The following abstract and conclusion is taken from:

Volume 4, Issue 4 - Fall 2009 - Special Issue: Key Issues in Value-Added Modeling

Would Accountability Based on Teacher Value Added Be Smart Policy? An Examination of the Statistical Properties and Policy Alternatives

Douglas N. Harris of University of Wisconsin Madison

Education Finance and Policy Fall 2009, Vol. 4, No. 4: 319–350.

Available here:

http://www.mitpressjournals.org/doi/pdfplus/10.1162/edfp.2009.4.4.319

Abstract

Annual student testing may make it possible to measure the contributions to student achievement made by individual teachers. But would these "teacher value added" measures help to improve student achievement? I consider the statistical validity, purposes, and costs of teacher value-added policies. Many of the key assumptions of teacher value added are rejected by empirical evidence. However, the assumption violations may not be severe, and value-added measures still seem to contain useful information. **I also compare teacher value-added accountability with three main policy alternatives: teacher credentials, school value-added accountability, and formative uses of test data.** I argue that using teacher value-added measures is likely to increase student achievement more efficiently than a teacher credentials-only strategy but may not be the most cost-effective policy overall. Resolving this issue will require a new research and policy agenda that goes beyond analysis of assumptions and statistical properties and focuses on the effects of actual policy alternatives.

6. CONCLUSION

A great deal of attention has been paid recently to the statistical assumptions of VAMs, and many of the most important papers are contained in the present volume. The assumptions about the role of past achievement in affecting current achievement (Assumption No. 2) and the lack of variation in teacher effects across student types (Assumption No. 4) seem least problematic. However, unobserved differences are likely to be important, and it is unclear whether the student fixed effects models, or any other models, really account for them (Assumption No. 3). **The test scale is also a problem and will likely remain so because the assumptions underlying the scales are untestable**. There is relatively little evidence on how administration and teamwork affect teachers (Assumption No. 1).

The assumptions are important, but even more significant are the statistical properties of the measures. **To what degree does teacher value added reflect true differences in teacher performance?** Kane and Staiger (2008) find that some value-added models can replicate teacher performance when teachers and students are randomly assigned. There is also evidence that teacher value added is positively correlated with principals' own confidential assessments of teachers (Harris and Sass 2007c; Jacob and Lefgren 2005). This evidence suggests that despite the problematic assumptions, teacher value added still provides useful information about teacher performance. **On the other hand, teacher value-added measures are somewhat unreliable, so clear distinctions can be made only between the very highest and very lowest levels of teacher value added by traditional statistical standards.** This imprecision partly explains why teacher value added is so unstable over time (Koedel and Betts 2007), although it appears that much of the instability problem can be addressed by using multiple years of data and adjusting for measurement error (McCaffrey et al. 2009).

Some potential policies that would use teacher value added for accountability also stack up well in the policy validity framework when compared with teacher credentials. This is hardly surprising, given that the assumed goal of education here is to raise student achievement. I sometimes use a tennis analogy to make this point. If we wanted to figure out who were the best tennis players, we could carefully observe the backhand technique, footwork, serve percentages, and so on, and from that we could draw conclusions about who is better. Or we could just see who wins the most games. If winning is the goal, trying to incorporate the winning percentage into the performance measurement system is a reasonable thing to do. Measuring the equivalent of the winning percentage is more difficult in education, but the evidence here suggests that it would be worth trying.

How, then, should researchers and policy makers proceed? Given the apparent potential of teacher value added, I recommend that federal and state governments provide funds to encourage local experimentation and learn which policies work in practice (Harris 2008b). Likewise, state and federal governments should avoid putting up legal barriers to experiments with teacher value added, as the New York State Legislature did when New York City Chancellor Joel Klein proposed using student test scores in tenure decisions. Whatever the weaknesses of our decentralized system, the ability to experiment on a small scale is a clear strength and one we should take advantage of. Doing so would be largely a waste of time, however, if such

experiments were not accompanied by rigorous evaluation. The federal government has already made this mistake in the TIF grants by imposing only minimal standards on the evaluations.

Ideally, local experimentation would be done through cooperation between local unions and district management. In contrast, some merit pay plans have been forced on teachers, and both sides share the responsibility in most of these cases. Teacher unions are right to call for collaboration, but not as an excuse to maintain the status quo. Likewise, **district administrators cannot expect their calls for reform to be embraced when their arguments are accusatory and their proposals ill informed.** Not all school districts have the leadership and capacity to lead the way in these policy changes. Experimentation should take place where success is most likely, providing potential examples for others to follow.

One of the possible alternatives to teacher value-added accountability is improved school-level accountability. As I have argued elsewhere (Toch and Harris 2008), NCLB and federal accountability will never reach their full potential and may even be substantially counterproductive if school performance continues to be measured by the percentage of students meeting proficiency. **Such measures largely reward schools for who they teach rather than how well they teach, and this does little to provide incentives for real improvement.** Further, it makes little sense for state and federal governments to intervene in "failing" schools if they have not correctly identified who is failing. The remedy should match the disease. School value added also solves a political problem because schools, under the current system, have a legitimate excuse to ignore federal accountability. This is ironic given that the present system was motivated by a desire among some advocates to stop the "no excuses" mentality of schools. More important here, an accountability system that starts with school level value added could improve the effectiveness of teacher-level value-added policies by aligning the entire school performance system. **Or it may turn out that teacher value added adds little once a robust school-level accountability system is put in place.**

The evidence in this volume is important because it suggests that teacher value added has the potential to improve educational policy and student achievement. It is now time to take this effort to the next stage with a new research and policy agenda focused on putting the idea into practice.

==========

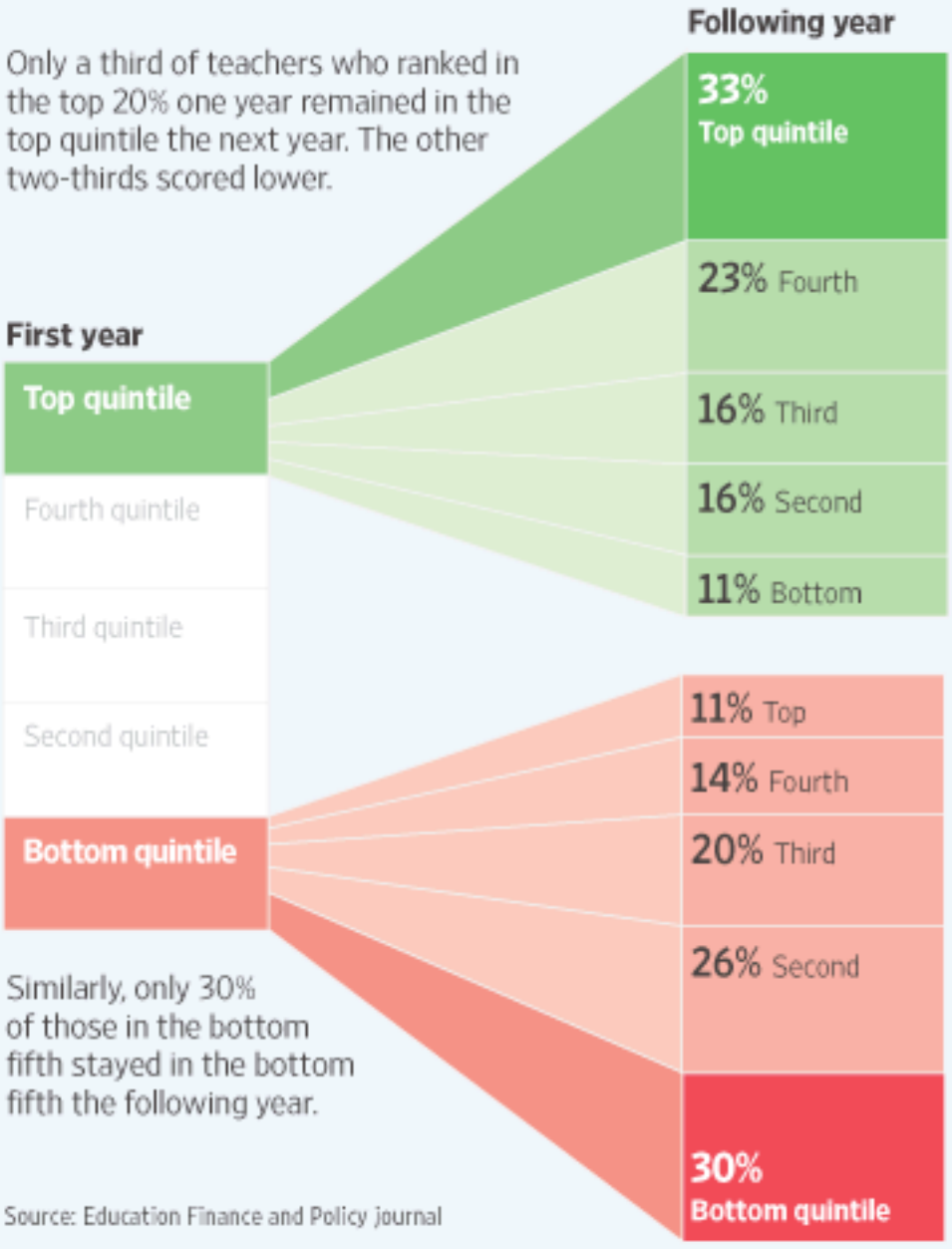
The following chart "Skipping Grades" shows the enormous variation in the Teacher Value Added approach when done using one year to assess with the particular test assessment used in Florida. Seattle's proposal to use the NWEA/MAP test looks even less reliable as the test makes no claim to be a value added tool. It certainly calls into question on what basis the Seattle administration claimed sole source supplier status for the MAP purchase. MAP has not been a useful tool for formative assessment. On what basis was it a uniquely valuable sole source tool?

Once again Seattle Admin's Performance Management proposals appear to be another Giant Seattle Experiment put forth as if the experiment is based on reliable proven relevant data, when in fact no such evidence exists. The similarity to the other Giant Experiments like the Cleveland three year math experiment, the District's k-12 math program, the 6-12 Denny/Sealth school, the coming NTN project based Cleveland STEM and so many other poorly research proposals can hardly be missed.

More here: http://www.mitpressjournals.org/toc/edfp/4/4

Skipping Grades

A study of teacher performance based on the test results of elementary school math teachers from 2000 to 2005 in Dade County, Fla.



Source: Education Finance and Policy journal