Beyond Value-Added Models...Getting the Mechanics of High-Stakes Teacher Effectiveness Policies Right

By Dan Goldhaber on January 7, 2011 8:00 AM | 

http://blogs.edweek.org/edweek/rick_hess_straight_up/2011/01/beyond_value-added_modelsgetting_the_mechanics_of_high-stakes_teacher_effectiveness_policies_right.html

Note: Dan Goldhaber, an economist and professor at the University of Washington, is guest-posting this week.

I've had a lot of opportunity lately to talk with both Race to the Top states and other states and localities that are working on incorporating student achievement measures into their teacher evaluations. Unless you've been living under a rock, you know that there is currently a major policy push to implement education reforms focused on teacher effectiveness. Programs like Race to the Top and the Teacher Incentive Fund are pushing states and localities to recognize and reward teachers in high-stakes ways, from linking pay to performance to deciding which teacher ought to be allowed to remain in the profession (through tenure and dismissal policies). Focusing on teachers makes sense. We know there is considerable variation in the effectiveness of teachers and that this variation has meaningful consequences for student achievement. Moreover, there is good evidence that today most teacher workforce policies fail to recognize differences among teachers when it comes to evaluation, professional development, and compensation.

It should come as no surprise to anyone who read Monday's blog entry that I believe we ought to be engaging in careful experimentation along the lines that many policymakers are headed. But emphasis should be put on careful, as poorly implemented policies could do great damage to the notion that paying greater attention to the strengths and weaknesses of individual teachers can lead to meaningful improvements in the quality of the teacher workforce.

When it comes to estimating teacher effectiveness based on students’ test achievement, part of being careful is a thoughtful selection of the means to translate student achievement results into teacher performance evaluations (or, more likely, a component of them). There are what might seem to some to be a dazzling array of different models--the Colorado Growth Model, the EVAAS system, the VARC system, or other "value-added" methodologies employed by economists and statisticians--and clearly there are choices to be made. Some of these models include student covariates (background variables) and some don't. Some rely on multiple years of prior student assessments, others only one.

Unfortunately, there is no right answer when it comes to picking a model--at least at this point, though we will hopefully know more when we begin to see how different models work in practice. Thus, it is no wonder that policymakers charged with implementing systems relying on
student assessments appear to be wrestling with how to strike the right balance between accurately identifying teacher contributions to student learning on the one hand and transparency (and stakeholder understanding of the method used for teacher assessments) on the other. The bottom line is that there is likely a tradeoff between these two in terms of policy options and no clear a priori reason to adopt a system that is closer to one or the other.

I would argue that far less attention is being focused on figuring out the student attribution rules that will determine which students in a teacher's classroom count toward a teacher's student-based effectiveness measure. This is unfortunate, for these rules are potentially far more important than the choice of model in influencing both measures of effectiveness and the incentives teachers will face if the measures are used for high-stakes purposes.

We know that schools today are far messier than the single room schoolhouse that once existed. Highly mobile students come and go. Pull out programs and co-teaching situations may imply that a teacher is not responsible for the entire class of students even if they are there for an entire year. So which students count towards a teacher's evaluation? Is a system that only considers 80 percent of students credible? Does it matter if there are significant differences in the proportion of higher and lower poverty kids who contribute toward teacher effectiveness measures? Is it OK for a teacher with 25 students in her class to have only 5 to 10 of them count toward her evaluation (this is not as inconceivable as it may appear at first blush when you consider the mobility levels in some schools and classrooms)?

There probably are not clear-cut right answers to these questions, but that does not mean that they should not be explored. The choices that are made could have profound implications for the proportion of the student population or of student subgroups that counts toward teachers' effectiveness measures. This in turn will have political implications and provide teachers with incentives once a system is adopted, and they merit exploration before student attribution rules are adopted.

So what might exploration mean? Here I am strongly urging policymakers to access the empirical work that could help them better understand the tradeoffs described above. In some cases this might simply be drawing on research that already exists. For example, there is terrific work by Dan McCaffrey and colleagues that illustrates the tradeoffs of some models. In other cases, e.g. with the student attribution issue, this may entail commissioning new studies. The bottom line is that policymakers should apply different attribution rules and assessment models to state and local data and see what the implications are. Find out whether there are impacts on the distribution of teachers or the percentages of students that contribute to effectiveness rankings. If there are dependencies on the choices that are made (and I suspect there will be, particularly in high-poverty classrooms with mobile student populations), policymakers should bring stakeholders into the decision-making processes. This will encourage a more robust, up-front discussion of what stakeholders value in the education system and, ultimately, better buy-in when it comes to the adoption of a particular set of rules and policies.

The kind of sensitivity analyses that I've described above is certainly a heavy lift, particularly for states and localities that are in a race against an implementation deadline. I'd argue it is essential,
however, as we need to know as best we can the tradeoffs inherent with implementing student assessment-based systems in order to best avoid pitfalls that could set the use of such systems back significantly. Mind you, I'm not arguing to slow down change when it comes to re-working teacher effectiveness policies. The last thing we need is for the perfect to be the enemy of the good--as I said before, I believe the status quo sets a low bar for change. But, a bit of empirical work (and this is not the sort of work that requires years to complete) would go a long way toward better implementation.

--Dan Goldhaber