unstable to be useful for any purpose, even low-stakes decisions. These are not myths, they are serious risks. Given that TNTP is actively involved in redesigning teacher quality policies in <u>dozens of states and large districts</u>, it is somewhat disturbing that they don't seem to know the difference.