NATIONAL
CENTER *for* ANALYSIS *of* LONGITUDINAL DATA *in* EDUCATION RESEARCH

TRACKING EVERY STUDENT'S LEARNING EVERY YEAR

A program of research by the American Institutes for Research with Duke University, Northwestern University, Stanford University, University of Missouri-Columbia, University of Texas at Dallas, and University of Washington

Teacher Quality Gaps and Student Outcomes: Assessing the Association Between Teacher Assignments and Student Math Test Scores and High School Course Taking

Dan Goldhaber
Roddy Theobald
Danielle Fumia

WORKING PAPER 185 • January 2018

Teacher Quality Gaps and Student Outcomes: Assessing the Association Between Teacher Assignments and Student Math Test Scores and High School Course Taking

Dan Goldhaber
American Institutes for Research/CALDER

Roddy Theobald
American Institutes for Research/CALDER

Danielle Fumia
Washington State Institute for Public Policy

Contents

Acknowledgements

This research was supported by the National Center for Analysis of Longitudinal Data in Education Research (CALDER), which is funded by a consortium of foundations. For more information about CALDER funders, see www.caldercenter.org/about-calder. This work is also supported by the William T. Grant Foundation (grant #184925). We wish to thank the State of Washington Office of Superintendent of Public Instruction (OSPI) for providing the data we utilize, Stephanie Liddle for excellent research assistance, and James Cowan and seminar participants at Boston University for comments that improved the manuscript. All opinions expressed in this paper are those of the authors and do not necessarily reflect the views of our funders or the institutions to which the author(s) are affiliated

CALDER working papers have not undergone final formal review and should be cited as working papers. They are intended to encourage discussion and suggestions for revision before final publication. Any opinions, findings, and conclusions expressed in these papers are those of the authors and do not necessarily reflect the views of our funders.

CALDER • American Institutes for Research
1000 Thomas Jefferson Street N.W., Washington, D.C. 20007
202-403-5796 • www.caldercenter.org

Teacher Quality Gaps and Student Outcomes: Assessing the Association Between Teacher Assignments and Student Math Test Scores and High School Course Taking

Dan Goldhaber, Roddy Theobald, Danielle Fumia

CALDER Working Paper No. 185
January 2018

Abstract

We use panel data in Washington State to study the extent to which teacher assignments between fourth and eighth grade explain gaps between advantaged and disadvantaged students—as defined by underrepresented minority status (URM) and eligibility for free or reduced price lunch (FRL)—in their eighth grade math test scores and high school course taking. We find some significant gaps between advantaged and disadvantaged students in the value added of the teachers to which they are assigned in these grades, although gaps in middle school grades are sensitive to the specification of value added. We then show that teacher assignments are highly predictive of both eighth-grade test scores and advanced course taking in high school, and that differences between advantaged and disadvantaged students in teacher assignments explain significant portions of student outcome gaps. In the case of eighth-grade test scores, the URM/non-URM gap drops by more than 15% and the FRL/non-FRL gap drops by more than 20% when we control directly for teacher assignments. That said, the reduction in the achievement gap is more modest when we control for measures of teacher value added that do not control for classroom characteristics (8% and 9%, respectively), while gaps actually increase slightly when we control for measures of value added that control for classroom characteristics. These patterns are qualitatively similar and even larger in magnitude when we consider the number of advanced math courses taken in high school as the outcome.

1. Introduction

There is a significant and growing body of evidence showing that disadvantaged students, typically measured by race/ethnicity or economic status, tend to be assigned to less credentialed and effective teachers.[1] Evidence of these teacher quality gaps (TQGs) has garnered high-level policy attention. Inequities in the distribution of measures of teacher quality across student subgroups has factored into the ruling in several educational adequacy lawsuits (e.g., Leandro, 1997). Additionally, as part of the U.S. Department of Education's Excellent Educators for All Initiative, all states were required in 2014 to create new Comprehensive Educator Equity Plans designed to reduce inequity in the distribution of teacher quality across public schools (Rich, 2014).

The argument for focusing on teacher equity is simple: a large body of empirical evidence shows that teachers have significant effects on students' test performance, educational attainment, and noncognitive outcomes, as well as long-term impacts on later life outcomes such as employment probabilities and labor market earnings.[2] The large impacts of teachers, combined with evidence of a considerable amount of heterogeneity in teacher effectiveness (Koedel et al., 2016; Nye et al., 2004; Rivkin et al., 2005), arguably make teachers the key schooling variable influencing the equality of educational opportunity.

In this paper, we document the extent to which teacher assignments in Grades 4–8 are associated with eighth grade math test scores and high school course taking. To our knowledge, this is

---

[1] Teacher quality gaps between advantaged and disadvantaged students show up whether teacher quality is measured by degrees, experience, or advanced credentials (e.g. Clotfelter et al., 2005; Goldhaber et al., 2015, 2018; Kalogrides and Loeb, 2013; Lankford et al., 2002) and/or by value-added measures of teacher effectiveness (e.g., Goldhaber et al., 2015, 2018; Isenberg et al., 2016; Sass et al., 2012).

[2] See, for instance: Aaronson et al. (2007), Bacher-Hicks et al. (2014), Goldhaber and Hansen (2013), Jacob et al. (2010), Kane et al. (2013), and McCaffrey et al. (2009) on the effects of teachers on student test performance; Gershenson (2016), Kraft (forthcoming), and Jackson (forthcoming) on teacher effects on noncognitive outcomes (e.g., student absences); and Chamberlain (2013) and Chetty, Friedman, and Rockoff (2014b) on their effects on longer-term outcomes (e.g., college-going behavior, labor market outcomes, etc.).

the first empirical evidence relating teacher value added in these grades to student course taking in high school. We also document teacher quality gaps between advantaged and disadvantaged students in these grades and related gaps in student math test achievement and advanced course taking. Finally, we explore the extent to which teacher assignments appear to explain these test and course-taking outcome gaps between advantaged and disadvantaged students.

Consistent with prior evidence (e.g., Betts et al., 2003; Clotfelter et al., 2009; Reardon, 2011), we find large gaps between traditionally advantaged and disadvantaged students in student math achievement in both the third and eighth grades, as well as significant differences in the number of advanced math courses students take while in high school that also align with historical figures (e.g., Gamoran, 1987; Lee, 2002). For instance, third grade math test scores for underrepresented minority (URM) students—defined as American Indian, Black, or Hispanic—are about 0.6 standard deviations lower than those of non-URM students, and this gap remains in eighth grade math test scores. These gaps are consistent for students eligible for free or reduced-price lunch (FRL) in third and eighth grade. Similarly, URM and FRL students are 15 and 20 percentage points less likely to take any advanced math courses in high school compared to non-URM and non-FRL students, respectively.

We explore the extent to which these persistent gaps might be explained by teacher quality gaps (TQGs) across grades using models that directly account for the assignment of students to particular teachers in Grades 4–8 and models that use estimates of value added as a measure of the quality of teachers to which students are assigned in these grades. As it turns out, the value-added measures of TQGs are somewhat sensitive to the specification of the value-added model. There is little difference in the estimates of the TQGs at the elementary level whether or not value-added models are specified with classroom-level covariates; for example, we find teacher quality gaps in fourth and fifth grade whereby FRL students have teachers about 0.02–0.03 and 0.013–0.017 standard deviations below non-FRL students, respectively, regardless of value-added model specification. But the estimates of

TQGs in Grades 6–8 are sensitive to the specification of the value-added model. For example, in eighth grade, teacher quality gaps estimated using specifications controlling for classroom characteristics suggest that FRL students have higher-quality teachers than non-FRL students, whereas models that do not include classroom controls suggest that FRL students have lower-quality teachers, again by about 0.03 standard deviations.

Finally, we demonstrate the importance of teacher assignment in models predicting end of eighth grade test scores and high school course taking. Models that directly control for teacher assignments suggest that the teachers to whom a student is assigned in Grades 4–8 explain about 16% and 21% of the eighth-grade math achievement gaps between advantaged and disadvantaged students and about 33% of the gaps in the number of advanced math courses taken in high school. Value added of the teachers to whom students are assigned explains a modest portion of these gaps—between 8% and 9% of gaps in eighth grade math achievement and number of advanced math courses taken—although value added does account for about half of the total effect of teacher assignments. We further show the importance of the value-added model specification because value-added measures that include classroom controls suggest that assignment to higher-quality teachers increases the gap between advantaged and disadvantaged students in math achievement and has almost no effect on gaps in advanced math course taking, in contrast to models that directly control for teacher assignment.

The remainder of the paper is organized as follows. In Section 2, we provide background on prior empirical evidence about student achievement gaps, teacher quality gaps, and the impact of teachers on subsequent student outcomes. We describe our data and analytic approach in Section 3, present results in Section 4, and offer concluding thoughts in Section 5.

2. Background on the Importance of Teacher Quality and Distribution

This study connects two different strands of literature. The first strand deals with the inequities between advantaged and disadvantaged students in achievement and, often, in access to educational resources. The second is related to the import of teacher quality in explaining future student test and other academic (and nonacademic) outcomes. As we describe below, there is relatively little work linking these two strands.

Considerable prior evidence documents persistent gaps in test achievement (e.g., Betts et al., 2003; Clotfelter et al., 2009; Reardon, 2011) and high school advanced course taking (e.g., Gamoran, 1987; Kelly, 2009; Lee, 2002) between advantaged and disadvantaged students. For instance, Clotfelter et al. (2009) document substantial achievement gaps in North Carolina by student race that largely persist from Grades 3 through 8 (e.g., gaps of about 0.8 standard deviations between Black and White students and about 0.5 standard deviations between Hispanic and White students in math). Likewise, Kelly (2009) finds that White students in the National Education Longitudinal Study are about 60% more likely to take an advanced class in high school than Black students. Given these gaps in K–12 outcomes, it is not surprising that disadvantaged students are far less likely to graduate high school, attend college, and graduate from college than more advantaged students (e.g., Aud et al., 2011; Kena et al., 2015).

Evidence suggests that advanced high school *math* course taking, specifically, can have effects on students' college readiness and attainment in general, and on success in college math courses, in particular. For example, Long et al. (2012) find that students taking advanced math courses were more likely to enroll in 4-year colleges rather than 2-year colleges even among those taking advanced courses in one or more other subjects. Rose and Betts (2004) find a similar positive effect of algebra and geometry courses on earnings even after controlling for course taking in other subjects. Math course taking in high school also predicts readiness for college-level math courses (Long, et al., 2009) and increases the likelihood of choosing a STEM major in college (Federman, 2007). Further evidence illustrates that substantial portions of URM/non-URM and FRL/non-FRL gaps in readiness for college-level math courses (Long et al., 2009), earnings (Rose and Betts, 2004), and URM/non-URM gaps in rates of STEM degree completion (Tyson, et al., 2007) can be explained by math course taking in high school.

There is some evidence that differences in teacher quality between advantaged and disadvantaged students may explain some portion of the above achievement gaps.[3] A significant amount of evidence shows teacher qualifications (e.g., experience and degrees) are inequitably distributed across students (Betts et al., 2003; Clotfelter et al., 2005; Lankford et al., 2002). More recent evidence has buttressed these findings (Kalogrides and Loeb, 2013) and also shows that there tend to be inequities when teacher quality is measured based on value added as well (Goldhaber et al., 2015, 2018; Isenberg et al., 2016; Sass et al., 2012). In particular, prior work in Washington State (Goldhaber et al., 2015, 2018), the setting of this study, illustrates the magnitude and consistency of these gaps. Specifically, Goldhaber et al. (2015) find that URM and FRL students tend to be assigned to teachers who are .02 to .05 standard deviations less effective in value added than their more advantaged peers in elementary, middle school, and high school, while Goldhaber et al. (2018) report that these teacher quality gaps have been consistent over the past decade.

There are good reasons to believe that TQGs have an impact on student achievement gaps. Evidence shows that teachers have significant effects on students' test performance (e.g., Aaronson et al., 2007; Bacher-Hicks et al., 2014; Goldhaber and Hansen, 2013; Kane et al., 2013; Kane & Staiger, 2008; McCaffrey et al., 2009), noncognitive outcomes (e.g., Gershenson, 2016; Kraft, forthcoming; Jackson, forthcoming), and longer-term educational attainment (e.g., Chamberlain, 2013; Chetty et al., 2014b). Importantly, while teachers' impacts on test achievement and noncognitive outcomes are only weakly correlated with their impacts on noncognitive outcomes (e.g., Kraft, forthcoming), teacher value

[3] There is a broader literature on whether schools or resources more generally help explain gaps (e.g., Hubbard, 2017; Jackson et al., 2016; LaFortune et al., 2016). Here we focus more narrowly on teacher quality, which is the most important schooling factor predicting student achievement (e.g., Rivkin et al., 2005).

added to student test achievement is quite predictive of students' long-term outcomes (e.g., Chetty et al. [2014b] find that higher value-added teachers influence later student outcomes like teen pregnancy, college attendance, and earnings).

On the other hand, studies also tend to find that the test score gains induced by teachers in one grade dissipate, or "fade out," in later grades. Specifically, estimates of the persistence of value added across grades suggest that 50–60% of teacher's value added is no longer detectable in terms of student test achievement 2 years after students have had a teacher, and upward of 80% has faded out after 3 years (Chetty et al., 2014a; Jacob et al., 2011; Kane and Staiger, 2008; Kinsler, 2012; Konstantopoulos & Chung, 2011; Lockwood et al., 2007; McCaffrey et al., 2004; Rothstein, 2010).[4]

To our knowledge, this is the first paper to link the above literature strands by exploring the extent to which teacher quality gaps appear to explain subsequent achievement and outcome gaps between advantaged and disadvantaged students. In the next section, we describe the data and analytic approach that allow us to connect these different strands of prior research.

3. Data, Measures, and Analytic Approach

3.1. Data

For our analysis, we use 11 years of administrative student-level data from Washington, provided by the Washington State Office of Superintendent of Public Instruction (OSPI). In 2005–06, the state began annual testing in both math and reading in Grades 3–8, which means that we observe both current and prior test performance for students in Grades 4–8 from 2006–07 through 2015–16. There

[4] One possible explanation for fade out is that teachers facing test-based accountability pressure focus more narrowly on students' test taking skills, crowding out some deeper learning that may be important for students' success in higher grades, college, or the workforce (Corcoran et al., 2011). There are, however, less pernicious explanations for fade out, such as variation in test content across grades and test scaling effects (Cascio and Staiger, 2012).

have been two test regime changes over this time period: the state transitioned from the Washington Assessment of Student Learning (WASL) to the Measures of Student Progress (MSP) in 2009–10, and then to the Smarter Balanced Assessment (SBA) in 2014–15[5]; see Backes et al. (2018) for evidence that estimates of teacher value added are largely unaffected by these test regime changes. We standardize all test scores across all test takers within grade and year to ensure that scores are comparable across years.

Between 2006–07 and 2008–09, we link students in Grades 4–6 in elementary schools to their classroom teachers through a proctor field in the state assessment file.[6] Since the 2009–10 school year, students can be linked to their teachers using a unique classroom ID in the state's CEDARS database.[7] For all school years, the student database contains additional information on individual student background variables including gender, race/ethnicity, learning disability status, and free or reduced-price lunch eligibility, as well as participation in the following programs: gifted/highly capable; limited English proficiency (LEP); and special education. These student-level variables are used as control variables in all our models, and two variables—whether the student is American Indian, Black, or Hispanic (i.e., underrepresented minority, URM) or is eligible for free or reduced-price lunch (FRL)—are our primary measures of student disadvantage. All years of data are also merged to the state's S-275 database, which contains information from OSPI's personnel-reporting process and includes school assignments of all certified employees in the state and the experience level of these employees.

3.2. Measures

[5] About one-third of schools in the state participated in a pilot of the SBA in 2013–14, and the state did not collect test scores from students in these schools for this school year. Thus, current test scores are missing for students in these schools in 2013–14, and prior test scores are missing for students in these schools in 2014–15.
[6] The proctor of the state assessment was used as the teacher-student link for at least some of the data used for analysis. The "proctor" variable was not intended to be a link between students and their classroom teachers so this link may not accurately identify those classroom teachers.
[7] CEDARS data include fields designed to link students to their individual teachers, based on reported schedules. However, limitations of reporting standards and practices across the state may result in ambiguities or inaccuracies around these links.

Value-Added Estimates

Our measure of teacher quality is based on value-added models that seek to isolate the contribution of individual teachers to student test score gains. The value-added model specifications we utilize rely on the value-added framework estimation that is described in Chetty et al. (2014a) because the value-added estimates from this specification have been validated as an out-of-sample predictor of both short-term and long-term student outcomes (Chetty et al., 2014a, 2014b).[8] Specifically, we use the following procedure in each grade from 4 through 8.[9] First, we create a residualized test score for each student *i* with teacher *j* in year *t* by estimating the following regression:

$$Y_{ijt} = \alpha_j + \delta Y_{i(t-1)} + \gamma X_{it} + \varepsilon_{ijt} \qquad (1)$$

In the model in equation 1, the outcome variable Y_{ijt} is the student's standardized test score in year *j*; our primary analysis focuses on student math performance. The predictor variables include: $Y_{i(t-1)}$, a vector of prior test scores in math and reading; X_{it}, a vector of student and/or classroom characteristics in year *t*; and a teacher fixed effect α_j. We use the estimated coefficients $\hat{\delta}$ and $\hat{\gamma}$—which are estimated from within-teacher variation due to the presence of the teacher fixed effect in equation 1—to create the residualized test scores:

$$Y_{ijt}^* = Y_{ijt} - \hat{\delta} Y_{i(t-1)} - \hat{\gamma} X_{it} \qquad (2)$$

Y_{ijt}^* can be interpreted as a student's residual test score adjusting for the student's prior performance and observable characteristics.

We then use the mean residual scores for teacher *j* in year *t*, $\overline{Y_{jt}^*}$, to calculate the teacher value-added estimates. We first calculate forecasting coefficients, ψ_s, where *s* is the number of years between the observed school year and the forecasting target:

[8] We replicate this specification using the vam STATA package (Stepner, 2013).
[9] We estimate all models separately by grade because the models described in Section 3.4 consider value-added in different grades as separate predictors.

$$\psi = \arg\min_{\{\psi_s\}} \sum_j \left(\overline{Y_{Jt}^*} - \sum_{s \neq t} \psi_s \overline{Y_{Js}^*}\right)^2 \qquad (3)$$

In other words, we estimate the forecasting coefficients to minimize the mean-squared error of the forecasts (see Chetty et al. [2014a] for additional details).

Finally, we use the estimates $\hat{\psi}_s$ from equation 3 and the mean residual scores $\overline{Y_{Jt}^*}$ to calculate teacher value added in year *t:*

$$\hat{\tau}_{jt} = \sum_{s \neq t} \hat{\psi}_s \overline{Y_{Jt}^*} \qquad (4)$$

The estimates $\hat{\tau}_{jt}$ produced by this procedure are "leave-one-out" or "jackknife" estimates of teacher value added in that they use data on students linked to a teacher in all years other than year *t* to estimate value added in year *t*. While this is essential for the predictive models described in Section 3.4, teachers who are linked to students in only one year do not have a value-added estimate, which has implications for the analytic sample described in Section 4.3.[10]

Within the framework described above, we calculate four different specifications of value added. First, we estimate models that do and do not include aggregated characteristics of a student's classmates in the vector X_{it} in equations 1 and 2. The objective in estimating value added is to get a *causal* estimate of the contribution that teachers make toward student achievement that is separate from the individual and joint influence of student background characteristics, but it is difficult to separate the influence of a student's peers from inequities in teacher quality across different types of classrooms (Zamarro et al. 2015). This is reflected in the fact that prior work has shown that estimated teacher quality gaps can be sensitive to the inclusion of these covariates (Goldhaber et al., 2016; Isenberg et al., 2016).

Unfortunately, it is difficult to know outside of an experiment whether these differences are because models without classroom controls misattribute peer effects to differences in value added

[10] As discussed in Chetty et al. (2014a), the estimates $\hat{\tau}_{jt}$ are implicitly shrunken by the forecasting coefficient estimates $\hat{\psi}_s$, and thus no additional corrections for measurement error are necessary.

across different types of classrooms, or because models with classroom controls over-control for the influence of these peer effects and remove true differences in teacher quality across different types of classrooms (Goldhaber et al., 2016; Isenberg et al., 2016). As discussed above, the peer effects are identified by within-teacher variation in classroom characteristics, but this model could over-control for peer effects if, for example, teachers who teach both advantaged and disadvantaged classes in the same year have different expectations for different types of classrooms or put more effort into their advantaged classes.[11] Given that we find substantial differences in estimated teacher quality gaps in middle school grades depending on whether the model controls for classroom covariates (see Section 4.3), and prior work has validated estimates from both types of specifications (Chetty et al., 2014a; Kane et al., 2013), we present all results in this paper separately for specifications that do and do not control for these classroom covariates.

Second, we estimate models that do and do not account for a teacher's experience in year *t.* This is important given evidence of substantial returns to early-career teaching experience (e.g. Kraft and Papay, 2014; Ladd and Sorenson, 2017; Rivkin et al., 2005; Rockoff, 2004) and the fact that prior work in Washington suggests that disproportionate assignment to novice teachers explains about one-third of the teacher quality gap in elementary grades (Goldhaber et al., 2018). Specifically, we create a vector of teacher indicators Exp_{jt} for whether a teacher has 1, 2, …, 8 years of experience or 9 or more years of experience in year *t* (0 years of experience is the reference category). In these "experience-adjusted" VAMs, this vector is first included in the first-stage regression:[12]

---

[11] Zamarro et al. (2015) use simulated data to show that that models with and without classroom covariates both understate true TQGs when there is limited variability in classroom composition and weak peer effects, and that classroom covariate models understate the true TQGs more. It is possible that these classroom covariates are picking up tracking (both informal and formal) within schools. We estimate models that control for formal tracking (e.g., advanced and remedial courses) and find very similar results. See Gershenson et al. (2016) for recent evidence about variation in teacher expectations for different students.

[12] In Appendix Table A1, we report the coefficients on each of the teacher experience dummies from this regression. The estimated returns to the first 9 years of experience range from 0.049 standard deviations in eighth grade to 0.143 standard deviations in sixth grade, and are comparable to other estimates in the literature.

$$Y_{ijt} = \alpha_j + \delta Y_{i(t-1)} + \gamma X_{it} + \theta Exp_{jt} + \varepsilon_{ijt} \quad (5)$$

The estimates from model 5 are then used to calculate the residualized test scores:

$$Y_{ijt}^{*} = Y_{ijt} - \hat{\delta} Y_{i(t-1)} - \hat{\gamma} X_{it} - \hat{\theta} Exp_{jt} \quad (6)$$

We can then use equations 3 and 4 to calculate the value-added estimates $\hat{\tau}_{jt}$. However, this estimate accounts for teacher experience in all years *other than* year *t*, when what matters is the teacher's experience in year *t* (i.e., the year in which the student had the teacher). As a last step, therefore, we add back in the expected returns to experience to the estimates $\hat{\tau}_{jt}$:

$$\hat{\tau}_{jt}^{Exp} = \hat{\tau}_{jt} + \hat{\theta} Exp_{jt} \quad (7)$$

We refer to the estimates $\hat{\tau}_{jt}^{Exp}$ as "experience-adjusted value-added estimates," and produce these estimates both with and without classroom controls.

Advanced Mathematics Courses

In an effort to consider longer-term outcomes other than test performance, we follow prior work in Washington (Goldhaber et al., 2017) and use the CEDARS student schedule files to create a measure of advanced math course taking in high school. We define high school courses as "advanced" following the procedure described in Gottfried (2015), which relies on a taxonomy outlined in Burkham et al. (2003).[13] In our primary results, advanced math courses include trigonometry, statistics, precalculus, and higher courses. Our longitudinal data allow us to track the first two cohorts of third graders (i.e., third graders in 2005–06 or 2006–07) through all four years of high school, so we create two measures for students in these cohorts who are observed in all four years of high school: an indicator for whether the student took an advanced math course in high school; and the number of advanced math courses the student took in high school.

3.3. Analytic Samples

[13] At the high school level, courses are classified via state course codes and state course names. In cases where a course is not mentioned in Burkham et al. (2003), we use out best judgment to determine which level a course aligns with, and delete observations in schools with all missing state course names.

We create two different analytic samples. The first, which we refer to as the "eighth-grade achievement sample," includes six cohorts of students that were enrolled in the third grade in 2005–06 through 2010–11, and for whom we observe their eighth-grade math test score, a baseline third-grade math score, information about URM and third-grade FRL status, and who are matched to at least one of their teachers in Grades 4–8. Students are matched to teachers who have a full-time teaching appointment in a single school within a given year and, in the case of value added, teachers for whom we can estimate value added based on 10 or more students in a classroom. We exclude from the value-added estimates student-year observations in which students are matched to multiple teachers in a year.

The second analytic sample, which we refer to as the "high school course-taking sample," includes two cohorts of students who were enrolled in the third grade in 2005–06 or 2006–07, and for whom we observe all four years of high school and at least one teacher in Grades 4–8.[14] We require students to be observed for all four years of high school in order to be included in the high school sample so that dropout is not confounded with course taking, but this decision likely means that our estimated relationships between teacher quality and course taking represent a lower bound if more effective teachers both have a positive impact on course taking and a negative impact on the probability of dropout.

After these restrictions, the eighth-grade achievement sample includes 330,539 student observations and 36,729 teacher-year observations (11,194 unique teachers), and the high school course-taking analytic sample includes 104,001 student observations and 12,829 teacher-year observations (7,874 unique teachers). Table 1 shows the implications of the above restrictions. Specifically, we report descriptive statistics for the unrestricted (column 1) and the two analytic samples

[14] Note that while the overwhelming proportion of students in the high school course-taking sample graduate from high school, about 93%, graduation from high school is not a requirement to be in this sample.

(columns 2–6 for the eighth grade achievement sample and 7–11 for the high school course-taking sample) across all the cohorts in each subsample.

Students in the restricted sample appear somewhat less disadvantaged than those in the unrestricted samples. For example, in comparing the unrestricted sample means to the eighth grade test sample (column 1 compared to column 2), we see that the students in the full analytic sample (column 2) are slightly less likely to be URM or FRL students, have substantially higher third grade math and reading test scores (by about 2–3% of a standard deviation), and are less likely to be English language learners or receive special education services.[15] The differentials are generally even more stark between the unrestricted and high school course-taking samples (column 1 compared to column 7), which is not surprising given that disadvantaged and lower achieving students are less likely to make it all the way through high school (e.g., Heckman & LaFontaine, 2010). Given the significant differences between the unrestricted and analytic samples, it is important then to note that the findings in this paper may not generalize to the entire population of students in these cohorts.

Table 1 also shows large differences in the baseline characteristics and achievement of the different subgroups of students in our sample. Both URM and FRL students are more likely than other students to be designated as having a learning disability or to be an English language learner. And they have far lower baseline achievement levels. Indeed, consistent with other evidence (e.g., Betts et al., 2003; Clotfelter et al., 2009), we see large gaps in the third grade between URM and non-URM and FRL and non-FRL students; these are in the neighborhood of 0.60 to 0.65 standard deviations on the third-grade tests. Moreover, the magnitudes of the eighth-grade math test score gaps are about as large they were in the third grade, indicating that disadvantaged students who stay in Washington do not substantially catch up with more advantaged third grade students.[16] Not surprisingly, given the

[15] Note that student test scores are standard normalized within the unrestricted sample.
[16] The finding that disadvantaged students do not catch up with their more advantaged peers while enrolled in school is consistent with Betts et al. (2003) and Clotfelter et al. (2009). Note, however, it is unclear whether this means that

significant difference in eighth-grade achievement, there are also large gaps in the number of advanced courses that students take while in high school: non-URM students on average take about 75% more advanced courses than URM students, and non-FRL students on average take about twice as many advanced courses as FRL students.[17]

There is also prima facie evidence that students are assigned to very different teachers. For example, disadvantaged students are more likely than advantaged students to have a novice teacher—one who has two or fewer years of teaching experience—in each grade from fourth through eighth grade. While not reported in Table 1, they are also more likely to repeatedly, across grades, have less-experienced teachers. For instance, URM students are 50% more likely than non-URM students to have two or more novice teachers in fourth through eighth grade; and FRL students are 25% more likely than non-FRL students to have two or more novice teachers. These findings are notable since early career teaching experience is strongly predictive of teacher effectiveness (e.g. Kraft and Papay, 2014; Ladd and Sorenson, 2017; Rivkin et al., 2005; Rockoff, 2004).[18] This also emphasizes the need (discussed above) to account for teacher experience in the value-added models so that we properly account for the fact that disadvantaged students are more likely to have teachers at a point in their careers when they are less effective.

Table 2 reports the correlation in teacher effectiveness estimates across the value-added specifications that include or exclude classroom covariates by grade. Consistent with Goldhaber et al. (2013), we find very high correlations across the two specifications, although the correlations are

these subgroups of students are stalling (or falling behind) in terms of their mathematical knowledge or skills. Casio and Staiger (2012), for instance, suggest that comparing where students fall in the test distribution across grades may fail to capture differences in accumulated knowledge due to the fact that tests *sample* a wider distribution of knowledge over time. On the other hand, Hill et al. (2008) show that students tend to gain less in standard deviation terms in higher grades, which suggests that a given achievement gap in 8th grade is actually larger than the comparable achievement gap in 3rd grade because it represents more learning at that grade level.

[17] As is shown in the table, disadvantage students are also substantially less likely to take any advanced math course.

[18] Goldhaber et al. (2015) also find differences in the licensure test scores of teachers assigned to advantaged and disadvantaged schools in Washington.

notably lower at the middle school level, particularly in the seventh grade where the correlation is 0.88 as opposed to more 0.98 in elementary grades. This could indicate that peer effects are more important in middle schools, or that there is systematically more sorting of students into classrooms according to teacher effectiveness. As we described above, it is not possible with nonexperimental data to distinguish the degree to which these two explanations might explain the differences in the estimates generated from the different value-added specifications.

Given the relatively high correlations across specifications, one might expect little sensitivity in the estimates of the value-added teacher quality gaps (TQGs) between advantaged and disadvantaged students, but this turns out not to be the case in some grades. We illustrate this in Table 3, which reports differences between disadvantaged and advantaged students in mean teacher value added by grade and value-added specification. At the elementary level, there is relatively little difference in the estimates of TQGs regardless of whether we utilize classroom covariate-adjusted or experience adjusted specifications to generate the value-added measures. In these grades, the estimated TQG is in the neighborhood of -0.01 to -0.03 standard deviations of student achievement on the student math test. These gaps are roughly equivalent to one half to two-thirds (depending on the gap used and the grade of the teacher; see Appendix Table A1) of the return to having a teacher with a year of experience as opposed to a first-year teacher; the estimates are also consistent with prior estimates from other states like Florida (Sass et al., 2010), Massachusetts (Cowan et al., 2017), and North Carolina (Goldhaber et al., 2018).

But at the middle school level, and in Grades 7 and 8 in particular, the estimated gaps are quite sensitive to the value-added specification. For each grade, the estimated value added TQG is larger when using value-added estimates that *do not* include classroom level covariates then when using value-added estimates that *do* include classroom covariates. This is true for both URM and FRL gaps and regardless of whether we do not account for teacher experience (column 1 vs. column 2) or use value-added estimates that account for the experience level of teachers at the time that they taught students in the sample (column 3 vs. column 4). In fact, the value-added measure with classroom covariates suggest that the TQGs *favor* both URM and FRL students in seventh and eighth grades (i.e., disadvantaged students are assigned to more effective teachers in these grades).

By contrast, there is little difference in the estimated TQGs when we adjust for teacher experience. This is likely because, while Table 1 shows gaps in exposure to novice teachers, only about 5% of teachers in the analytic sample have two or fewer years of experience, and thus the experience adjustment does not change the overall gaps substantially. Given that the TQG is sensitive to whether the value-added estimates include classroom covariates, but not to whether we adjust for teacher experience, henceforth we only report findings from the two different experience-adjusted value-added specifications (with and without classroom covariates).

3.4. Analytic Approach

The central goal of our analytic models is to estimate the extent to which teacher assignments between fourth and eighth grade explain gaps between advantaged and disadvantaged students in their eighth grade math test scores and high school course taking. We do this by first estimating models that predict these outcomes as a function of observable student characteristics, and then estimating models that also include controls for the student's teacher assignments between fourth and eighth grade. We then use the differences between the estimated achievement gaps from these two models to estimate the extent to which student achievement gaps would change if we could make the assignment of teachers completely equitable.

Specifically, our baseline model that does *not* account for teacher characteristics is the following:

$$A_i = \alpha_0 + \alpha_1 URM_i + \alpha_2 FRL_{3i} + \alpha_3 X_i + \varepsilon_i \tag{8}$$

In equation 8, the outcome A_i is either the student's eighth grade test score or the number of advanced courses they take in high school. We estimate these outcomes as a function of the student's URM indicator URM_i, the student's third-grade FRL indicator FRL_{3i}, and a vector of student controls X_i that, in different specifications, may include third grade test scores and other observable third grade characteristics (e.g., special education or gifted status). We opt only to control for third-grade student characteristics because observed characteristics may be endogenous to the teacher measures we introduce in later specifications.[19]

We then add controls to the model in equation 8 that account for the teachers to whom the student was assigned in Grades 4–8. In our first specification that does this, we define the vector ρ_{ig} as an indicator for the student's teacher in Grades 4–8 (i.e., *g*=4,...,8), and we directly control for the sequence of teachers to whom the student was assigned in Grades 4–8 by including ρ_{ig} as fixed effects in the model.[20]

$$A_i = \beta_0 + \beta_1 URM_i + \beta_2 FRL_{3i} + \beta_3 X_i + \sum_{g=4}^{8} \rho_{ig} + \varepsilon_i \qquad (9)$$

We are primarily interested in the differences between the estimated coefficient $\hat{\alpha}_1$ from equation 8 and the estimated coefficient $\hat{\beta}_1$ from equation 9, and between the estimated coefficient $\hat{\alpha}_2$ from equation 8 and the estimated coefficient $\hat{\beta}_2$ from equation 9. These differences tell us the extent to which the regression-adjusted achievement gaps between URM and FRL students are explained by the assignment of students to teachers in Grades 4–8.

The model in equation 9 has the advantage of accounting for all the ways that a sequence of teachers may affect future outcomes but it is perhaps not very useful from a policy perspective because

[19] For example, teachers may influence students' program placements (e.g., special education and gifted) in subsequent grades. This endogeneity may not apply to FRL status in later grades (i.e., teacher assignment should not generally affect whether a student receives free or reduced-price lunch) and in models available upon request, we estimate all specifications including FRL status in third through eighth grade.

[20] We also control for teacher experience indicators in these models to account for returns to teaching experience, but estimates from models that do and do not include these experience controls are nearly identical.

it is difficult to imagine designing an intervention that influences the entire sequence of teachers to whom a student is assigned in Grades 4–8. Moreover, it is likely necessary to have some measure of teacher quality that is used to determine teacher assignments. The teacher value-added estimates described in Section 3.2 offer an observable teacher characteristic on which policy makers could plausibly intervene to close achievement gaps. Thus, we estimate a second variant of equation 8 that includes teacher value-added in Grades 4–8 as predictors:

$$A_i = \gamma_0 + \gamma_1 URM_i + \gamma_2 FRL_{3i} + \gamma_3 X_i + \sum_{g=4}^{8} \gamma_g \hat{\tau}_{ig} + \varepsilon_i \qquad (10)$$

In the model in equation 10, the estimated coefficients $\hat{\gamma}_g$ can be interpreted as the partial correlation between the value added of the student's teacher in grade *g* and the outcome A_i. As before, we are interested in the differences between the estimated coefficient $\hat{\alpha}_1$ from equation 8 and the estimated coefficient $\hat{\gamma}_1$ from equation 10, and between the estimated coefficient $\hat{\alpha}_2$ from equation 8 and the estimated coefficient $\hat{\gamma}_2$ from equation 10. These differences tell us the extent to which the regression-adjusted achievement gaps between URM and FRL students are explained by the value added of the student's teachers in Grades 4–8.

Because only about 12% of students in the analytic sample are matched with a single math teacher in every grade from fourth grade through eighth grade, we allow every student in the analytic sample to contribute to the estimates in equations 8–10 by creating a vector of indicators in each grade from Grades 4 through 8 of whether the student is matched to a teacher in grade *g* and including this vector in *all* models (not just models that include teacher indicators or value added). This ensures that differences in estimates across specifications are driven solely by the observed teachers or teacher value-added and not by non-random patterns of missingness in the student-teacher links. We include a school-by-grade indicator (or the mean value added for the school-grade cell) for students not linked to a teacher in a given grade and year.

We interpret the findings from the models described above as descriptive rather than causal given that there are potential sources of bias in the estimates from each of these models. First, the estimates from the model in equation 8 are likely biased because this model does not account for differences in teacher assignments across students, and thus the inequities documented in Table 3 are attributed to the student characteristics in the model. If URM or FRL students tend to be assigned to less effective teachers (as suggested by all value-added estimates at the elementary level and the estimates that do not control for classroom variables in middle school in Table 3), this biases the coefficients on student URM and FRL down in these models. On the other hand, if URM or FRL students tend to be assigned to more effective teachers (as suggested by the estimates that control for classroom variables in middle school in Table 3), this biases the coefficients on student URM and FRL up in these models. That said, this source of bias is not a major concern in our application because we are specifically interested in how the coefficients change when we do and do not control for measures of teacher quality.

A potentially bigger concern is whether the models that do control for teacher assignments are still biased. Specifically, while the models in equations 9 and 10 control for teacher assignments, they only control for students' third grade characteristics. Thus the influence of any time-varying factors from fourth through eighth grade that are correlated with teacher assignments (for the model in equation 9) or with teacher value added (for the model in equation 10) are attributed to the teachers in these models.[21] It is therefore possible that the estimates from equations 9 and 10 may over-attribute outcomes to teacher assignments.

[21] See Rothstein (2010) and Chetty et al. (2014a) for a more extensive discussion of this issue as it relates to value added.

4. Results

4.1 Eighth-Grade Test Scores

Table 4 shows the coefficient estimates for various model specifications that predict students' eighth-grade test scores. We begin with relatively sparse specifications that include only indicators for URM and FRL (column 1), then add in baseline (end of third grade) math and reading test results (column 2), then add to these baseline third-grade student characteristics (column 3),[22] and finally include measures of teacher quality: either direct controls for the teachers to whom students were assigned in Grades 4–8 (column 4); an estimate of teacher value added that does not include classroom covariates (column 5); or an estimate of teacher value added that does include classroom covariates (column 6).[23]

The results from column 1 reflect large differences in end-of-grade eighth grade math achievement by URM and FRL status, about -0.3 and -0.5 standard deviations on the eighth grade test. These achievement gaps are somewhat smaller than the magnitudes of the gaps reported for these subgroups of students in Table 1 because of the overlap between the two groups in URM and FRL status. Controlling for baseline third-grade test scores (column 2) shrinks these gaps considerably, by 0.24 standard deviations for URM and by 0.30 standard deviations for FRL. Put another way, 63% to 75% of the achievement gaps in eighth-grade math achievement appear to be associated with differences between students' third-grade achievement levels. The addition of student baseline covariates (column 3) has little effect on these gaps.[24]

---

[22] As described in Section 3.4, in some specifications we add a vector of fourth grade classroom controls, but these results, available upon request, are strikingly similar to the results that only include third grade student characteristics so we omit them for the sake of brevity.
[23] Note that the additional information about the covariates included in these models is reported in notes at the bottom of the table.
[24] In results not reported but available upon request, we also experiment with adding fourth grade classroom covariates to the models. This addition has little impact on the URM, FRL, third grade test coefficients. We also estimate all models including FRL indicators for third through eighth grade. Patterns with these models are identical to those discussed; however, the coefficient on FRL in third grade is about half as large as the specifications that do not include FRL in additional grades. Including FRL indicators in additional grades has a little impact on the coefficient on URM.

Next, we turn to describing the specifications that account for the quality of teachers to which students are assigned. Including indicators for teacher assignments in Grades 4–8 (column 4) leads to small (compared to adding baseline student test scores), but significant reductions in the estimated advantaged-disadvantaged gaps in eighth-grade achievement.[25] Specifically, accounting for the fourth–eighth grade assignment of teachers leads to a reduction of the URM coefficient of about 4% (-0.079 to -0.076) from the regression adjusted measure of the URM gap; the reduction in the FRL coefficient is much larger (-0.187 to -0.107), about 43% less than the regression adjusted measure of the FRL gap.

Column 5 shows the findings when we replace the teacher indicators with teacher value added (estimated without classroom covariates). The relationship between the value added of teachers across all five grades and eighth-grade achievement are statistically significant and positive, but far larger in the later grades, consistent with the notion that a significant portion of value added fades out over time.[26] It is also worth noting that the estimate of the effect of having a higher value-added teacher in the eighth grade (1.035) is not statistically different than one and is within the range of estimates from Kane et al. (2013).

As was the case with teacher indicators, the inclusion of value added in the model leads to a reduction in the regression-adjusted achievement gaps, but the magnitude of the reduction is much smaller for FRL (-0.187 to -0.148). The fact that value added captures far less of the full effect of teachers could be a product of a downward bias of the coefficients due to measurement error in the value-added estimates (e.g., Schochet and Chiang, 2010).[27] Alternatively, this may reflect the fact that

---

[25] Not surprisingly, an f-test of these teacher indicators suggests that they improve the explanatory power of the model.

[26] Our estimates of fadeout—that the predictive power of sixth grade value added about a third of the predictive power of eighth grade value added on eighth grade test scores—are comparable to estimates of fadeout in teacher effects estimated with different methodologies elsewhere in the literature (Chetty et al., 2014a; Jacob et al., 2011; Kane and Staiger, 2008; Kinsler, 2012; Konstantopoulos & Chung, 2011; Lockwood et al., 2007; McCaffrey et al., 2004; Rothstein, 2010).

[27] Note that, if the model is correctly specified, the shrinkage embedded in the Chetty et al. specification should account for measurement error.

the measure fails to fully account for ways that teachers contribute to students—such as their impacts on student "grit" (Kraft, forthcoming)—that, while not highly correlated with value added, may impact their on future achievement.

In column 6, we report the results with the alternative measure of teacher value added that adjusts for classroom covariates. As was the case with the first specification, the estimates suggest that having a higher value-added teacher in each grade is beneficial for students' eighth-grade achievement, although there is some variation in the point estimates for specific grades across the two specifications. However, unlike the case with the prior value-added estimates, the estimates that include classroom covariate controls show that accounting for the quality of teacher assignment leads to increases in the achievement gaps, represented by the fact that the coefficients on URM and FRL increase in column 6 relative to column 3. This finding is counterintuitive but is consistent with the finding reported above that the classroom covariate specification of value added suggests that disadvantaged students tend to be assigned to higher-quality teachers in Grades 7 and 8.[28] As we discuss below, we believe the dichotomy between what the two value-added specifications suggest about TQGs and eighth-grade achievement has important implications about the validity of the specifications.

There are two reasons why we should be careful about not over-interpreting what the changes in the estimated achievement gaps from any of these models mean for understanding how teacher assignment influences the overall achievement gaps reported in Table 1. First, the coefficients in the models that exclude teacher assignment (e.g. column) may be biased by this exclusion. Second, the strong correlations between URM, FRL, and third grade baseline test scores suggest the changes to the regression-adjusted gap measures may not reflect the changes to the overall gaps associated with teacher assignment. This is discussed more extensively in Section 4.3 below.

---

[28] Note that the coefficients on the import of value added predicting eighth grade achievement are much larger in Grades 7 and 8, indicating that the estimated TQGs *favoring* disadvantaged students in those grades matter much more than the earlier grade TQGs favoring advantaged students.

4.2 High School Course Taking

The findings for models predicting advanced high school math course taking are reported in Table 5. The structure of the table parallels that of Table 4. Columns 1–3 are sparse specifications, column 4 adds a vector of teacher indicators, and columns 5 and 6 replace the vector of teacher indicators with the two different value-added measures. Not surprisingly given the mean differences in advanced course taking, URM students are estimated to take about 0.1 less math courses, and FRL 0.3 less, than students who don't fall into those categories; these differences are about 11% to 33% of the course-taking variable.[29]

As was the case with eighth-grade achievement, we observe that much of the raw advantaged-disadvantaged student gaps in advanced math course taking in high school reported in column 1 is explained by the inclusion of third-grade test scores. In fact, in the case of URM students, what was a gap of about 0.1 courses favoring non-URM students in the unadjusted model (column 1) becomes a gap of 0.03 advanced math courses favoring URM students when we control for third grade tests (column 2). While the positive regression-adjusted gap for URM students may seem surprising, this finding is in line with literature showing that, conditional on prior achievement and/or socioeconomic status, the gaps between URM and non-URM students in educational attainment (e.g., Alexander et al., 1987; Alexander et al., 1982; Bennett & Lutz, 2009; Bennett & Xie, 2003; Kane, 1994; Perna, 2000) and high school course taking (Congar, et al., 2009) disappear or advantage URM students. Including additional student controls (column 3) causes the gap favoring URM students to grow slightly, but, surprisingly, makes the regression-adjusted gap for FRL students even more negative.

The findings when we consider teachers in the model (in both columns 4 and 5) are consistent with the notion that being assigned to better math teachers (as measured by value added) in

[29] One concern with these course-taking models is that there may be substantial differences in the availability of advanced math courses across different schools. We therefore estimate alternative specifications of the course-taking models that include high school fixed effects and report these results in Appendix Table A2. These results are qualitatively similar to the results discussed for Table 5.

elementary and middle schools increases the number of advanced courses students take in high school. Interestingly, the coefficients on value added do not increase monotonically by grade, as was the case for eighth grade test achievement. This may reflect the fact that the skills acquired from teachers (or encouragement from them) in some earlier grades tend to be as important or more important than those acquired in later grades.[30] The change in the coefficients on URM and FRL between columns 3 and columns 4–6 all suggest that URM and FRL students would have done better with more equitable teacher assignments, but the magnitudes of these changes are difficult to interpret. We return to an estimate of the change in the overall gap in high school course taking in Section 4.3.

4.3 Estimating the Import of TQGs in Explaining Student Outcome Gaps

The estimates in Tables 4 and 5 show how the *regression-adjusted* gaps between disadvantaged and advantaged students change between models that do and do not control for teacher assignments, but it is difficult to relate these changes to changes in the *overall* gaps documented in Table 1. We therefore use the estimates in Tables 4 and 5 to compare the average predicted outcomes for advantaged and disadvantaged students under the different approaches to accounting for teacher assignments.

We report the differences in these predicted outcomes between disadvantaged and advantaged students in Table 6. While models in Tables 4 and 5 net out differences in student characteristics between URM/non-URM and FRL/non-FRL students, the gaps in the predicted outcomes reported in Table 6 account for differences in student characteristics between URM/non-URM and FRL/non-FRL students and the returns to those student characteristics. These gaps in predicted outcomes better reflect the observed gaps that are actually realized between URM/non-URM and FRL/non-FRL students and how teacher assignments impact those observed gaps.

[30] See, for instance, the relatively large coefficients on value added in the fifth grade as compared to the sixth.

The estimates in column 1 of Table 6 are the average predicted gaps from the estimates in column 3 of Table 4 (for Panel A) and column 3 of Table 5 (for Panel B) and are very close to the average outcome gaps reported in Table 1 (as we would expect). The estimates in column 2 of Table 6 are derived from the estimates in column 4 of Table 4 (for Panel A) and column 4 of Table 5 (for Panel B) and show that the small changes in the regression-adjusted gaps in these tables translate into large changes in the overall outcome gaps between advantaged and disadvantaged students. Specifically, the eighth grade URM achievement gap drops from about 0.6 to about 0.5 in models that control for fourth through eighth grade teacher assignments, while the eighth grade FRL achievement gap drops from about 0.66 to 0.52. These changes represent 16% and 21% reductions in the overall achievement gap, respectively. The percent reductions in the overall high school course taking gaps (Panel B of Table 6) are even larger, about 33% for both URM and FRL gaps.

The estimates in column 3 of Table 6 are derived from the estimates in column 5 of Table 4 (for Panel A) and column 5 of Table 5 (for Panel B) and show that accounting for teacher value added without classroom controls in Grades 4–8 drops the predicted outcome gap between 6% and 8% for both URM and FRL students and for both outcomes (eighth grade achievement and high school advanced course taking). These decreases are more modest than the decreases resulting from directly accounting for teacher assignments but do account for about half of the total relationship between teacher assignments and students' later outcomes. Moreover, the absolute drop in the overall achievement gap for URM and FRL students—.04 and .06 standard deviations, respectively—are within the range of estimates in Table A1 for the returns to the first year of teaching experience, and thus are certainly educationally meaningful reductions. Finally, the estimates in column 6 of Tables 4 and 5, from models including value added with classroom controls, suggest that test gaps narrowly increase and course-taking gaps only narrowly decrease for both URM and FRL students.

5. Discussion and Conclusions

One of the most consistent goals of policy makers is identifying ways of eliminating or mitigating the persistent gaps in student achievement between advantaged and disadvantaged students. While not a focus of the paper, it is worth reiterating that our findings show that much of the gaps we observe in eighth-grade tests and high school course taking can be explained by achievement differentials that existed in the third grade. This suggests that policy makers wishing to alleviate later achievement gaps either need to intervene earlier in a student's academic career or be far more aggressive after the third grade to address academic deficiencies.

In demonstrating gaps in student achievement and course taking, we provide further evidence that gaps for FRL students are 1.5 to 2 times larger than those for URM students in regression-adjusted models predicting eighth grade math achievement.[31] And models predicting the number of advanced math courses that control for third-grade test scores show a similar sharp reduction in the outcome gap for FRL students, and a *conditional advantage for URM students*.

The estimates from models predicting student outcomes generally support the importance of considering teacher assignment as one policy lever for addressing achievement gaps. Specifically, these models predict that if URM and non-URM students had equivalent teachers throughout Grades 4–8, gaps in eighth grade math test scores would be 16% smaller, while for FRL students the gap would fall by 21%. Our models predict that the reduction in the gaps would be even larger with respect to the number of advanced math courses taken in high school, dropping by 33% for both URM and FRL students.

Our findings also provide more evidence that value-added measures of teacher quality are a useful measure in that they predict not only later student test scores, but high school course taking as well. Unfortunately, using value added as a lever for promoting equity is not as simple as it might seem,

[31] This finding is consistent with research showing that the achievement gap between poor and non-poor students has widened over time going from half the size of the Black-White achievement to nearly double the size (Reardon, 2011).

as we find that the conclusions from models that consider value added are somewhat sensitive to the model specification. The results that consider value-added models that *exclude* classroom covariates are more closely aligned with the findings that account for teacher quality directly through the inclusion of classroom indicators. This, combined with the fact that the TQGs based on models without classroom controls are consistent with observed gaps in teacher experience in middle school and observed gaps in all measures in elementary school, suggests to us that the non-classroom covariate value added is likely to be a more valid indicator of teacher quality for this application. Clearly, however, we need more evidence on this issue—particularly at the middle school level where the two value-added teacher quality measures diverge—as well as other issues such as the extent to which teachers in one subject influence student outcomes in other subjects (e.g., Koedel, 2009).

When we consider our preferred specification of teacher value added, the models suggest that about half of the influence of teachers on future test scores can be explained by the value added of those teachers, while a quarter of teachers' influence on high school course taking can be explained by the value added of those teachers. This suggests that interventions intended to eliminate TQGs between advantaged and disadvantaged students could have educationally meaningful impacts on student achievement gaps.[32] But while teacher quality gaps appear to be an important contributor to gaps in longer-term student outcomes, eliminating them should not be viewed as a panacea for eliminating student outcome gaps; for example, even taking our results as causal, no more than 10% of the eighth-grade achievement gap between FRL and non-FRL students can be explained by differences in the value added of students' teachers in fourth through eighth grade.

[32] This finding is timely because equalizing teacher quality across student types is the focus of the Department of Education's Excellent Educators for All Initiative in 2014, which required all states to create plans to reduce equity gaps in education. For instance, strategies included improving teacher preparation and cultural competency; providing financial incentives for effective teachers, often targeted at teachers in majority-minority and/or high-poverty schools; providing teacher preparation programs or school districts with data on teacher performance of newly certified teachers or teachers moving in or out of schools and districts to aid in staffing decisions; or developing strategies to recruit new teachers, including minority teachers, to reduce teacher shortages. A few use value added; New Mexico, for instance, provides bonuses to highly effective teachers.

References

Aaronson, D., Barrow, L., & Sander, W. (2007). Teachers and Student Achievement in the Chicago Public High Schools. *Journal of Labor Economics*, 25(1), 95–135.

Alexander, K. L., Holupka, S., & Pallas, A. M. (1987). Social Background and Academic Determinants of Two-Year versus Four-Year College Attendance: Evidence from Two Cohorts a Decade Apart. *American Journal of Education*, 96(1), 56-80.

Alexander, K. L., Riordan, C., Fennessey, J., & Pallas, A. M. (1982). Social Background, Academic Resources, and College Graduation: Recent Evidence from the National Longitudinal Survey. *American Journal of Education*, 90(4), 315-333.

Aud, S., Planty, M., & Hussar, W. J. (2011). Condition of Education 2011. Government Printing Office.

Bacher-Hicks, A., Kane, T. J., & Staiger, D. O. (2014). Validating Teacher Effect Estimates using Changes in Teacher Assignments in Los Angeles (No. 20657). Cambridge, MA: National Bureau of Economic Research.

Backes, B., Cowan, J., Goldhaber, D., Koedel, C., Miller, L. C., & Xu, Z. (2018). The common core conundrum: To what extent should we worry that changes to assessments will affect test-based measures of teacher performance? *Economics of Education Review*, 62, 48-65.

Bennett, P. R., & Lutz, A. (2009). How African American Is the Net Black Advantage? Differences in College Attendance Among Immigrant Blacks, Native Blacks, and Whites. *Sociology of Education*, 82(1), 70-99.

Bennett, P. R., & Xie, Y. (2003). Revisiting Racial Differences in College Attendance: The Role of Historically Black Colleges and Universities. *American Sociological Review*, 68(4), 567-580.

Betts, J. R., Zau, A., & Rice, L. (2003). Determinants of student achievement: New evidence from San Diego (pp. 1-5821). San Francisco: Public Policy Institute of California.

Burkam, D. T., Lee, V. E., & Smerdon, B. A. (2003). Mathematics, foreign language, and science coursetaking and the NELS: 88 transcript data. US Department of Education, Institute of Education Statistics, National Center for Education Statistics.

Cascio, E. U., & Staiger, D. O. (2012). Knowledge, tests, and fadeout in educational interventions (No. w18038). National Bureau of Economic Research.

Chamberlain, G. E. (2013). Predictive effects of teachers and schools on test scores, college attendance, and earnings. *Proceedings of the National Academy of Sciences*, 110(43), 17176–17182. doi:10.1073/pnas.1315746110

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *The American Economic Review*, 104(9), 2633-2679.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. (2005). Who teaches whom? Race and the distribution of novice teachers. *Economics of Education review*, *24*(4), 377-392.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2009). The academic achievement gap in grades 3 to 8. *The Review of Economics and Statistics*, *91*(2), 398-419.

Conger, D., Long, M. C. and Iatarola, P. (2009), Explaining race, poverty, and gender disparities in advanced course-taking. *Journal of Policy Analysis and Management*, 28: 555–576.

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

Cowan, J., Goldhaber, D., & Theobald, R. (2017). Teacher equity gaps in Massachusetts. http://www.doe.mass.edu/research/reports/2017/10teacher-equity.pdf.

Gamoran, A. (1987). The stratification of high school learning opportunities. *Sociology of Education*, 135-155.

Gershenson, S. (2016). Linking teacher quality, student attendance, and student achievement. *Education Finance and Policy* 11(2), 125-149.

Gershenson, S., Holt, S. B., & Papageorge, N. W. (2016). Who believes in me? The effect of student–teacher demographic match on teacher expectations. *Economics of Education Review*, 52, 209-224.

Goldhaber, D., & Hansen, M. (2013). Is it just a bad class? Assessing the long-term stability of estimated teacher performance. *Economica*, 80(319), 589-612.

Goldhaber, Dan, Gabele, Brian, and Walch, Joe. (2013). Does the Model Matter? Exploring the Relationship Between Different Student Achievement-Based Teacher Assessments. *Statistics and Public Policy*. 1(1): 28–39.

Goldhaber, D., Gratz, T., and Theobald, R. (2017). What's in a teacher test? Assessing the relationship between teacher licensure test scores and student secondary STEM achievement and course taking. *Economics of Education Review* 61, 112-129.

Goldhaber, D., Lavery, L., and Theobald, R. (2015). Uneven playing field? Assessing the teacher quality gap between advantaged and disadvantaged students. *Educational Researcher*, 44(5), 293-307.

Goldhaber, D., Quince, V., & Theobald, R. (2016). Reconciling different estimates of teacher quality gaps based on value added. CALDER Policy Brief 14.

Goldhaber, D., Quince, V., & Theobald, R. (2018). Has it always been this way? Tracing the evolution of teacher quality gaps in U.S. public schools. *American Educational Research Journal*, doi 10.3102/0002831217733445.

Gottfried, M. A. (2015). The Influence of Applied STEM Coursetaking on Advanced Mathematics and Science Coursetaking. *The Journal of Educational Research*, 108(5), 382-399.

Heckman, J. J., & LaFontaine, P. A. (2010). The American high school graduation rate: Trends and levels. *The Review of Economics and Statistics*, 92(2), 244-262.

Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives*, 2(3), 172-177.

Hubbard, D. (2017). More gains than score gains? High school quality and college success. University of Michigan Working Paper.

Isenberg, E., Max, J., Gleason, P., Johnson, M., Deutsch, J., & Hansen, M. (2016). Do low-income students have equal access to effective teachers? Evidence from 26 districts. NCEE 2017-4007. Institute of Education Sciences.

Jackson, C. K. (Forthcoming). What do test scores miss? The importance of teacher effects on non-test-score outcomes. *Journal of Political Economy*.

Jackson, C. K., Johnson, R. C., & Persico, C. (2016). The effects of school spending on educational and economic outcomes: Evidence from school finance reforms. *The Quarterly Journal of Economics*, 131(1), 157-218.

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

Kalogrides, D., & Loeb, S. (2013). Different teachers, different peers: The magnitude of student sorting within schools. *Educational Researcher*, 42(6), 304-316.

Kane, T. J. (1994). Race, College Attendance and College Completion: Brookings Institution, Washington, DC; Kennedy School of Government, Cambridge, MA.

Kane, T. J., McCaffrey, D. F., Miller, T., & Staiger, D. O. (2013). Have We Identified Effective Teachers? Seattle, WA: Bill and Melinda Gates Foundation.

Kane, T. J., & Staiger, D. O. (2008). Estimating teacher impacts on student achievement: An experimental evaluation. National Bureau of Economic Research Working Paper Series, 14607.

Kelly, S. (2009). The black-white gap in mathematics course taking. *Sociology of Education*, 82(1), 47-69.

Kena, G., Musu-Gillette, L., Robinson, J., Wang, X., Rathbuun, A., Zhang, J., Wilkinson-Flicker, S., Barmer, A., and Dunlop Velez, E. (2015). The Condition of Education 2015. (NCES 2015-144) U.S. Department of Education, National Center for Education Statistics. Washington, DC.

Kinsler, J. (2012). Beyond levels and growth: Estimating teacher value-added and its persistence. *Journal of Human Resources*, 47(3), 722–753.

Koedel, C. (2009). An empirical analysis of teacher spillover effects in secondary school. *Economics of Education Review*, 28(6), 682-692.

Koedel, C., Mihaly, K., & Rockoff, J. E. (2015). Value-added modeling: A review. *Economics of Education Review*, 47, 180-195.

Konstantopoulos, S., & Chung, V. (2011). The persistence of teacher effects in elementary grades. *American Educational Research Journal*, 48(2), 361-386.

Kraft, M. A. (Forthcoming). Teacher Effects on Complex Cognitive Skills and Social-Emotional Competencies. *Journal of Human Resources*.

Kraft, M. A., & Papay, J. P. (2014). Can professional environments in schools promote teacher development? Explaining heterogeneity in returns to teaching experience. *Educational Evaluation and Policy Analysis*, 36(4), 476-500.

Ladd, H.F. & Sorenson, L.C. (2017). Returns to teacher experience: student achievement and motivations in middle school. *Education Finance and Policy*. 12(2), 241-279.

Lafortune, J., Rothstein, J., & Schanzenbach, D. W. (2016). School finance reform and the distribution of student achievement (No. w22011). National Bureau of Economic Research.

Lankford, H., Loeb, S., & Wyckoff, J. (2002). Teacher sorting and the plight of urban schools: A descriptive analysis. *Educational Evaluation and Policy Analysis*, 24(1), 37-62.

Leandro v. State, 88 S.E.2d 249 (N.C. 1997).

Lee, J. (2002). Racial and ethnic achievement gap trends: Reversing the progress toward equity? *Educational Researcher*, 31(1), 3-12.

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

Long, M. C., Conger, D., & Iatarola, P. (2012). Effects of high school course-taking on secondary and postsecondary success. *American Educational Research Journal*, 49(2), 285-322.

Long, M. C., Iatarola, P., & Conger, D. (2009). Explaining gaps in readiness for college-level math: The role of high school courses. *Education*, 4(1), 1-33.

McCaffrey, D.F., Lockwood, J. R., Koretz, D., Louis, T. A., & Hamilton, L. (2004). Models for value-added modeling of teacher effects. *Journal of Educational and Behavioral Statistics*. 29(1).

McCaffrey, D. F., Sass, T. R., Lockwood, J. R., & Mihaly, K. (2009). The intertemporal variability of teacher effect estimates. *Education Finance and Policy*, 4(4), 572-606.

Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis*, 26(3), 237-257.

Perna, L. W. (2000). Differences in the Decision to Attend College among African Americans, Hispanics, and Whites. *Journal of Higher Education*, 71(2), 117-141.

Reardon, S. F. (2011). The widening academic achievement gap between the rich and the poor: New evidence and possible explanations. *Whither Opportunity*, 91-116.

Rich, M. (2014). U.S. to focus on equity in assigning of teachers. *The New York Times*, November 10, 2014.

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

Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. The American Economic Review, 94(2), 247-252.

Rose, H., & Betts, J. R. (2004). The effect of high school courses on earnings. *The Review of Economics and Statistics*, 86(2), 497-513.

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

Sass, T. R., Hannaway, J., Xu, Z., Figlio, D. N., & Feng, L. (2012). Value added of teachers in high-poverty schools and lower poverty schools. *Journal of Urban Economics*, 72(2), 104-122.

Schochet, P. Z., & Chiang, H. S. (2010). Error Rates in Measuring Teacher and School Performance Based on Student Test Score Gains. NCEE 2010-4004. National Center for Education Evaluation and Regional Assistance.

Stepner, M. (2013). vam [STATA package].

Tyson, W., Lee, R., Borman, K. M., & Hanson, M. A. (2007). Science, technology, engineering, and mathematics (STEM) pathways: High school science and math coursework and postsecondary degree attainment. *Journal of Education for Students Placed at Risk*, 12(3), 243-270.

Zamarro, G., Engberg, J., Saavedra, J. E., & Steele, J. (2015). Disentangling Disadvantage: Can We Distinguish Good Teaching From Classroom Composition? *Journal of Research on Educational Effectiveness*, 8(1), 84-111.

Table 1. Student Baseline Summary Statistics

	Unrestricted sample	eighth Grade Achievement Analytic Sample					High School Course Taking Analytic Sample				
		All	URM	non-URM	FRL	non-FRL	All	URM	non-URM	FRL	non-FRL
	1	2	3	4	5	6	7	8	9	10	11
Baseline characteristics											
URM student	0.258	0.256	1.000	0.000	0.447	0.106	0.225	1.000	0.000	0.443	0.092
Third-Grade FRL	0.449	0.441	0.769	0.328	1.000	0.000	0.379	0.746	0.273	1.000	0.000
Third-Grade Math Score	-0.000	0.032	-0.421	0.188	-0.316	0.307	0.095	-0.402	0.240	-0.289	0.330
	(1.000)	(0.977)	(0.934)	(0.942)	(0.931)	(0.923)	(0.963)	(0.962)	(0.914)	(0.951)	(0.892)
Third-Grade Reading Score	0.000	0.026	-0.408	0.176	-0.328	0.306	0.084	-0.385	0.221	-0.302	0.320
	(1.000)	(0.976)	(0.936)	(0.945)	(0.940)	(0.912)	(0.970)	(0.961)	(0.929)	(0.960)	(0.898)
Female	0.490	0.492	0.495	0.491	0.495	0.490	0.499	0.507	0.496	0.507	0.493
Third-Grade Special Education	0.127	0.118	0.124	0.116	0.141	0.100	0.105	0.113	0.102	0.126	0.092
Third-Grade English Language Learner	0.094	0.096	0.278	0.033	0.190	0.021	0.091	0.297	0.030	0.207	0.019
Third-Grade Gifted	0.030	0.029	0.009	0.036	0.009	0.045	0.031	0.011	0.037	0.010	0.044
N	437123	330539	84744	245795	145811	184728	104001	23438	80563	39464	64537
Teacher experience											
Fourth Grade <= 2 Years Experience	0.048	0.051	0.065	0.047	0.056	0.048	0.068	0.082	0.064	0.075	0.063
Fifth Grade <= 2 Years Experience	0.038	0.040	0.050	0.037	0.045	0.036	0.048	0.057	0.046	0.054	0.044
Sixth Grade <= 2 Years Experience	0.034	0.037	0.043	0.034	0.038	0.035	0.024	0.026	0.023	0.024	0.024
Seventh Grade <= 2 Years Experience	0.041	0.044	0.055	0.040	0.049	0.039	0.046	0.061	0.042	0.055	0.041
Eighth Grade <= 2 Years Experience	0.038	0.044	0.052	0.042	0.047	0.042	0.038	0.048	0.035	0.043	0.035
Outcomes											
Eighth-Grade Math Score	0.042	0.066	-0.374	0.218	-0.301	0.356	0.155	-0.279	0.286	-0.200	0.378
	(0.989)	(0.978)	(0.880)	(0.965)	(0.896)	(0.943)	(0.959)	(0.858)	(0.950)	(0.882)	(0.939)
N	344537	330539	84744	245795	145811	184728	97635	22605	75030	37667	59968
Any advanced math course in HS	0.419	0.408	0.292	0.443	0.279	0.489	0.421	0.296	0.457	0.285	0.504
Number of advanced math courses in HS	0.631	0.597	0.393	0.659	0.372	0.739	0.634	0.404	0.702	0.385	0.787
	(0.877)	(0.843)	(0.692)	(0.874)	(0.672)	(0.906)	(0.878)	(0.706)	(0.912)	(0.690)	(0.944)
N	106525	97635	22605	75030	37667	59968	104001	23438	80563	39464	64537

*NOTE: Sample sizes for teacher experience range from 223,938 in sixth grade to 265,400 in fifth grade.

Table 2. Correlations Between VAM Estimates With and Without Classroom Controls

	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
Correlation	0.992	0.983	0.971	0.883	0.914

Table 3. Estimated Teacher Quality Gaps by Grade and VAM Specification

		1	2	3	4
non-URM vs. URM	Grade 4	-0.029*** (0.001)	-0.025*** (0.001)	-0.030*** (0.001)	-0.024*** (0.001)
	Grade 5	-0.012*** (0.001)	-0.009*** (0.001)	-0.012*** (0.001)	-0.008*** (0.001)
	Grade 6	-0.030*** (0.001)	-0.003** (0.001)	-0.031*** (0.001)	-0.002* (0.001)
	Grade 7	-0.025*** (0.001)	0.017*** (0.001)	-0.026*** (0.001)	0.017*** (0.001)
	Grade 8	-0.026*** (0.001)	0.003*** (0.001)	-0.026*** (0.001)	0.002*** (0.001)
non-FRL vs. FRL	Grade 4	-0.031*** (0.001)	-0.025*** (0.001)	-0.031*** (0.001)	-0.024*** (0.001)
	Grade 5	-0.017*** (0.001)	-0.014*** (0.001)	-0.017*** (0.001)	-0.013*** (0.001)
	Grade 6	-0.033*** (0.001)	-0.000 (0.001)	-0.033*** (0.001)	0.000 (0.001)
	Grade 7	-0.041*** (0.001)	0.006*** (0.001)	-0.042*** (0.001)	0.006*** (0.001)
	Grade 8	-0.030*** (0.001)	0.003*** (0.001)	-0.030*** (0.001)	0.003*** (0.001)
VAMs include classroom controls			X		X
VAMs include experience adjustment				X	X

*NOTE: p-values from two-sided t-test: *p<.05; **p<.01; ***p<.001.

Table 4. Regressions Predicting Eighth-Grade Math Performance

	(1)	(2)	(3)	(4)	(5)	(6)
URM student	-0.319***	-0.083***	-0.079***	-0.076***	-0.074***	-0.094***
	(0.007)	(0.005)	(0.005)	(0.003)	(0.004)	(0.004)
Third-Grade FRL	-0.477***	-0.177***	-0.187***	-0.107***	-0.148***	-0.190***
	(0.007)	(0.004)	(0.004)	(0.003)	(0.003)	(0.004)
Third-Grade Math Score		0.473***	0.455***	0.400***	0.436***	0.449***
		(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
Third-Grade Reading Score		0.197***	0.190***	0.157***	0.178***	0.186***
		(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Fourth-Grade Teacher Value Added				Teacher Indicators	0.048***	0.080***
					(0.011)	(0.012)
Fifth-Grade Teacher Value Added					0.171***	0.227***
					(0.015)	(0.015)
Sixth-Grade Teacher Value Added					0.344***	0.415***
					(0.015)	(0.016)
Seventh-Grade Teacher Value Added					0.693***	0.695***
					(0.022)	(0.027)
Eighth-Grade Teacher Value Added					1.035***	1.025***
					(0.026)	(0.031)
N	330539	330539	330539	330539	330539	330539
Third-grade student controls			X	X	X	X
Value added without class controls					X	
Value added with class controls						X

*NOTE: p-values from two-sided t-test: *p<.05; **p<.01; ***p<.001. All models also include indicators for missing teacher links in Grades 4–8. Third-grade student controls include gender, gifted status, special education status, learning disability status, and English Language Learner status. Standard errors clustered at the eighth-grade classroom level.

Table 5. Regressions Predicting Number of Advanced Math Courses in High School

	(1)	(2)	(3)	(4)	(5)	(6)
URM student	-0.102***	0.034***	0.045***	0.027***	0.048***	0.040***
	(0.011)	(0.010)	(0.010)	(0.007)	(0.010)	(0.009)
Third-Grade FRL	-0.333***	-0.174***	-0.188***	-0.127***	-0.171***	-0.187***
	(0.011)	(0.008)	(0.008)	(0.006)	(0.008)	(0.008)
Third-Grade Math Score		0.291***	0.276***	0.183***	0.266***	0.271***
		(0.007)	(0.006)	(0.005)	(0.006)	(0.006)
Third-Grade Reading Score		0.079***	0.087***	0.060***	0.082***	0.085***
		(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Fourth-Grade Teacher Value Added				Teacher Indicators	0.139***	0.148***
					(0.032)	(0.032)
Fifth-Grade Teacher Value Added					0.221***	0.281***
					(0.044)	(0.043)
Sixth-Grade Teacher Value Added					0.043	0.070
					(0.049)	(0.049)
Seventh-Grade Teacher Value Added					0.484***	0.381***
					(0.060)	(0.074)
Eighth-Grade Teacher Value Added					0.289***	0.117
					(0.082)	(0.090)
N	104001	104001	104001	104001	104001	104001
Third-grade student controls			X	X	X	X
Value added without class controls					X	
Value added with class controls						X

*NOTE: p-values from two-sided t-test: *p<.05; **p<.01; ***p<.001. All models also include indicators for missing teacher links in Grades 4–8. Third-grade student controls include gender, gifted status, special education status, learning disability status, and English Language Learner status. Standard errors clustered at the eighth-grade classroom level.

Table 6. Predicted Outcome Gaps

	(1)	(2)	(3)	(4)
	Observed teacher assignments	Under same teacher assignments	Under same assignments by teacher VA, no class controls	Under same assignments by teacher VA, class controls
Panel A. Student achievement in eighth grade				
Predicted URM math achievement gap	-0.593	-0.498	-0.548	-0.607
Predicted FRL math achievement gap	-0.657	-0.519	-0.596	-0.662
Panel B. Number of advanced high school courses				
Predicted URM advanced math courses gap	-0.298	-0.201	-0.276	-0.300
Predicted FRL advanced math courses gap	-0.402	-0.265	-0.374	-0.400

Table A1. Returns to Teacher Experience by Grade

	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
1 year	0.067*** (0.005)	0.021*** (0.006)	0.071*** (0.006)	0.065*** (0.006)	0.037*** (0.007)
2 years	0.087*** (0.006)	0.059*** (0.006)	0.075*** (0.006)	0.064*** (0.007)	0.035*** (0.007)
3 years	0.098*** (0.006)	0.049*** (0.006)	0.091*** (0.006)	0.086*** (0.008)	0.058*** (0.008)
4 years	0.108*** (0.006)	0.060*** (0.006)	0.104*** (0.007)	0.057*** (0.008)	0.041*** (0.008)
5 years	0.113*** (0.006)	0.069*** (0.006)	0.109*** (0.007)	0.091*** (0.008)	0.088*** (0.008)
6 years	0.107*** (0.006)	0.071*** (0.006)	0.092*** (0.007)	0.092*** (0.008)	0.073*** (0.008)
7 years	0.109*** (0.007)	0.061*** (0.007)	0.119*** (0.007)	0.068*** (0.008)	0.035*** (0.008)
8 years	0.118*** (0.007)	0.067*** (0.007)	0.131*** (0.007)	0.099*** (0.008)	0.053*** (0.009)
9 or more years	0.120*** (0.007)	0.082*** (0.006)	0.143*** (0.007)	0.087*** (0.008)	0.049*** (0.009)

*NOTE: p-values from two-sided t-test: *p<.05; **p<.01; ***p<.001.

Table A2. Regressions Predicting Number of Advanced Math Courses in High School with School Fixed Effects

	(1)	(2)	(3)	(4)	(5)	(6)
URM student	-0.139***	-0.011	0.016**	0.020**	0.018***	0.015**
	(0.008)	(0.007)	(0.007)	(0.008)	(0.007)	(0.007)
Third-Grade FRL	-0.284***	-0.153***	-0.165***	-0.119***	-0.158***	-0.164***
	(0.008)	(0.006)	(0.006)	(0.007)	(0.006)	(0.006)
Third-Grade Math Score		0.280***	0.266***	0.183***	0.258***	0.263***
		(0.006)	(0.005)	(0.006)	(0.005)	(0.005)
Third-Grade Reading Score		0.081***	0.085***	0.060***	0.082***	0.084***
		(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Fourth-Grade Teacher Value Added					0.094***	0.097***
					(0.018)	(0.019)
Fifth-Grade Teacher Value Added					0.156***	0.196***
					(0.027)	(0.027)
Sixth-Grade Teacher Value Added				Teacher Indicators	0.125***	0.135***
					(0.030)	(0.030)
Seventh-Grade Teacher Value Added					0.638***	0.482***
					(0.049)	(0.061)
Eighth-Grade Teacher Value Added					0.310***	0.053
					(0.057)	(0.063)
N	104001	104001	104001	104001	104001	104001
Third-grade student controls			X	X	X	X
Value added without class controls					X	
Value added with class controls						X