CEDR
Center for Education
Data & Research

DO DIFFERENT VALUE-ADDED MODELS TELL US THE SAME THINGS?*

Dan Goldhaber & Roddy Theobald
Center for Education Data & Research
University of Washington Bothell

Carnegie Knowledge Network Webinar
January 9, 2013

Context

- Measures of student growth (on standardized tests) currently being incorporated into state and local teacher evaluation systems
  - Understandably, people want to know what is the "right" model
- Idea of using student growth is conceptually simple, but there is no universally agreed upon statistical methodology for translating student achievement measures into teacher performance
  - Prominent examples of where modeling approach seems to lead to relevant differences in classification of teachers
  - Disagreement about validity and reliability/stability of student growth models (not our focus, but addressed by other CKN briefs)
  - Large vendors use different approaches
  - May be a tradeoff between accuracy and transparency of measures (theory of action/usage of information & behavioral response matter)

Questions

1. What do we currently know about value-added teacher effect estimates generated by different models?
2. What more needs to be known about this issue (question 1)?
3. What can't be *directly* resolved by empirical evidence on this issue?
4. What are the practical implications of the research on this issue for decision making?

Background on Using Student Growth as a Metric for Teacher Performance

- All methods of estimating teacher performance (based on student test achievement) entail estimating student achievement
  - Important differences between models in the variables used to predict student test achievement (e.g., test scores, student characteristics, classroom factors), which yield different "expectations" of teacher "performance"
- VAMs generally have more covariate controls than SGPs so are thought to be a better causal measure, but also less transparent (teachers can't easily aggregate up)
  - Controversy over methods especially contentious around race/ FRL

How Correlated Are Teacher Effects Generated From Different VAMs

- Numerous studies show high correlations (over 0.9) between models with and without student covariates
- Lower correlations (around 0.5) between measures that compare teachers within and across schools to those that only compare teachers within schools
  - Unequal sorting of teacher effectiveness across schools and/or some school-wide effect being attributed to teachers
- Less is known about how SGPs/MGPs compare to VAM effectiveness measures of performance
  - SGPs typically do not directly account for student background (other than test performance) but may implicitly through functional form of model

What Do We Know? Impact of Model Choice on Teacher Effectiveness Measures in Math

		VAM with prior test score (Correlation = 0.97)	Student Growth Percentiles (Correlation = 0.91)		VAM within-school (Correlation = 0.55)	
		Q1 (Lowest)	Q2	Q3	Q4	Q5 (Highest)
VAM with prior test score and student covariates	Q1 (Lowest)	17.2%	2.7%	0.0%	0.0%	0.0%
	Q2	2.7%	13.7%	3.6%	0.1%	0.0%
	Q3	0.1%	3.4%	12.9%	3.5%	0.0%
	Q4	0.0%	0.2%	3.4%	13.8%	2.6%
	Q5 (Highest)	0.0%	0.0%	0.1%	2.6%	17.3%

What Do We Know? Impact of Model Choice on Teacher Effectiveness Measures in Math

		VAM with prior test score (Correlation = 0.97)		Student Growth Percentiles (Correlation = 0.91)		VAM within-school (Correlation = 0.55)	
		Q1 (Lowest)	**Q2**	**Q3**	**Q4**	**Q5 (Highest)**
VAM with prior test score and student covariates	**Q1 (Lowest)**	15.4%	4.2%	0.4%	0.0%	0.0%
	Q2	4.1%	10.2%	5.0%	0.6%	0.0%
	Q3	0.6%	5.0%	9.4%	4.7%	0.4%
	Q4	0.0%	0.9%	5.3%	10.2%	3.6%
	Q5 (Highest)	0.0%	0.0%	0.5%	4.2%	15.3%

What Do We Know? Impact of Model Choice on Teacher Effectiveness Measures in Math

		VAM with prior test score (Correlation = 0.97)		Student Growth Percentiles (Correlation = 0.91)		VAM within-school (Correlation = 0.55)	
		Q1 (Lowest)	**Q2**	**Q3**	**Q4**	**Q5 (Highest)**
VAM with prior test score and student covariates	**Q1 (Lowest)**	9.0%	5.6%	3.1%	1.6%	0.8%
	Q2	5.0%	5.4%	4.6%	3.3%	1.7%
	Q3	3.1%	4.4%	5.0%	4.5%	3.0%
	Q4	1.9%	3.0%	4.5%	5.5%	5.0%
	Q5 (Highest)	0.9%	1.5%	2.7%	5.2%	9.4%

What Do We Know? Teacher Effectiveness Measures and Classroom Characteristics

- Different teachers benefit from different modeling decisions
- Table shows the average percentile rank for teachers in advantaged and disadvantaged classrooms (as measured by poverty level and prior performance)
  - Teachers in advantaged classrooms tend to benefit when models do not control for student background

Panel 1: Math	**Advantaged**	**Disadvantaged**
Student Growth Percentiles	60.7	41.1
VAM with prior test score	65.1	38.2
VAM with prior test score and student covariates	57.8	47.7
VAM with prior test score, student, and classroom covariates	60.1	46.6
VAM with within-school comparison	51.9	48.7
Panel 2: Reading	**Advantaged**	**Disadvantaged**
Student Growth Percentiles	66.6	33.8
VAM with prior test score	71.8	29.0
VAM with prior test score and student covariates	58.2	43.6
VAM with prior test score, student, and classroom covariates	60.3	42.8
VAM with within-school comparison	51.0	49.4

What More Needs to be Known?

- Do the findings we have presented generalize across different states/contexts?
  - There is some evidence that the answer is no at the school level

- What student growth-based teacher evaluation systems lead to increased teacher effectiveness/student performance
  - Depends both on the model, how the information it generates is used, and teachers' behavioral response to model/use

What Can't Be Resolved Based on Empirical Evidence?

- Which measures are "fairest" to teachers and students?
  - Tradeoff of Type I versus Type II errors
- What is the right balance between accuracy and transparency for evaluation systems using student growth measures?
  - Depends on impacts of system and decisions about fairness
- Is it more appropriate to compare teachers only within schools or within and across schools?
  - Empirical models cannot entirely separate a school (e.g. principal, culture) contribution from a teacher contribution to student learning

What Are the Practical Implications of the Research on Decision-making?

- (In our subjective opinion) modeling choices do have important impact on teacher rankings
  - This is true in the classroom composition tails even when two models are highly correlated overall
- Policymakers should take the long-view when thinking about implementing new student growth-based evaluation systems as shifting rankings associated with models could undermine confidence in the system
  - Transparency about assumptions/tradeoffs is essential

Summary

- Correlations across many models in use/under consideration are high overall
  - High correlations can mask important implications of model choice for advantaged/disadvantaged classrooms
- When correlations are lower (in the case of within/ between school comparison choice), there is not a "right" answer about choice
- No empirical answer to some questions about fairness
- The impact on student achievement ultimately depends not only on model choice, but also use of measures and teachers' behavioral response