VALIDATING TEACHER EFFECTS ON STUDENTS' ATTITUDES AND BEHAVIORS: EVIDENCE FROM RANDOM ASSIGNMENT

David Blazar

University of Maryland College Park

November 2, 2017

Email: dblazar@umd.edu

Twitter: @dblazar

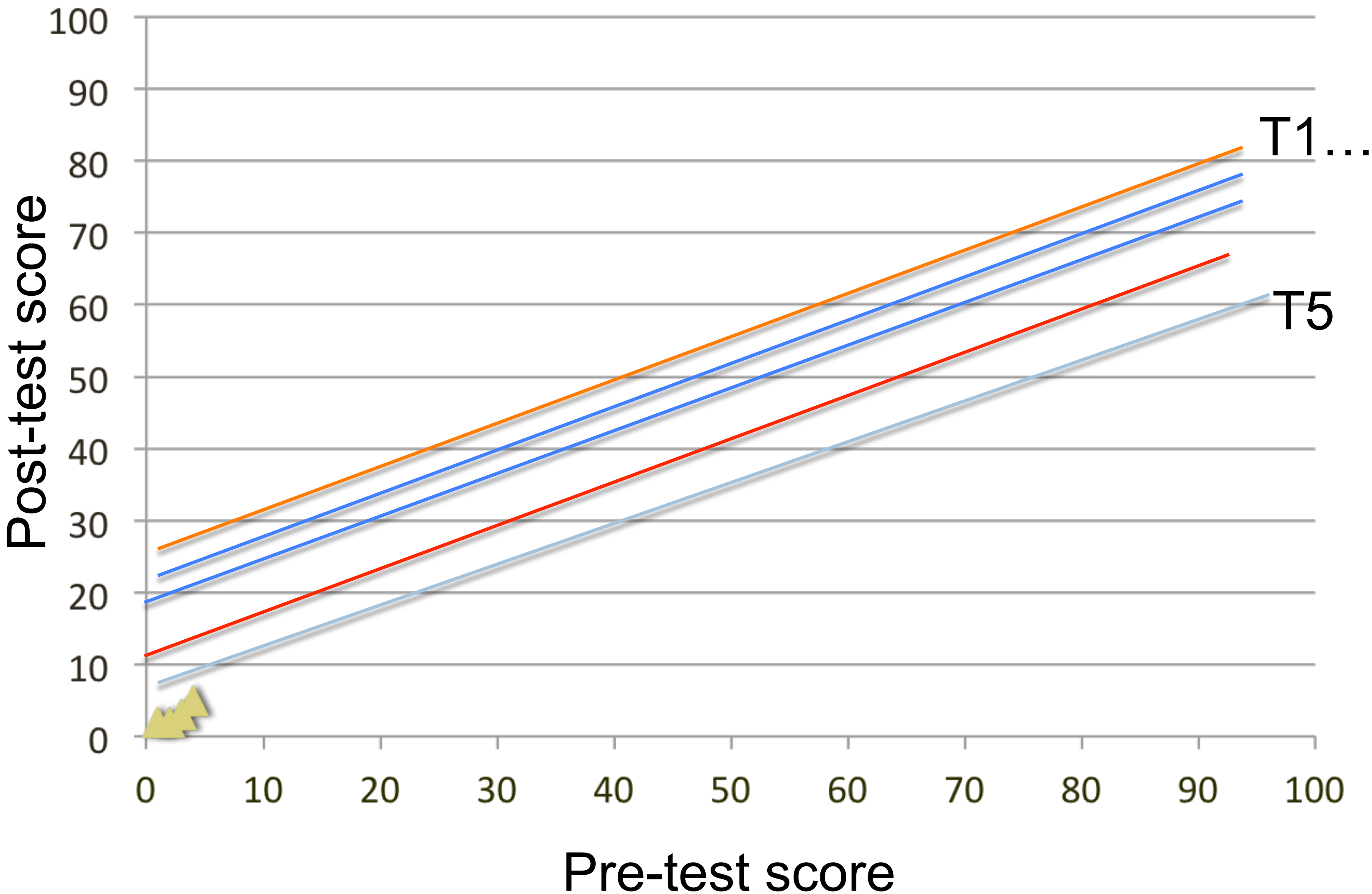
Research support came from the Mathematica Policy Research summer fellowship and the Smith Richardson Foundation.

Background and Motivation

Identifying effective teachers

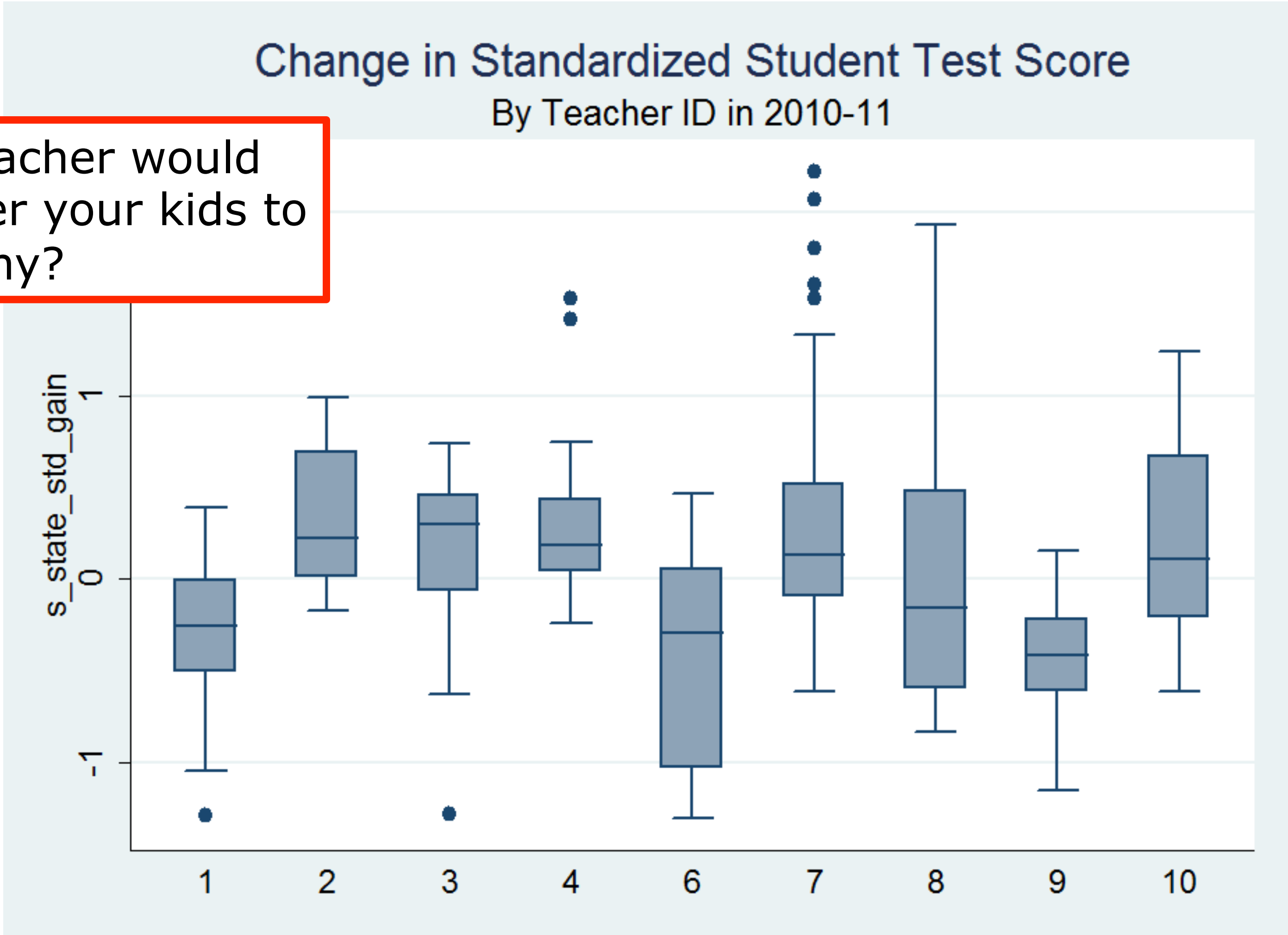
- One of the major outcomes of longitudinal data systems has been an ability to identify effective teachers using performance on the job.
  - Students tests in grades 3 through 8 in math and ELA, linked to teachers
- How do we do it?

Basics of "value-added" modeling

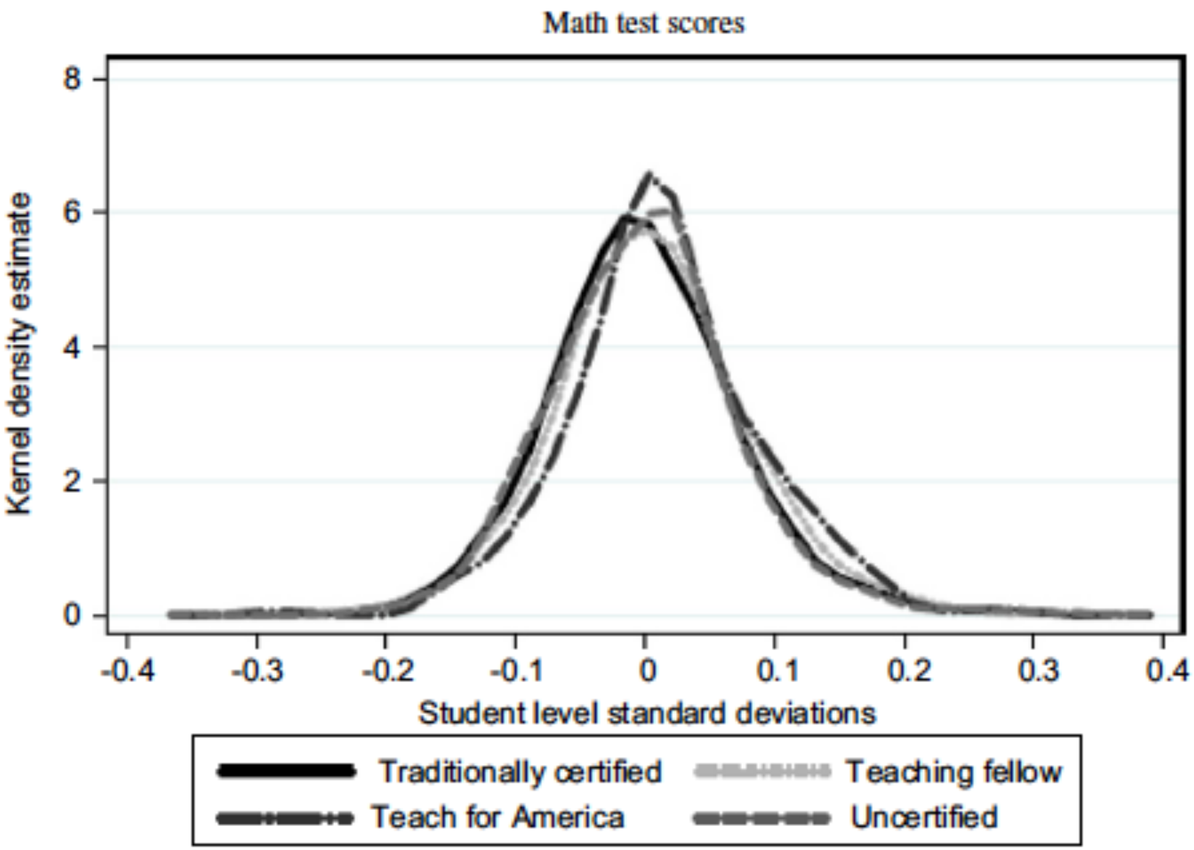


Gains in student performance vary considerably across teachers

Which teacher would you prefer your kids to have? Why?



Result: Distribution of teacher effectiveness



Note: Shown are estimates of teachers' Impacts on average student performance, controlling for teachers' experience levels and students' baseline scores, demographics and program participation; Includes teachers of grades 4-8 hired since the 1999-2000 school year.

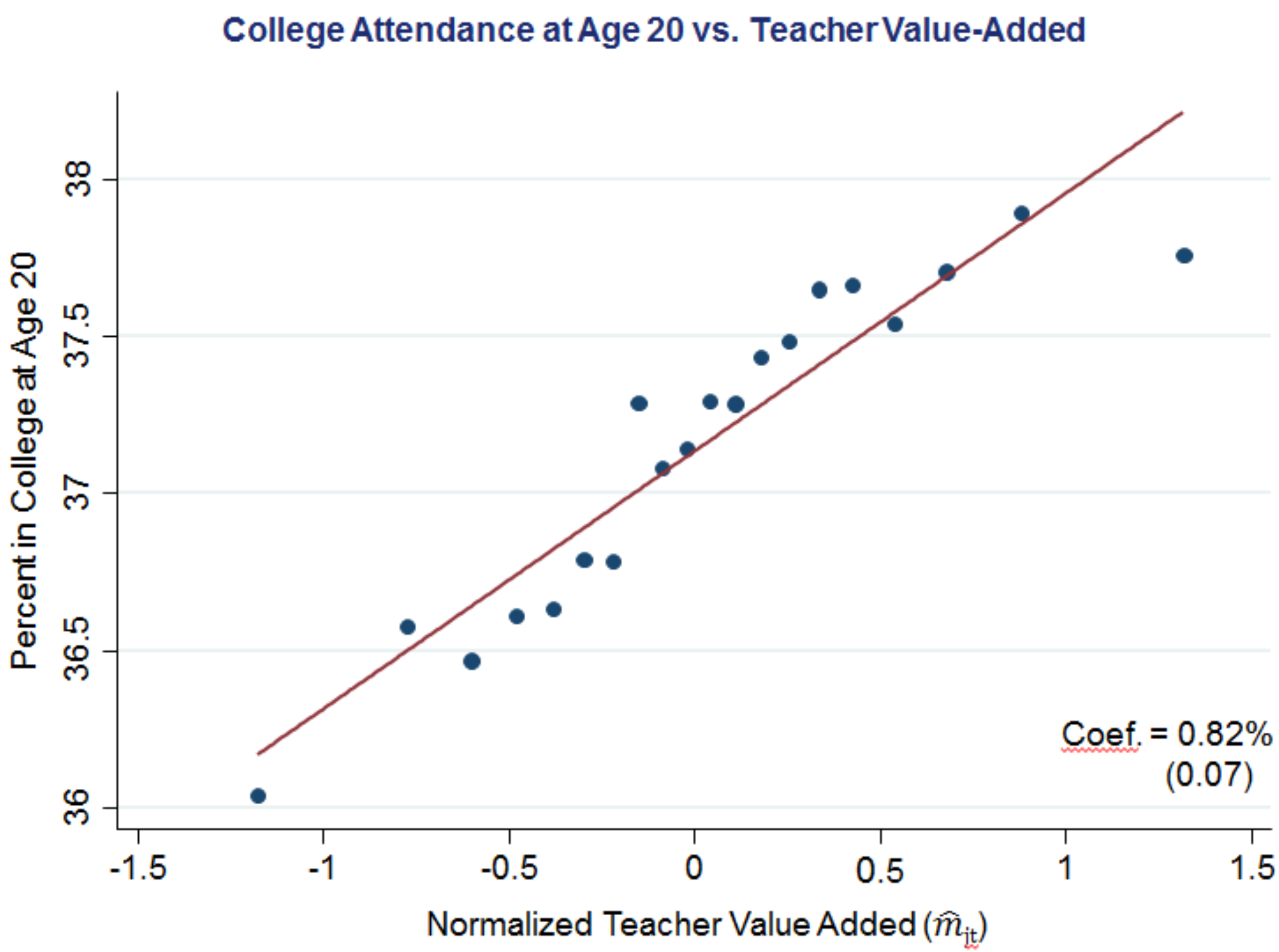
Source: Kane, Rockoff, & Staiger, 2006

What do we know from several decades of research?

- Teachers vary substantially in their “effects” on student test scores, which in turn influences a variety of long-term outcomes.

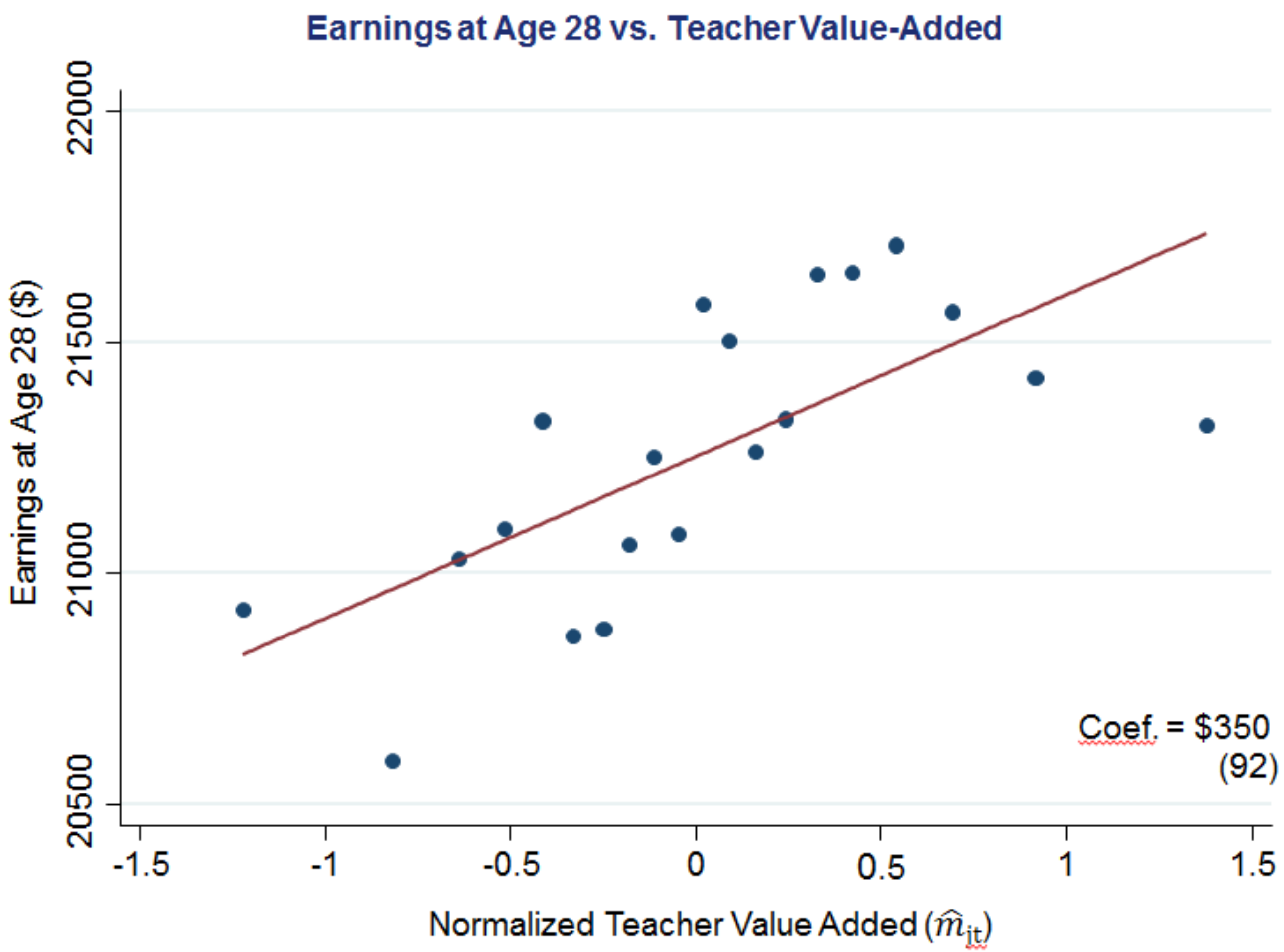
Evidence from Tax Data

Chetty, Friedman, and Rockoff, 2014



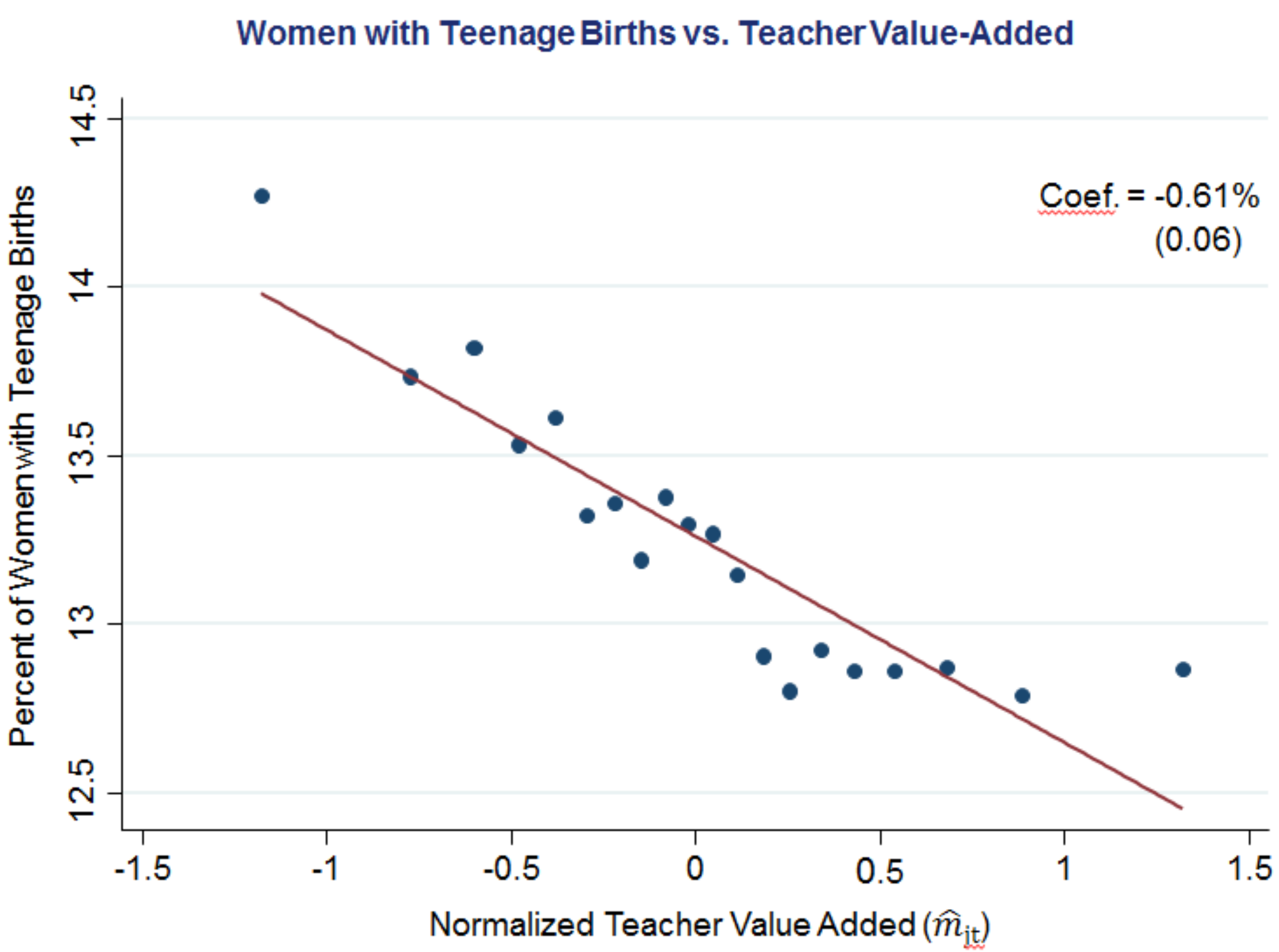
Evidence from Tax Data

Chetty, Friedman, and Rockoff, 2014



Evidence from Tax Data

Chetty, Friedman, and Rockoff, 2014

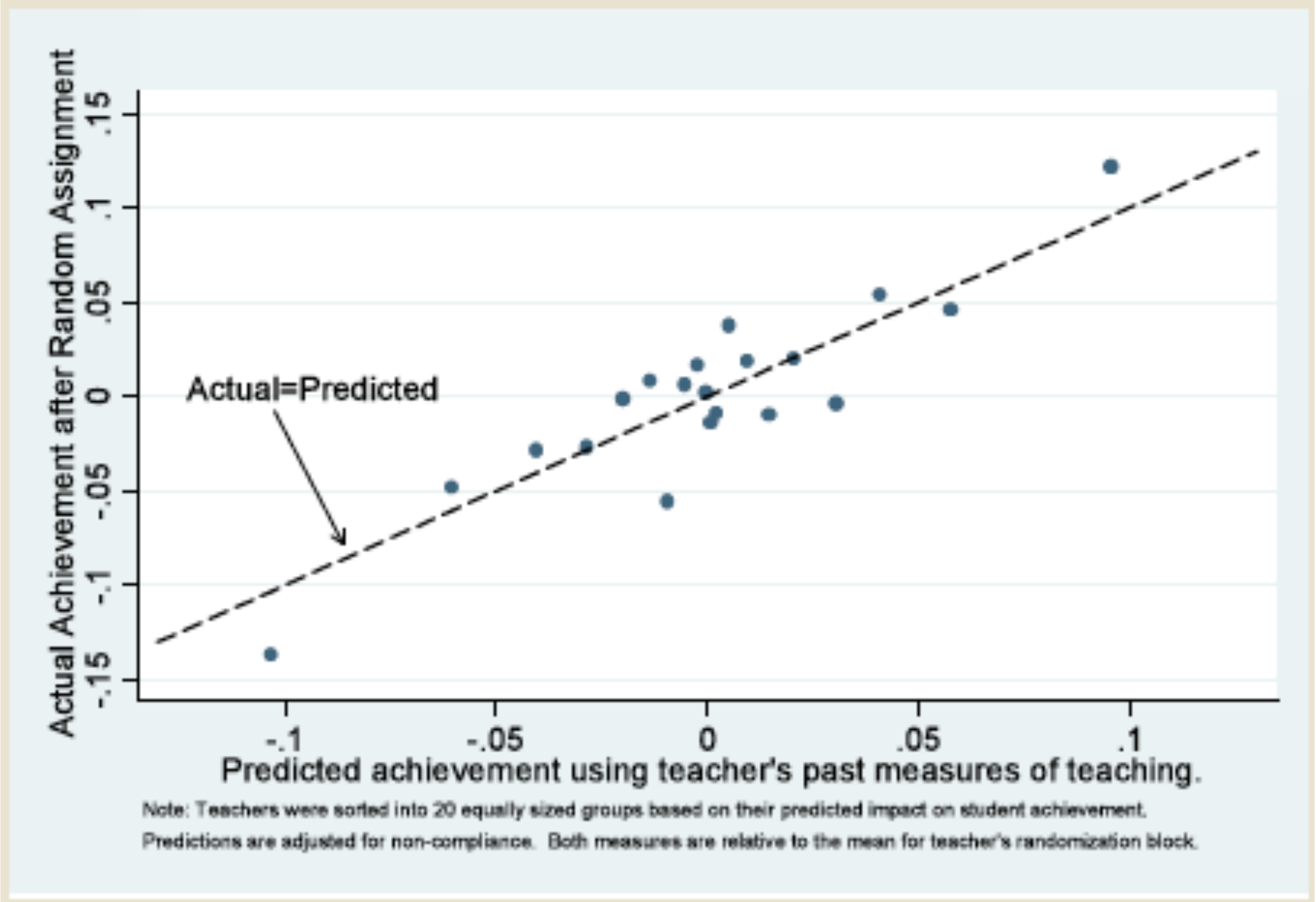


What do we know from several decades of research?

- Teachers vary substantially in their "effects" on student test scores, which in turn influences a variety of long-term outcomes.
- Value-added approaches produce estimates of teacher effects that are unbiased – i.e., not influenced by non-random sorting of students to teachers, principals' preferential treatment of teachers, and other factors beyond teachers control.

Evidence from Random Assignment

Measures of Effective Teaching (MET) Project, 2013



What do we know from several decades of research?

- Teachers vary substantially in their "effects" on student test scores, which in turn influences a variety of long-term outcomes.
- Value-added approaches produce estimates of teacher effects that are unbiased – i.e., not influenced by non-random sorting of students to teachers, principals' preferential treatment of teachers, and other factors beyond teachers control.
- Whether or not you agree with value-added approaches to identifying effective teachers, it is clear that research in this area has had a large influence on policy.
  - Incentives from the Obama administration to evaluate teachers using student achievement data. Still in effect in many states, despite ESSA regulations that feds no longer can create such incentives.
  - Pushes to evaluate teacher preparation programs using similar approaches.

Current Study

One concern with "value-added" to test scores is that it only captures single dimension of teacher quality

- My own experience as a classroom teacher: Effective teachers not only raise test scores but also manage the classroom environment, build positive relationships with students, etc.
- Growing interest among policymakers, researchers, and practitioners in using longitudinal data to identify teachers who are skilled at improving student outcomes beyond test scores.
- Observational estimates indicate that teachers do indeed vary in their contributions to a range of attitudes and behaviors captured on surveys and observed school behaviors.
- However, questions remain about the validity of these teacher effect estimates.

Research Questions

- In the value-added literature more broadly, researchers have asked about:
  - Sensitivity of teacher effects to different model specifications
  - Most appropriate ways to calculate these scores in light of measurement error
  - Bias in teacher effect estimates due to non-random sorting
- **How do teacher effects on students' attitudes and behaviors hold up to these tests?**
- **In turn, in which policy settings are these estimates most useful?**

Data

- National Center for Teacher Effectiveness study of upper-elementary math instruction
- Participants were 4th and 5th grade teachers in four school districts over the course of three school years (2010-11 through 2012-13)
- Administered survey asking about students' *Behavior in Class, Self-Efficacy in Math*, and *Happiness in Class*
- Administrative data include current and prior-year test scores, demographic characteristics
- In the third year, teachers (N = 41) were randomly assigned to class rosters within schools. Class rosters constructed by principals to be comparable.
  - Similar to the MET study, but on a smaller scale → much higher rates of compliance

Analyses

$$OUTCOME_{idsgjt} = \alpha f(A_{it-1}) + \zeta OUTCOME_{it-1} + \pi X_{it} + \varphi \bar{X}_{it}^{c} + \varphi \bar{X}_{it}^{s} + \boxed{\varepsilon_{idsgjt}}$$

- Examine the extent to which teachers vary in their contribution to students' attitudes and behaviors, even after random assignment.
- Examine the sensitivity of teacher effects on students' attitudes and behaviors to different model specifications, including those that control for students' prior academic performance versus prior attitudes and behaviors.
- Examine whether non-experimental estimates of teacher effects on these attitudes and behaviors predict these same outcomes following random assignment. → If no bias, there should be 1:1 relationship.

Questions Before Results?

Results (1)

Teachers vary in their contributions to students' attitudes and behaviors in addition to their math performance, even after random assignment.

Put all metrics on a common scale (i.e., standard deviation units)

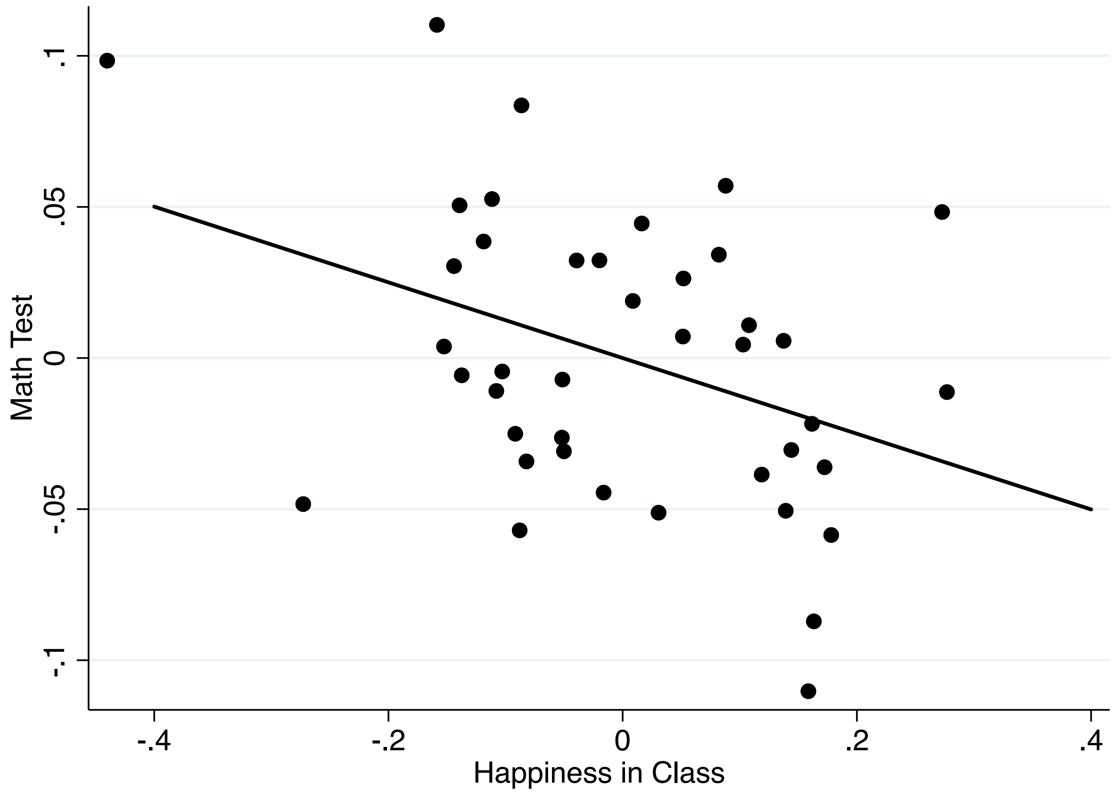
	Standard Deviation of Teacher-Level Variance
State Math Test	0.13
Behavior in Class	0.05
Self-Efficacy in Math	0.08
Happiness in Class	0.34
Teachers	41
Students	531

A 1 standard deviation (SD) increase in teacher effectiveness results in a 0.13 SD increase in students' math test scores.

Relative to the average teacher, teachers at the 84th percentile in the distribution of effectiveness move the median student up to the 55th percentile of math performance.

Results (2)

Teachers who improve test scores often are not the same as those who improve students' attitudes and behaviors.



Results (3)

In observational data, teacher effects on students' attitudes and behaviors are not particularly sensitive to controlling for prior achievement versus prior attitudes and behaviors.

Pairwise Correlations Between Teacher Effects Across Model Specifications

	ρ _(Model 1,Model 2)	ρ _(Model 1,Model 3)	ρ _(Model 2, Model 3)
Teacher Effects on Behavior in Class	0.90***	0.91***	1.00***
Teacher Effects on Self-Efficacy in Math	0.86***	0.90***	0.97***
Teacher Effects on Happiness in Class	0.96***	0.96***	0.99***

Notes: ~ p< .10, * p<.05, ** p<.01, ***p<.001. Model 1 calculates teacher effectiveness ratings that only control for students' prior achievement in math and reading. Model 2 only controls only for a prior measure of students' attitude or behavior. Model 3 controls for prior scores on both prior achievement and prior attitude or behavior. Samples includes 51 teachers.

Model 1: Control for prior achievement

Model 2: Control for prior attitude or behavior

Model 3: Control for both

Results (4)

Non-experimental methods for estimating teacher effects on students' attitudes and behaviors have predictive validity following random assignment but do not remove bias in all cases.

Remember – we are looking for a 1:1 relationship between non-experimental and experimental teacher effect estimates.

Relationship Between Current Student Outcomes and Prior, Non-Experimental Teacher Effect Estimates

	State Math Test	Behavior in Class	Self-Efficacy in Math	Happiness in Class
Teacher Effects Calculated from Model 1	0.960***	1.003***	0.514	0.427*
	(0.078)	(0.266)	(0.369)	(0.177)
Teacher Effects Calculated from Model 4	0.995***	1.090***	0.507	0.438*
	(0.084)	(0.268)	(0.372)	(0.175)
Teacher Effects Calculated from Model 5	1.055***	1.240***	0.557	0.416*
	(0.100)	(0.305)	(0.404)	(0.167)
Teacher Effects Calculated from Model 6	1.079***	1.472***	0.557	0.487**
	(0.101)	(0.368)	(0.410)	(0.174)
Teacher Effects Calculated from Model 7	1.084***	1.789***	0.582	0.522**
	(0.102)	(0.458)	(0.389)	(0.172)
Teachers	41	41	41	40
Students	531	531	531	509

Notes: ~ p< .10, * p<.05, ** p<.01, ***p<.001. Cells include estimates from separate regression models that control for students' prior achievement in math and reading, student demographic characteristics, classroom characteristics from randomly assigned rosters, and fixed effects for randomization block. Robust standard errors clustered at the class level in parentheses. Model 1 calculates teacher effectiveness ratings that only control for students' prior achievement in math and reading; Model 4 adds student demographic characteristics; Model 5 adds classroom characteristics; Model 6 adds school characteristics; Model 7 replaces school characteristics with school fixed effects.

Where Do We Go From Here?

Synthesis of Results

- Teachers have causal effects on their students' self-reported behavior in class, self-efficacy in math, and happiness in class.
- Weak correlations between teacher effects indicate that these measures capture unique skills that teachers bring to the classroom.
- Teacher effects calculated in non-experimental data are related to these same outcomes following random assignment, revealing that they contain important information content on teachers.
- However, for some non-experimental teacher effect estimates, large and potentially important degrees of bias remain.
- Teacher effects are not particularly sensitive to different sets of control variables. Given that these are the tools and data typically available to the econometrician, not clear what else could be used to reduce bias.

Policy Implications

- How might teacher effects on students' attitudes and behaviors be used in light of bias?
- In high-stakes policy settings?
  - Some (not me) may argue: Moderate relationships between non-experimental and experimental estimates indicates that a teacher de-selection policy using biased measures still would improve outcomes on average.
  - Incorporating these measures would create clear incentives for improving these skills in school.
  - <u>However</u>, we already have observed substantial pushback from teachers and schools for using unbiased measures of teacher effects on test scores.
- Biased measures may be less concerning when used for professional growth and allocation of professional development resources. Principals could use these measures to:
  - Identify teachers most in need of support.
  - Identify specific skills teachers need to improve in.
  - Create matches between these teachers and a coach, teacher mentor, PD services, etc.

Implications for MD/MLDS

- Longitudinal data systems are the cornerstone of this line of research → many opportunities in MD through MLDS!
- Connecting teachers to students and then to student outcomes is feasible.
  - Big hurdle is cleaning the teacher-student links through course data.
- While MLDS doesn't have access to student survey measures, it does have observable school behaviors (e.g., absences) that also fall into the "non-tested outcome" bucket.
- In turn, can answer several teacher-related questions:
  - Are effective teachers distributed in the districts and schools that need them most?
  - In which areas of practice do MD teachers most need improvement?
  - Which resources best support teacher improvement efforts?

Thank you!

Website: https://www.education.umd.edu/directory/david-blazar

Email: dblazar@umd.edu

Twitter: @dblazar