Title: Learning that Lasts: Understanding Variation in Teachers' Effects on Students' Long-Term Knowledge

Authors: Benjamin Master[1], Susanna Loeb[1], James Wyckoff[2]

Affiliations:

[1]Stanford University

[2]University of Virginia

Abstract: Measures of teachers' "value added" to students' current test performance feature prominently in ongoing reforms to teacher evaluation systems. However, this immediate effect may not capture teachers' more meaningful longer-term impact on student learning. Set in New York City, this study investigates the persistence of teachers' effects. Two findings emerge. First, a teacher's effect on students' English Language Arts achievement has substantial crossover effects on long-term math performance. Second, the persistence of teachers' value-added is considerably lower in schools that serve more low-achieving students or that hire fewer academically skilled teachers. The results provide evidence that teachers' effects on students' long-term skills can vary substantially as a function of both instructional content and quality.

Main Text:

Teachers play an important role in students' academic achievement gains during the year students are in the teachers' class (*1*). Policymakers and researchers frequently operationalize teacher effectiveness empirically as value-added. Value-added attempts to statistically isolate a teacher's ability to improve students' performance, often using the year-over-year gains in individual student standardized achievement tests in math and English Language Arts (ELA), controlling for other factors that may influence achievement. Value-added has proven reasonably robust in many applications (*2, 3*) and is increasingly widely employed in practice (*4*), although most stakeholders agree it has important limitations (*5, 6*).

Researchers have expended substantial effort towards understanding teachers' short-term effects on student achievement as measured by value-added; however, little effort has gone into understanding what constitutes effective teaching for learning that accumulates and persists over time. On average, teachers who improve student test performance also benefit students' long-term life outcomes, such as college attendance and quality, lifetime income, and the likelihood of becoming a teen parent. For example, Chetty, Friedman and Rockoff (*7*) find that a one standard deviation improvement in value-added for just one year increases students' earnings, on average, by 1.3 percent by age 28. However, not all teachers who add value to students short-term test performance also benefit students in the long run. Some teachers may appear effective using value-added measures merely by teaching short-run test taking strategies and providing information tailored to the particular test, while other teachers who focus on high-order learning skills may appear less productive immediately but their students may develop skills that promote the accumulation and persistence of knowledge over time, resulting in greater total achievement. Recent research finds that only a small portion of measured teacher value-added persists (*8-11*), and that teachers' proximal-year effects on achievement are only moderately correlated with their longer-term effects (*8*, *12*, *13*).

In this study we assess the extent to which teachers' short-run effects on student test performance lead to longer-term achievement benefits for students. We look at the effect of value-added in math and ELA both on students' longer-term performance in that same subject and on their cross-subject learning. We also assess heterogeneity in the persistence of teachers' contributions to student achievement asking whether teachers with stronger qualifications show greater persistence and whether teachers of higher achieving students or in schools serving more high achieving students show greater persistence. The results shed light on the variance in long-run effects of teachers, the importance of proximal ELA learning on long-run math achievement, the inequities in long-run learning across schools, and, more generally, the importance of considering persistence when measuring teachers' effectiveness.

Background

We use a model of student learning, following Jacob et al. (*9*), that distinguishes short-term, test specific, knowledge from longer-term learning that accumulates. In general, achievement in period t is a function of long-term (l) and short-term (s) learning.

$$Y_t = y_{l,t} + y_{s,t} \tag{1}$$

Longer-term learning persists over time but may do so only partially at rate δ and includes additions to long-term learning by the current teacher, θ, as well as other factors, η. We extend this model to include both long-term learning that is subject or context specific and other learning that transcends subjects, developing generic skills useful in many contexts. Thus both prior and contemporaneous learning have context specific (c) and generic (g) components.

$$y_{l,t} = \delta_c y_{l,c,t-1} + \delta_g y_{l,g,t-1} + \theta_{l,c,t} + \theta_{l,g,t} + \eta_{l,t} \quad (2)$$

While, by definition, short-term learning completely decays by the next period, $\delta=0$, it includes short-term teacher inputs, $\theta_{s,t}$, and idiosyncratic factors η. Contemporaneous teacher inputs that influence long or short-term learning may overlap but need not.

$$y_{s,t} = \theta_{s,t} + \eta_{s,t} \quad (3)$$

$$Y_t = \delta_c y_{l,c,t-1} + \delta_g y_{l,g,t-1} + \theta_{l,c,t} + \theta_{l,g,t} + \theta_{s,t} \quad (4)$$

Most would agree that individuals and society value the development of knowledge and skills that lead to greater individual and social well-being, not necessarily the maximization of test performance in a specific subject or in a specific grade. Policymakers use standardized achievement tests as proximal measures of whether students have mastered skills linked to learning standards that accumulate over time, with the goal that realization of these learning standards provides students with knowledge and skills useful beyond high school. Yet, teachers are time constrained and often encouraged to improve student performance on annual exams. To the extent that teachers view short-term learning as more productive in realizing improved scores on this year's standardized tests, they may favor the development of context specific short-term knowledge. In the extreme, short-term learning may dominate such that students excel on the test, but know little that increases their performance next year or beyond.

Our understanding of the factors associated with the development long-term learning is limited. There is some evidence that English skills are particularly important. Chetty et al. (*7*) find that teacher-driven increases in students' English language arts (ELA) test scores predict improvements in life outcomes 1.7 times as large as comparable teacher-driven improvement in math test scores. ELA skills may be particularly relevant to students' long-term success, perhaps because they are foundational to student learning in other subject areas. There is also some evidence that learning persists more for some students than others. Chetty et al. (*7*) find that students from higher and lower income families make equivalent learning gains when taught by equally high value-added teachers, but that teacher effects on later life outcomes are greater for higher income students. However, Jacob et al. (*9*) do not find meaningful variation in persistence as a function of students' race or free-lunch status in North Carolina.

Persistence of teachers' effects on student academic achievement may arise through mechanisms associated with the instruction, the students or the tests. First, teachers' instructional practices may vary, with some teachers emphasizing students' long-term content specific knowledge and/or long-term generic knowledge, while others focus on short-term tested knowledge, e.g., teaching to the test. Differences in teachers' instructional practices may reflect differences in teachers' abilities or preferences, or they may reflect contextual factors, such as school accountability, that create pressure for short-run gains in achievement. Second, students may differentially forget long-term knowledge, regardless of the quality of instruction that they initially received. Such differential forgetfulness could result from students' innate abilities or from instructional contexts that support different rates of knowledge retention. Similarly, students may also differ in their ability to acquire long-term knowledge. For example, some students understand the underlying principles, while others memorize the outcome and merely develop a superficial grasp of the material. Finally, we would expect persistence to mechanically vary according to the degree of overlap in content between one test administration and the next.

Employing administrative data on students and teachers in New York City we examine the persistence of learning in ELA and mathematics across a variety of teachers and school contexts to better under the extent and nature of learning persistence. More specifically, we consider the following research questions:

1) What is the persistence of teachers' value-added within and across subjects?
2) To what extent does persistence vary as a function of teacher or student attributes?
3) To what extent does variation in persistence stem from the broader school context?

Students, Teachers, and Tests in NYC

We employ administrative data about students, teachers, classrooms, and schools from the New York City Department of Education (NYCDOE) and the New York State Education Department (NYSED). The data include information on students and their teachers in 3rd through 8th grade in school years (SY) 2003-04 through SY 2010-11. Because the analysis focuses on the persistence of teacher value-added effects from a prior school year, our analytical sample is limited to students in grades 5 through 8. A summary of these data are shown in Table 1. Most are self-explanatory, but we note the heterogeneity of students and teachers. For example, about 70 percent of students are eligible for free or reduced price lunch, a measure of low income, and about two-thirds of students are black or Hispanic.

The content assessed by the annual New York state ELA and math student achievement tests has been aligned with the state's content standards in grades three through eight, and exams in each subject area include a mix of multiple choice and open response questions. The ELA exams primarily assess students' comprehension of reading passages and writing ability, while math exams address a range of topics including number sense, algebra, probability, and geometry, with overlapping topics across grades.

Short-term and Long-term Teacher Effects on Achievement

Teacher Value-Added Measures

To examine variation in the persistence of teachers' effects, we first estimate the annual contribution of each teacher to their students' gains on standardized achievement, employing an empirical model that many states and school districts, including NYC, use (*14*). Conceptually, this model compares teachers to other "similarly circumstanced" teachers by first predicting students' achievement with both prior achievement measures and a range of observable student, classroom, and school characteristics that may influence their achievement, and then attributing the remaining unexplained variation in student performance to individual teachers. Because the method by which value-added is calculated may affect persistence estimates, we also consider two other approaches which make alternative assumptions in apportioning student learning gains between teachers and other sources. The details of all three value-added models may be found in the Technical Appendix (*15*).

Estimating the Persistence of Teacher Value-Added Effects

Next, we estimate the persistence of teachers' value-added using an instrumental variables approach described by Jacob et al. (*9*). As shown in equation 4 above, attempting to recover estimates of long-term knowledge by regressing current achievement on lagged achievement will be biased due to the presence of short-term knowledge. Jacob et al. (*9*) use twice lagged achievement Y_{t-2} as an instrument for Y_{t-1}, purging the short-term knowledge component. This approach demonstrates that nearly all of a student's previously assessed long-term knowledge continues to persist across time (i.e., the δ nearly equal 1) (*16*).

Following a similar approach, we can estimate the proportion of teachers' effects that generate long-term knowledge by instrumenting each student's lagged knowledge Y_{t-1} with their lagged teacher's contribution to that knowledge (value-added). The prior-year teacher's total contribution to a student's lagged knowledge is a combination of her contribution to lagged long-term knowledge (context specific and general) and her contribution to short-term lagged knowledge, expressed as $M_{t-1} = \mu_{l,c,t-1} + \mu_{l,g,t-1} + \mu_{s,t-1}$. Given that the persistence of long-term knowledge, δ, is close to 1, $\hat{\delta}_{Mc}$ and $\hat{\delta}_{Mg}$ approximate the fraction of teacher value-added that is attributable to long-term context specific and long-term generic, rather than short-term, knowledge creation. The Technical Appendix (*15*) provides details of the estimation strategy for these parameters.

Our approach to measuring persistence does not generate individual teacher-level persistence estimates, but instead estimates the average persistence of teacher value-added effects across a particular sample of students. While this approach presents some limitations, a key benefit of our approach is that it allows us to estimate teachers' value-added persistence both within and across subject areas.

Results

The Persistence of Teacher Effects Within and Across Subjects

We find that approximately 20 percent of a teacher's within-subject (context specific) contributions to students' learning persists into the following school year for both ELA (0.19) and math (0.20) (Table 2). Consistent with an intuitive understanding of long-term knowledge, we also find that nearly all of previously assessed long-term knowledge (i.e. knowledge that is relevant across two prior school years) also persists into a third year for both subjects (0.95 in ELA and 0.93 in math). Each of these estimates is consistent with those found in the literature (*8-11*). On average, only a moderate portion of teachers' short-term influence on student achievement persists beyond the proximal year, but learning effects that do persist into the next school year subsequently decay slowly (*7, 9, 12, 17*).

In contrast to the similarity of context specific (within subject) persistence of teacher effects between math and ELA, teachers' contributions to generic learning differ substantially across subjects. Sixteen percent of the gains attributable to ELA teachers persist in the long-term gains of students in math. This crossover is notable in two respects. First, it is more than 80 percent of the within-subject long-term persistence for ELA (0.16/0.19), suggesting that instruction in ELA may be mostly generic, generalizing across subjects, and less context specific. In contrast, only four percent of the gains attributable to math instruction persist in the long-term gains of students in ELA, suggesting that instruction in math is much more context specific with relatively little generic learning.

The differential persistence of *teacher* effectiveness across contexts contrasts with the similarity of cross-subject, long-term *student* knowledge for ELA and math. Previous long-term ELA knowledge is associated with a 0.64 increase in math achievement, while long-term math knowledge predicts a 0.61 increase in ELA achievement. The abilities or skills that allow students to perform well in one subject also predict success to a reasonable degree in the other subject, suggesting that many abilities or skills are portable across subjects. ELA teachers appear to develop more of this generic long-term knowledge in students than do math teachers.

Heterogeneity in the Persistence of Teachers' Effects

The persistence of ELA teachers' effects on future achievement varies greatly across teachers, even among teachers with equivalent short-run effectiveness. Much of this heterogeneity appears driven by variation in school-level characteristics.

Heterogeneity across teacher and student attributes. As shown in table 3, the persistence of ELA teachers' long-term contributions to student achievement varies significantly across a variety of observable teacher characteristics, with academically more able teachers demonstrating greater persistence. For instance, the within-subject value-added persistence of ELA teachers who attended a more competitive undergraduate institution exceeds 26 percent while that for teachers who attended a less competitive institution is just 16 percent. Differences in persistence are similarly large for ELA teachers when comparing teachers whose SAT Verbal exam scores or LAST licensure exam scores are in the top third of the teacher distribution, in comparison to all other teachers.

Observable student characteristics also predict substantial variation in the persistence of their ELA teachers' contributions to their achievement. The persistence of achievement gains from being taught by a higher value-added teacher is far lower for students who are eligible for free lunch (a measure of poverty), are black or Hispanic, or whose twice-lagged ELA achievement scores are below the mean. For example, same-subject ELA value-added persistence for students who are eligible for free lunch is approximately 15 percent, in comparison to 27 percent for other students.

The persistence of teacher effectiveness varies less in math than in ELA. In analyses similar to those in Table 3, differences are small, e.g., frequently less than 0.02 and rarely significant (*18*). The lack of variability in teacher value-added persistence in math may be due to the content of the math exams. If long-term knowledge in math consists primarily of content that is explicitly assessed over multiple grade levels, teachers' effects may persist similarly even when instruction focuses narrowly on tested content.

Heterogeneity in Persistence Across Schools. To investigate whether observed differences in persistence are due to instructional factors at the school level, we examine whether school-level characteristics predict differences in persistence, independent of teacher and student characteristics. Variation in persistence across schools, rather than across student characteristics within similar schools, would provide evidence that school-level context may be a driver of differential persistence. For example, school-level differences may reflect school-wide curriculum and instructional practices that influence the degree to which teachers focus on either short- or long-term knowledge. To examine this possibility, we compare ELA value-added persistence across each teacher and student characteristic of interest, within samples of schools that rate either above or below the mean in terms of school-level averages for the same characteristic. Our goal is to identify whether school-level characteristics explain much of the variation that we observe in value-added persistence, distinct from associations with the characteristics of the individual teachers or students within those schools. In these analyses, we exclude schools where the student or teacher population is extremely homogenous (i.e. >95% or <5%) with regard to our characteristic of interest.

As shown in Table 4, in samples of schools that are "better-staffed" or that serve fewer traditionally at-risk students, we find that both teacher and student-level characteristics continue to predict some differential ELA persistence independent of school characteristics. For example, within a sample of schools where more than a quarter of teachers attended a competitive undergraduate institution, persistence remains significantly higher for those students whose teacher attended a competitive institution (0.32) than for students whose teacher did not (0.23). The same pattern holds true for student-level characteristics. For example, within a sample of schools that primarily serve high-achieving students, the persistence of teacher effectiveness for previously low-achieving students (0.13) was significantly lower than for previously high achieving students (0.27) (*19*).

In marked contrast to the results for better-staffed schools and for schools serving more advantaged students, schools that have few high-ability teachers or that serve more poor, black, Hispanic, or previously low-achieving students demonstrate low value-added persistence across all individual teacher and student characteristics. For example, in a sample of schools where greater than 50 percent of students were previously low scoring, students of all ability levels show very low persistence of learning gains from their prior teachers' value-added. The coefficient for persistence within lower-achieving schools is 0.08 for previously low-scoring students, not significantly different from the 0.09 for previously high-scoring students. Similarly, in schools where less than a quarter of the teaching staff attended competitive undergraduate institutions, persistence is low regardless of the characteristics of the teacher a student is assigned. In these less-competitively staffed schools, the persistence coefficient for teachers who graduated from competitive institutions is just 0.10, which is nearly identical to the persistence coefficient for teachers from less competitive institutions (0.09). Within these schools, there is no evidence of variation in persistence as a function of either individual teacher or student characteristics. Instead, all teacher value-added effects in these schools exhibit similarly low rates of persistence.

The associations between value-added persistence and schools' student and teacher characteristics represent distinct trends. Both the composition of a school's students and the makeup of their teaching staff independently predict differences in persistence rates. As shown in Table 5, NYC schools that serve a majority of students with above average prior scores and who also hire an above-average proportion of teachers who attended competitive undergraduate institutions have the highest persistence rates, with a coefficient of 0.31. Schools that fit only one of those two criteria have moderate rates of persistence. Schools that serve primarily students with low prior achievement and who also hire few teachers from competitive institutions have persistence rates that are indistinguishable from zero, with a coefficient of 0.01. Thus, for more than one quarter of students in this very large district, measures of ELA teachers' value-added effectiveness show no persistence—they provide no information about their students' long-term academic performance.

Overall, the variation in persistence that we observe as a function of school-level characteristics suggests that school-wide curricular or instructional factors may be influencing students' long-term knowledge gains in ELA. This influence is most dramatically apparent in the very low persistence that we observe in schools serving high proportions of low-scoring students, even among previously high-achieving students. The persistence of value-added for those high-achieving students is substantially lower than even that of previously low-achieving students who attend high-achieving schools. If differences in student characteristics were the primary driver of differences in value-added persistence, then previously high-scoring students should have retained more long-term knowledge regardless of the quality of instruction that they received. Instead, our results are consistent with a hypothesis that instructional quality at the school level is the primary mechanism driving differences in ELA teachers' value-added persistence.

Conclusions and Discussion

Test-based accountability in K-12 education has led to a focus on short-term teacher effectiveness, potentially obscuring the mechanisms by which teachers create lasting impacts for students or causing us to overlook important heterogeneity in their capacity to do so. In a low-stakes context, teachers' value-added to short-term achievement has been shown to be predictive of students' life-long outcomes (*7*). However, it is unlikely that the delivery of transient or test-specific achievement gains is the primary mechanism by which teachers improve students' lives. Research on teacher effectiveness indicates that only a portion of the knowledge and skills that teachers impart yields persistent academic gains for students (*8-11*). In this study, we expand upon the extant research addressing teachers' contributions to academic learning that persists over time, distinguishing context specific knowledge from general, cross-subject knowledge and describing differences across students, teachers and schools.

Our investigation yields two main findings. First, the knowledge that English language arts teachers impart to students persists not only on future ELA exams, but also on future Mathematics exams. The same is not true for math teachers who have negligible effects on students' future ELA achievement. This finding is consistent with findings from Chetty et al. (*7*) that show that differential student learning from teachers in ELA has larger effects on college quality than does learning from teachers in math. Second, we find that there is substantial variation in the persistence of ELA teachers' effects across different school contexts in NYC. Schools that serve predominantly poor, minority, or low-scoring students, or that hire fewer academically skilled teachers have substantially lower value-added persistence, distinct from variation at the teacher or student level.

The research presented here has three primary implications for policy. First, instruction in English is crucially important. The skills developed by ELA teachers and measured on ELA exams have meaningful long-term effects on students' learning in both math and ELA. Failure to account for the development of generic learning effects of ELA teachers would lead to an undervaluing of ELA teachers in students' success. Second, some teachers and schools develop long-term learning more effectively than others. As a result, a substantial portion of teachers and schools may in fact be less effective than they appear to be, when they are assessed only in terms of students' short-term test performance. Similarly, some teachers may in fact be more effective for long-term learning than they appear to be using measures of proximal value-added. Failure to accurately identify teacher and school quality is not only inefficient, but could send perverse signals to teachers and school leaders about how best to support students. The substantial heterogeneity that we observe in ELA teachers' value-added persistence underscores the importance of closely monitoring the relationship between short- and long-term effects, particularly when high stakes are attached to measures of teachers' short-term effectiveness. Third, and finally, school–level factors are crucial to the development of long-term learning. In this district, schools that serve more disadvantaged students or that hire fewer academically skilled teachers have drastically lower value-added persistence in ELA for all of their students. In particular, students – regardless of their own prior ability levels – who attend schools with many low-performing students, demonstrate lower persistence of learning gains from having a high value-added teacher. The persistence of teachers' effects in low-achieving schools is less than half the rate of that in other schools. In light of prior research on educators' responses to high stakes accountability pressures (*20*) one plausible explanation for our findings could be that under-resourced schools and those serving lower-performing students systematically prioritize gains in short-term tested achievement in ways that detract from teachers' focus on long-term knowledge generation.

The utility of test-based measures of teacher performance depends in large part on the presumption that short-term instructional effects will correspond to lasting benefits for students. Expanding upon previous research (*8,12*), we demonstrate that this presumption is uncertain at best. Teachers that appear similarly effective according to short-term or subject-specific performance can differ substantially and systematically in their longer-term academic impacts on students. The variation that we observe in value-added persistence reflects drastic differences in the quality of student learning, with lower quality, short-term learning apparent in low-achieving and less competitively staffed schools in particular. As policymakers increasingly seek to evaluate and respond to differences in teacher performance, these findings highlight the importance of better understanding which specific knowledge and skills are most relevant to students' long-term success, and which instructional practices contribute most effectively to learning that lasts.

References and Notes:

1. S. G. Rivkin, E.A. Hanushek, T. F. Kain. Teachers, schools, and academic achievement. *Econometrica*, *73*(2), 417-458 (2005).

2. T. J. Kane,D.F. McCaffrey, T. Miller, D.O. Staiger. Have we identified effective teachers? Validating measures of effective teaching using random assignment. (*Bill and Melinda Gates Foundation*, Seattle, WA, 2013).

3. T. J. Kane,D.F. D.O. Staiger. *Estimating teacher impacts on student achievement: An experimental evaluation* (National Bureau of Economic Research, working paper 14607, 2008).

4. J. Hull. *Trends in Teacher Evaluation: How states are measuring teacher performance* (Center for Public Education Report, 2013). Retrieved on August 1st, 2014 at http://www.centerforpubliceducation.org/Main-Menu/Evaluating-performance/Trends-in-Teacher-Evaluation-At-A-Glance/Trends-in-Teacher-Evaluation-Full-Report-PDF.pdf.

5. S. P. Corcoran. Can Teachers Be Evaluated by Their Students' Test Scores? Should They Be? The Use of Value-Added Measures of Teacher Effectiveness in Policy and Practice. Education Policy for Action Series. (*Annenberg Institute for School Reform at Brown University (NJ1), 2010)*.

6. D.F. McCaffrey, J.R. Lockwood., D. M. Koretz, L. Hamilton. *Evaluating Value-Added Models for Teacher Accountability. Monograph*. (RAND Corporation. PO Box 2138, Santa Monica, CA 90407-2138, 2003).

7. R. Chetty, J. N. Friedman, J. E. Rockoff. (2013). *Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood.* (National Bureau of Economic Research working paper 19424, 2013).

8. J. Rothstein. Teacher quality in educational production: Tracking, decay, and student achievement. *The Quarterly Journal of Economics*, *125*(1), 175-214 (2010).

9. B. A. Jacob, L. Lefgren, D. P. Sims. The persistence of teacher-induced Learning. *Journal of Human Resources, 45*(4): 915-943, (2010).

10. D.F. McCaffrey, J.R. Lockwood., D. M. Koretz, T. A. Louis., L. Hamilton. Models for value-added modeling of teacher effects. *Journal of educational and behavioral statistics*, *29*(1), 67-101 (2004).

11. J.R. Lockwood, D.F. McCaffrey, L.T. Mariano, C. Setodji. Bayesian methods for scalable multivariate value-added assessment. *Journal of Educational and Behavioral Statistics*, *32*(2), 125-150 (2007).

12. L.T. Mariano, D.F. McCaffrey, J.R. Lockwood. A model for teacher effects from longitudinal data without assuming vertical scaling. *Journal of educational and behavioral statistics*, *35*(3), 253-279 (2010).

13. Across different value-added models and datasets, researchers have identified correlations ranging from 0.3 to 0.6 between teachers' proximal and future-year effects.

14. Value-Added Research Center. "NYC Teacher Data Initiative: Technical Report on the NYC Value-Added Model." (Madison Wisconsin: Wisconsin Center for Education Research, 2010). Retrieved on August 1, 2014 at http://schools.nyc.gov/NR/rdonlyres/A62750A4-B5F5-43C7-B9A3-F2B55CDF8949/87046/TDINYCTechnicalReportFinal072010.pdf.

15. A Technical Appendix detailing methods is available online as part of the Supplemental Materials section.

16. We observe little variation in the persistence of long-term knowledge across student characteristics. For example, the estimated persistence rate for long-term ELA knowledge is 0.95 for both low-income and higher-income students. Thus, differential student forgetfulness of long-term knowledge is unlikely to explain any differences in persistence that we observe.

17. In NYC, we find that around 19 percent of ELA teachers' effects persist two years after the year of instruction, virtually the same as the persistence rate we observe one year following instruction. Math teachers' effects in NYC decay substantially, however, with around 7 percent of proximal-year teachers' effects persisting two years after instruction. This is consistent with our hypothesis that math teachers' instructional effects are more content-specific.

18. Results available from authors.

19. Differential persistence rates associated with student and teacher characteristics are apparent in more "advantaged" schools across a range of specific cut points used to define our school samples.

20. S. P. Corcoran, J. L. Jennings, A. A. Beveridge. Teacher Effectiveness on High-and Low-Stakes Tests. *Society for Research on Educational Effectiveness,* (2011).

Additional References and Notes from the Supplemental Materials Section:

21. Prior to SY 2009-10, Math tests were administered in March of each year, and ELA tests in January of each year. From SY 2009-10 onward, both tests were administered in either April or May. As a specification check, we re-reran our analyses using TVA measures derived only from student data in the latter period. We found the same pattern of heterogeneity in persistence across observable student and teacher covariates as in our full sample.

22. W.A. Fuller. *Measurement error models*. (Wiley, New York, 1987).

23. The standard errors computed under this approach ignore error that comes from having used estimates of λ, β, C, and K to control for pre-tests, student-level variables, classroom-level variables, and school-level variables rather than the true values. However, given our instrumental variables approach to estimating persistence, we are not concerned with adjusting for error in our value-added measures.

24. Because our estimation of persistence includes controls for contemporaneous classroom fixed effects, our approach yields persistence point estimates that are virtually identical regardless of whether we include school fixed effects in our value-added modeling specification.

25. Jacob et al. (*9*) note, however, that our estimates of persistence may still be biased if schools adjust the instructional inputs (other than classroom assignments) that students receive, as a response to the quality of their lagged teacher. This could occur, for instance, if effective teacher raise students' achievement and this in turn leads schools to provide fewer instructional supports to the student.

26. We estimate two-year persistence for a sub-sample of our data that includes students in grades 6-8 who had been in the district for three consecutive years. In order to provide a

truly apples-to-apples comparison of one-year and two-year teacher persistence, we focus on a common sample of these students' teachers. So, for example, we observe the rate of persistence of a teacher's 6th grade value-added effect on her students' 7th grade achievement to estimate one-year persistence, and observe the persistence rate of the same 6th grade teacher's value-added effect on her students' 8th grade test scores to estimate two-year persistence.

TABLE 1

Summary statistics for students, teachers, and schools in our analytical sample

Variables	New York City
A. Students	
% Free price lunch	63.8
% Reduced price lunch	8.0
% Black	30.2
% Hispanic	36.3
% White	16.4
% Asian	16.5
% Female	51.8
N of Distinct Students	473004
B. Teachers	
% from a competitive undergraduate institution (Barron's rating)	30.0
Average LAST score	252.8 (21.4)
Average SAT verbal score	488.3 (93.2)
Average SAT math score	472.4 (93.3)
Average years of experience in the district	7.3 (6.3)
N of Distinct ELA Teachers	13660
N of Distinct Math Teachers	13368
C. Schools	
Average % of students eligible for free lunch	69.1 (24.0)
Average % of students Black	36.5 (30.1)
Average % of students Hispanic	39.3 (26.2)
Average % of teachers from a competitive institution	24.6 (15.0)
N of Distinct Schools	1169

Note: Analytical sample consists of students in grades 5 through 8 in school years 2005-06 through 2010-11, for whom prior-year teacher value-added data is available. The LAST is the Liberal Arts and Sciences Test for teacher licensure in New York State. Teacher SAT data is available for a subset (45%) of our teacher sample. Summary statistics for teachers reflect only teachers for whom value-added measures are available from more than one school year and for whom persistence rates can be estimated, as detailed in the Technical Appendix (*15*).

TABLE 2

Estimates for the Persistence of Observed Knowledge, Long-Term Knowledge, and Teacher Value-Added Effects, Within and Across Subjects

	Persistence Estimates		
	Observed Knowledge	Long-Term Knowledge	Teacher Value-Added Effects
Same-Subject Persistence			
Predicting ELA with ELA			
Coefficient on lagged ELA achievement	0.645	0.946	0.194
	(0.001)	(0.003)	(0.018)
First-stage F-statistic	-	75008	2109
N of students		933185	
Predicting math with math			
Coefficient on lagged math achievement	0.771	0.931	0.195
	(0.001)	(0.002)	(0.008)
First-stage F-statistic	-	169317	8725
N of students		981161	
Cross-Subject Persistence			
Predicting math with ELA			
Coefficient on lagged ELA achievement	0.581	0.639	0.160
	(0.001)	(0.003)	(0.017)
First-stage F-statistic	-	73089	2105
N of students		921511	
Predicting ELA with math			
Coefficient on lagged math achievement	0.609	0.614	0.044
	(0.001)	(0.002)	(0.010)
First-stage F-statistic	-	170557	8626
N of students		959327	

Note: Coefficient for Observed Knowledge from a regression of current achievement on prior achievement. Coefficient for Long-Term Knowledge from an instrumental variables (IV) regression of current achievement on prior achievement instrumented with twice-lagged achievement. Coefficient for Teacher Value-Added Effects from a regression of current achievement on prior achievement instrumented with teacher value-added quality. IV models include controls for current student characteristics and classroom fixed effects.

TABLE 3

Heterogeneity in the persistence of ELA teachers' value-added effects on student achievement in ELA and math in the subsequent school year

Subgroup interactions	Same-Subject: ELA on ELA			Cross-Subject: ELA on Math		
	In-group persistence	Out-group persistence	F-test of equal coefficients [p value]	In-group persistence	Out-group persistence	F-test of equal coefficients [p value]
A. Prior teacher characteristics						
Teacher attended a competitive institution	0.264	0.160	0.000***	0.214	0.136	0.000***
	(0.025)	(0.020)		(0.022)	(0.018)	
Teacher LAST score in top third	0.247	0.161	0.003**	0.220	0.137	0.001**
	(0.031)	(0.020)		(0.027)	(0.019)	
Teacher SAT math score in top third	0.199	0.157	0.219	0.211	0.126	0.006**
	(0.032)	(0.025)		(0.029)	(0.022)	
Teacher SAT verbal score in top third	0.242	0.144	0.009**	0.211	0.132	0.021*
	(0.036)	(0.024)		(0.032)	(0.022)	
B. Student characteristics						
Student eligible for free lunch in prior year	0.154	0.247	0.000***	0.122	0.211	0.000***
	(0.019)	(0.023)		(0.018)	(0.019)	
Student is black or Hispanic	0.154	0.270	0.000***	0.146	0.190	0.021*
	(0.018)	(0.026)		(0.018)	(0.021)	
Student’s twice-lagged test score below mean	0.111	0.208	0.000***	0.085	0.180	0.000***
	(0.021)	(0.021)		(0.023)	(0.018)	
N of students		933185			921511	

Note: Coefficients shown for interaction terms representing teacher value-added persistence for in-group and out-group samples across each teacher or student characteristic. Models include interaction terms (not shown) for value-added persistence across observations that are missing data for the particular characteristic of interest. Teachers' undergraduate institution competitiveness is based on Barron's rankings. The LAST is the Liberal Arts and Sciences Test for teacher licensure in New York State. ~p<.1 *p < .05, **p < .01, ***p < .001.

TABLE 4

ELA Teacher persistence estimates, with interactions of teacher and student characteristics across school-level characteristics

School Samples	Teacher or Student Subgroups [interactions]	In-group persistence	Std. Error	Out-group persistence	Std. Error	F-test of equal coefficients [p value]	N of students
A. Above mean ratio of teachers from competitive schools (1) vs below (2)							
(1) >24% competitive	Teacher attended competitive institution (versus others)	0.323	0.032	0.229	0.032	0.002**	441719
(2) <=24% competitive		0.096	0.072	0.093	0.030	0.960	463450
B. Above mean ratio of teachers with high LAST scores (1) vs below (2)							
(1) >23% with high LAST scores	Teacher LAST score in top third (versus others)	0.285	0.044	0.200	0.032	0.024*	430812
(2) <=23% with high LAST scores		0.153	0.063	0.155	0.029	0.977	479106
C. Above mean ratio of free-lunch eligible students (1) vs below (2)							
(1) <66% free-lunch eligible	Student eligible for free lunch (versus others)	0.278	0.044	0.352	0.041	0.016*	332216
(2) >=66% free-lunch eligible		0.168	0.034	0.200	0.040	0.323	318020
D. Above mean ratio of black and hispanic students (1) vs below (2)							
(1) <54% Black or Hispanic	Student is Black or Hispanic (versus others)	0.231	0.044	0.325	0.038	0.002**	320151
(2) >=54% Black or Hispanic		0.189	0.034	0.226	0.050	0.359	310013
E. Below mean ratio of low-scoring students (1) vs above mean (2)							
(1) <50% low scoring students	Student's prior ELA score below mean (versus others)	0.133	0.039	0.266	0.029	0.000***	457123
(2) >=50% low scoring students		0.075	0.027	0.090	0.033	0.634	459470

Note: School samples are determined by the characteristics of the school students attended in the prior year (i.e. in the year of instruction). All models exclude schools whose student or teacher population is overwhelmingly homogenous (i.e >95% or <5%) in terms of the selected student or teacher characteristic. Teachers' undergraduate institution competitiveness is based on Barron's rankings. The LAST is the Liberal Arts and Sciences Test for teacher licensure in New York State. Students' prior scores are twice-lagged ELA test scores. ~p<.1 *p < .05, **p < .01, ***p < .001.

TABLE 5

ELA Teacher persistence estimates across multiple school-level characteristics

School Samples	Schools' Student Characteristics	Schools' Staff Characteristics	Persistence estimate	Std. Error	N of students
A. High achieving and "better staffed"	>50% of students with prior scores above the mean	Above mean ratio of teachers (>24%) attended a competitive undergraduate school	0.309	0.038	255962
B. Low achieving but "better staffed"	<=50% of students with prior scores above the mean	Above mean ratio of teachers (>24%) attended a competitive undergraduate school	0.167	0.041	190519
C. High achieving but less well staffed	>50% of students with prior scores above the mean	Below mean ratio of teachers (<=24%) attended a competitive undergraduate school	0.171	0.046	201192
D. Low achieving and less well staffed	<=50% of students with prior scores above the mean	Below mean ratio of teachers (<=24%) attended a competitive undergraduate school	0.012	0.043	257396

Note: School samples are determined by the characteristics of the school students attended in the prior year (i.e. in the year of instruction). Students' prior scores are twice-lagged ELA test scores. Teachers' undergraduate institution competitiveness is based on Barron's rankings.

Supplemental Materials:

Technical Appendix

Estimating Teacher Value-Added

Following the model specification used by the Value-Added Research Center (*14*) in NYC, we compute value-added scores in three stages. In order to account for possible differences in tests, we estimate teacher value-added separately for each grade, subject, and year. In the first stage we estimate the coefficients λ for students' pretests and β for student-level characteristics on students' posttest scores. To estimate these coefficients, we regress posttest Y_t of student i in classroom c with teacher j in school s at time t on their same-subject pretest Y_{t-1}, other-subject pretest Y_{t-1}^{alt}, a vector of student-level covariates X, and a set of indicator variables representing individual classroom fixed effects π, which can be expressed as:

$$Y_{icjst} = \lambda Y_{it-1} + \lambda^{alt} Y_{it-1}^{alt} + \beta X_{it} + \pi_{cjst} + \varepsilon_{icjst} \qquad (1)$$

Our student-level characteristics include students' gender, race, an indicator for whether the student's home language is English, student eligibility for free or for reduced price lunch, student disability status, English language learner status, an indicator for whether the student switched schools in the prior year, and the number of prior-year absences for the student. Our student achievement measures come from annual state tests in Math and English language arts (ELA) (*21*). We standardize students' achievement test scores within each grade, subject, and year.

The first-stage regression is estimated using an errors-in-variables approach (following Fuller (*22*)) that accounts for measurement error in pretests Y_{it-1} and Y_{it-1}^{alt}. This removes the variance in the pretests that is attributable to measurement error. To facilitate this approach, we rely on reliability information reported in the technical manuals for the New York State assessments.

In the second stage, we use the estimated coefficients λ and β from our first stage to compute a new left-hand side variable q_{icjst}, where $q_{icjst} = Y_{icjst} - \lambda Y_{it-1} - \lambda^{alt} Y_{it-1}^{alt} - \beta X_{icjst}$. q_{icjst} is, then, the difference between the student's actual score and what we would predict it to be given student background characteristics and prior test performance. We then regress q_{icjst} on a vector C of classroom-level characteristics and time-varying school-level characteristics K:

$$q_{icjst} = \gamma C_{cjst} + \eta K_{st} + w_{icjst} \qquad (2)$$

Classroom-level characteristics include the racial and home language composition of the classroom, class size, the percent of students who are eligible for free or reduced price lunch, percent of students who are English language learners, the class average number of prior year absences, the class average prior year test scores in the same and alternate subject, and the standard deviation of classroom test scores in each subject. School characteristics include total enrollment, the percent of black, white, and Hispanic students in the school, and a control for the percent of students eligible for free or reduced price lunch. When running this regression, we specify a classroom random effect to take into account that errors are correlated within classrooms. From this regression, we obtain an estimate of w_{icjst}, that represents the residual post-test score variation for each student that is not explained by our observable student, classroom, or school characteristics.

In our third stage, we estimate individual teacher value-added measures in each year, τ_{jt}, by attributing all remaining variation in students' post-test scores to a combination of the individual teacher effects and error. This can be expressed as

$$w_{icjst} = \tau_{jt} + \varepsilon_{icjst} \quad (3)$$

We obtain estimates of the error term ε_{icjst} by subtracting each teachers' mean effect, τ_{jt}, from the estimates of w_{icjst} (*23*). We include in our analysis only teacher-by-year effects that are based on at least 5 students, and fewer than 100 students. We standardize our teacher-by-year effect estimates to have a mean of zero and a standard deviation of one.

Estimating the Persistence of Teacher Value-Added

We do not directly observe long-term knowledge, but rather the sum of long-term and short-term knowledge assessed in the prior period, Y_{t-1}. In light of this, as described by Jacob and colleagues (*9*) an ordinary least squares (OLS) coefficient θ_{OLS} for a regression of current achievement on prior achievement converges to the following:

$$\text{plim}\left(\hat{\theta}_{OLS}\right) = \theta\left(\frac{\sigma_{yl}^2}{\sigma_{yl}^2+\sigma_{ys}^2}\right) \quad (5)$$

This equation shows that because prior knowledge consists of a mix of long- and short-term knowledge, the OLS coefficient will be attenuated to the extent that Y_{t-1} consists of short-term, rather than long-term knowledge. In lieu of an OLS estimate of the persistence of observed knowledge, Jacob et al. (*9*) use an instrumental variables approach to estimate the decay of prior long-term knowledge, using twice lagged achievement Y_{t-2} as an instrument for Y_{t-1}. This estimator, which we refer to as $\hat{\theta}_{LT}$, purges Y_{t-1} of its short-term knowledge component. Jacob et al. (*9*) estimate (and we find) that almost all of a student's previously assessed long-term knowledge persists between one year and the next, with a value of $\hat{\theta}_{LT}$ close to 1. This serves as a benchmark for our subsequent estimation of teachers' effects on long-term knowledge.

Following a similar approach, we can estimate the proportion of a teacher's effect that consists of total long-term knowledge (both generic and context specific) by instrumenting each student's lagged knowledge Y_{t-1} with their lagged teacher's contribution (value-added) to that knowledge. The lagged teacher's total contribution to a student's lagged knowledge is a combination of her contribution to long- and short-term lagged knowledge, expressed as $\mathrm{M}_{t-1} = \mu^{l}_{t-1} + \mu^{s}_{t-1}$. Thus, the second stage estimator $\hat{\theta}_M$ converges to:

$$\operatorname{plim}\left(\hat{\theta}_M\right) = \theta\left(\frac{\sigma^2_{\mu l}}{\sigma^2_{\mu l}+\sigma^2_{\mu s}}\right) \qquad (6)$$

Given an estimate of θ that is close to 1, $\hat{\theta}_M$ approximates the fraction of teacher value-added that is attributable to long-term, rather than short-term, knowledge creation.

In practice, student assignment to teachers is nonrandom, and therefore the measured quality of a student's lagged teacher may be correlated with the quality of their current teacher. To minimize possible bias in our teacher persistence estimates due to nonrandom assignment, we include in our IV estimates of $\hat{\theta}_M$ additional controls for both student level covariates χ and for contemporaneous classroom fixed effects π (which subsume school, year and grade fixed effects). In addition, because teachers' value-added scores in any given year include estimation error that is correlated with other classroom-specific learning shocks in that year, we calculate, for each student in each period, their lagged teachers' average value-added score across all years *other* than the year in which they taught that student, expressed as $\mathrm{T}_{ijt-1} = \sum_{y \neq t-1} \mathrm{M}_{jy}$. The second-stage equation for estimating the persistence of teacher value-added then becomes:

$$Y_{icjt} = \theta_{\mathrm{M}} Y_{it-1} + X_{it} + \pi_{cjt} + \varepsilon_{ijt} \qquad (7)$$

Where the values of T_{ijt-1} for the lagged teachers serve as the excluded instruments for prior test scores, Y_{it-1}, in the first stage. In this formulation, persistence is a function of variation in the quality of the lagged teacher, distinct from the effects of the student's teacher or school in the current year.

In order to examine how much of teachers' effects consist of generalizable long-term knowledge that is relevant across subject areas, we modify equation 7 by replacing our outcome measure, Y_{icjt}, with a student's achievement in the alternate subject, Y_{icjt}^{alt}. Thus, for example, we model students' current math achievement as a function of their prior-year ELA achievement, instrumented by their lagged ELA teacher's value-added score. In addition, when predicting current math achievement, we include classroom fixed effects corresponding to their current-year math classroom assignment, rather than their ELA classroom. We do the reverse when estimating the persistence of lagged math teachers' value-added on students' current ELA achievement. This allows us to isolate the portion of the prior-year teachers' effects on long-term knowledge that reflect generalizable learning (i.e. learning effects that improve students' achievement in the alternate subject area), distinct from content-specific learning. We follow a similar procedure to estimate the generalizability of students' overall long run knowledge (i.e. including knowledge that is not teacher-driven) across subjects.

Estimating Heterogeneity in Teacher Value-Added Persistence

In order to test for heterogeneity in the persistence of teachers' value-added effects across our teacher and student characteristics of interest, we modify the first stage of our instrumental variables equation by replacing T_{ijt-1} with two interaction terms. The first term is set equal to T_{ijt-1} when the binary teacher or student characteristic is equal to 1, and 0 otherwise, while the other equals T_{ijt-1} when the characteristics is equal to 0, and is 0 otherwise. We similarly replace our lagged achievement measure Y_{it-1} from equation 7 with two interacted terms, following the same logic. For example, when investigating persistence across poor and non-poor students, we separately instrument poor students' lagged achievement scores with their lagged teachers' value-added, while also instrumenting non-poor students' achievement scores with their own lagged teachers' value-added. In cases where we are missing data on a student or teacher characteristics of interest for certain observations, we include an additional instrument and lagged achievement measure interacted with an indicator for the missing data. In practice, our approach yields very similar results to the alternative method of estimating persistence separately across in-group and out-group samples, which we run as a specification check. We opt to use interactions terms to facilitate a more succinct presentation of our findings. For each teacher or student characteristic of interest, we conduct F-tests to assess whether teacher value-added persistence coefficients for the in-group and out-group are significantly different from each other.

Estimating Persistence Using Alternative Value-Added Model Specifications

Because we are concerned with heterogeneity in the persistence of value-added effects as they are commonly measured in districts, we have intentionally used a value-added modeling approach that has been employed in NYC and in other school districts (*14*). Nonetheless, because different model specifications make different assumptions about what constitutes a teacher's effect on achievement, it is possible that our value-added model specification may influence our resulting estimates of teacher value-added persistence and heterogeneity in persistence. To account for this possibility, we test the robustness of our findings to two alternative model specifications. First, we consider an alternative model (specification "A") that includes student covariate controls, but no classroom or school characteristic controls. This approach attributes more of the variation in student achievement gains to teachers that might otherwise have been accounted for by the characteristics of students' classroom and school peers. Second, we consider a more extreme modeling approach (specification "B") that predicts student post-tests purely as a function of pretest scores with no additional control variables at the student, classroom, or school level. This model allows us to compare results for our baseline model that evaluates "similarly circumstanced" teachers against results for a model that instead attributes all of a student's achievement gains to individual teachers' effects, without adjustments for their school context or the characteristics of the students they teach.

We find that value-added models that include fewer controls for student and classroom characteristics yield higher persistence estimates than those for our baseline model, with average ELA persistence rates of 0.27 for alternative specification A (that includes student covariates) and 0.29 for specification B (that includes no classroom or student characteristics as control covariates). This indicates that including more contextual factors as part of our measures of teachers' value-added effects yields estimates that include more persistent contributors to students' future learning. Appendix Table 1-A provides an overview of observed heterogeneity in ELA teacher persistence and for each of our alternative value-added specifications. Overall, we continue to see significant differences in value-added persistence as a function of observable student and teacher characteristics. However, the magnitude of these differences is somewhat smaller across teacher characteristics in both alternative specifications.

Estimating 2-Year Persistence and Heterogeneity in Persistence

Research on teachers' value-added persistence has typically shown that the portion of teachers' effects that influences long-term knowledge continues to raise student achievement for several years (*7, 9, 12*). As noted previously, we observe overall two-year persistence rates for ELA teacher value-added at around 19 percent of teachers' proximal year effects, virtually the same as the 20 percent one-year persistence rate we observe in an identical sample of students and teachers (*26*). In addition, in order to confirm that the heterogeneity that we observe in ELA teachers' one-year value-added persistence continues into the second year following instruction, we compare heterogeneity in one-year and two-year value-added persistence in Appendix Table A-2.

Overall, we see that the heterogeneity apparent in teachers' one-year persistence estimates carries over into two-year persistence as well. In this restricted sample of teacher and student data, teachers' academic characteristics are less predictive of differences in persistence rates, relative to results from our full sample. However, results corresponding to student characteristics are very similar between our full and restricted samples. Moreover, variation in one-year persistence estimates corresponding to each student or teacher characteristic is also apparent in two-year persistence estimates. For example, teachers of students who are eligible for free lunch have both low one-year value-added persistence (0.17) and low two-year persistence (0.16), relative to the rates for students who are not eligible for free lunch (0.27 and 0.23, respectively). Variation in one-year value-added persistence appears to be a reasonable proxy for teachers' varying influence on student performance over an extended period.

Table A-1

Heterogeneity in ELA teacher persistence estimates when using alternative value-added models

Subgroup interactions	A. Model including only pre-tests and student-level covariates			B. Model including only pre-tests, with no student, classroom, or school covariates		
	In-group persistence	Out-group persistence	F-test of equal coefficients [p value]	In-group persistence	Out-group persistence	F-test of equal coefficients [p value]
A. Teacher characteristics						
Teacher from a competitive undergraduate school	0.302	0.256	0.001**	0.325	0.282	0.000***
	(0.020)	(0.017)		(0.019)	(0.017)	
Teacher LAST score in top third	0.293	0.249	0.005**	0.317	0.278	0.004**
	(0.022)	(0.017)		(0.020)	(0.016)	
Teacher SAT math score in top third	0.251	0.247	0.834	0.281	0.274	0.679
	(0.022)	(0.020)		(0.020)	(0.019)	
Teacher SAT verbal score in top third	0.262	0.238	0.193	0.290	0.268	0.179
	(0.023)	(0.019)		(0.021)	(0.018)	
B. Student characteristics						
Student eligible for free lunch in prior year	0.237	0.307	0.000***	0.262	0.330	0.000***
	(0.017)	(0.019)		(0.017)	(0.018)	
Student is black or hispanic	0.236	0.336	0.000***	0.263	0.354	0.000***
	(0.017)	(0.022)		(0.016)	(0.021)	
Student's test score prior to instruction was below mean	0.129	0.277	0.000***	0.127	0.314	0.000***
	(0.020)	(0.019)		(0.020)	(0.018)	
N of students		933429			933429	

Note: Model A predicts includes controls for prior-year test scores and student-level covariates only. Model B includes only prior-year test score controls, and attributes all remaining variance in post-test scores to individual teachers. Teachers' undergraduate institution competitiveness is based on Barron's rankings. ~p<.1 *p < .05, **p < .01, ***p < .001.

TABLE A-2

Heterogeneity in the persistence of ELA teachers' value-added effects, one and two years after instruction, across a common sample

Subgroup interactions	One-Year Persistence			Two-Year Persistence		
	In-group persistence	Out-group persistence	F-test of equal coefficients [p value]	In-group persistence	Out-group persistence	F-test of equal coefficients [p value]
A. Prior teacher characteristics						
Teacher attended a competitive institution	0.260	0.193	0.030*	0.227	0.180	0.045*
	(0.032)	(0.024)		(0.024)	(0.018)	
Teacher LAST score in top third	0.160	0.208	0.403	0.184	0.187	0.938
	(0.047)	(0.031)		(0.028)	(0.019)	
Teacher SAT math score in top third	0.216	0.185	0.504	0.234	0.183	0.125
	(0.044)	(0.031)		(0.031)	(0.022)	
Teacher SAT verbal score in top third	0.234	0.182	0.326	0.247	0.183	0.078~
	(0.053)	(0.029)		(0.036)	(0.022)	
B. Student characteristics						
Student eligible for free lunch two years prior	0.165	0.265	0.000***	0.163	0.225	0.003**
	(0.024)	(0.028)		(0.018)	(0.022)	
Student is black or Hispanic	0.165	0.299	0.000***	0.156	0.257	0.000***
	(0.023)	(0.035)		(0.017)	(0.027)	
Student's thrice-lagged test score below mean	0.097	0.255	0.000***	0.103	0.202	0.000***
	(0.026)	(0.026)		(0.020)	(0.021)	
N of students		512280			512280	

Note: Results from a sub-sample of students in grades 6-8 who had been in the district for three consecutive years. Coefficients shown for interaction terms representing teacher value-added persistence for in-group and out-group samples across each teacher or student characteristic. Models include interaction terms (not shown) for value-added persistence across observations that are missing data for each particular characteristic of interest. Teachers' undergraduate institution competitiveness based on Barron's rankings. The LAST is the Liberal Arts and Sciences Test for teacher licensure in New York State. ~p<.1 *p < .05, **p < .01, ***p < .001.