Research Article

Educational Evaluation and Policy Analysis
June 2022, Vol. 44, No. 2, pp. 234–256
DOI: 10.3102/01623737211040511
Article reuse guidelines: sagepub.com/journals-permissions

Do Low-Income Students Have Equal Access to Effective Teachers?

Eric Isenberg
Westat

Jeffrey Max
Philip Gleason
Jonah Deutsch
Mathematica

We examine access to effective teachers for low-income students in 26 geographically dispersed school districts over a 5-year period. We measure teacher effectiveness using a value-added model that accounts for measurement error in prior test scores and peer effects. Differences between the average value added of teachers of high- and low-income students are 0.005 standard deviations in English/language arts and 0.004 standard deviations in math. Differences between teachers of Black, Hispanic, and White students are also small. Rearranging teachers to obtain perfect equity would do little to narrow the sizable student achievement gap between low- and high-income students. We also show that a higher proportion of novice teachers in high-poverty schools contributes negligibly to differences in access to effective teachers.

Keywords: *achievement gap, equity, teacher characteristics, teacher research, teacher assessment, econometric analysis, regression analyses*

Inequality in educational outcomes is substantial and persistent in the United States. Recent evidence suggests that students from high-income families outperform those from low-income families on achievement tests by considerable amounts, and that this achievement gap has widened over the past 30 years (Reardon, 2011). Inequality persists in long-term educational outcomes as well, with high-income students more likely to attend college (Chetty, Hendren, et al., 2014) and obtain college degrees (U.S. Department of Education, National Center for Education Statistics, 2014).

One possible explanation for the substantial achievement gap is that high-income students may attend better schools and have more effective teachers. Evidence shows that teachers vary a great deal in their effectiveness (Kane et al., 2008; Koedel & Betts, 2011). In addition, students taught by the best teachers not only achieve higher test scores but also have better outcomes in the long run, including greater likelihood of college attendance and higher wages (Chetty, Friedman, & Rockoff, 2014a).

Under the Every Student Succeeds Act (ESSA) of 2015, states are required to track whether low-income students have equitable access to effective teachers and develop plans to ensure that “low-income and minority children enrolled in [Title I] schools . . . are not served at disproportionate rates by ineffective, out-of-field, or inexperienced teachers” (ESEA Section

1111(g)(1)(B)). ESSA also requires districts receiving Title I, Part A funds to submit a plan describing how they will curb these disparities (ESEA Section 1112(b)(2)), and ESSA provides funding for states and districts to address inequitable access to effective teachers, including the use of funds provided under Title II, Part A (ESEA Section 2101(c)(4)(B)(iii), SEA Section 2103(b)(3)(B)).

But it is not obvious that there are inequities in access large enough to contribute meaningfully to achievement gaps, whether defined as the differences between income groups or between Black and White students or Hispanic and non-Hispanic White students. Although several past studies have found that low-income students generally have less qualified teachers than high-income students—based on measures such as years of teaching experience, teacher test scores, certification status, and educational attainment (Boyd et al., 2008; Clotfelter et al., 2007; Schultz, 2014)—this does not necessarily imply that they also have substantially less effective teachers. Most research has found no consistent link between teachers' effectiveness in increasing student learning and these types of qualifications, except for teaching experience (Constantine et al., 2009; Hanushek & Rivkin, 2006; Harris & Sass, 2011; Kane et al., 2008). And while low-income, Black, and Hispanic students are more likely to be taught by novices, the extent to which this contributes to the student achievement gap is unknown.

Recent studies have compared the effectiveness of high- and low-income students' teachers based on value-added estimates. Most studies find that teachers of low-income students and teachers in high-poverty schools are less effective on average, but the magnitude of inequity is usually modest, especially in the core subjects of English/language arts (ELA) and math. The most equitable results were obtained in Chetty, Friedman, and Rockoff (2014b) and Steele et al. (2014). Chetty, Friedman, and Rockoff (2014b), using data from a large urban district, reported that a US$10,000 increase in parental income is associated with an increase in teacher value added of 0.00084 standard deviations of student achievement in ELA and math. Steele et al. (2014) examined three school districts and a charter school consortium and found that in two of the districts and the charter school consortium, disadvantaged students (defined as being both low-income and minority) have more effective teachers, though the differences are modest. Three studies that used data from North Carolina appear to find slightly more inequity. Sass et al. (2012) measured differences in average value added between high- and low-poverty elementary schools in Florida and North Carolina. The authors defined high-poverty schools as those with more than 70% of students eligible for a free or reduced-price lunch (FRL). The differences in value added range from 0.019 to 0.044 for the two states. Mansfield (2015) reported a difference of 0.03 between high schools in the top and bottom poverty quartile in North Carolina. This study combined information across a range of end-of-course tests. Goldhaber et al. examined access to effective teachers in Washington state (Goldhaber et al., 2015, 2016b) and, in one case, also in North Carolina (Goldhaber et al., 2016a). In both states, across ELA and math, in a given year, the average value added of the teachers of FRL students is lower than the average value added of teachers of non-FRL students in a range that extends from about 0.025 to 0.035.[1] Using a more geographically dispersed sample of 10 large districts in six states, Glazerman and Max (2011) found that teachers in the top quintile of value added are less common in high-poverty middle schools, but equally present in high- and low-poverty elementary schools. For example, 15% of math teachers in the highest poverty middle schools were highest performing, compared with 29% in the lowest poverty schools.

Several studies also examined differences between minority and nonminority students in access to effective teachers, with a similar pattern of findings to those that focused on students' income status. Alternate ways of defining student groups for measuring the difference in access to effective teachers are likely to be correlated, given the correlation between race, ethnicity, and income. However, due to past and current discrimination and segregation in housing and schools for minority students, definitions based on race and ethnicity are useful to examine apart from strictly income-based differences in access. Chetty, Friedman, and Rockoff (2014b) found very small differences by race—Black and

Hispanic students had teachers with value-added scores 0.001 lower than White students, on average. Both Hanselman (2018) and Mansfield (2015) found somewhat larger gaps by race, but each concluded that these gaps contributed modestly to overall differences by race in student achievement. Goldhaber et al. (2015) found that the average value added of fourth- and seventh-grade teachers of underrepresented minority students was 0.023 to 0.046 lower than those of other students. Finally, Steele et al. (2015) compared average teacher value added in a large, urban, southern district across schools with different proportions of Black and Hispanic students. When comparing schools in the top and bottom quartile based on the proportion of minority students, differences in average value added are 0.062 in ELA and 0.044 in math. The authors found much larger gaps in science (0.188) and social studies (0.163).

Two sets of questions emerge from this literature. First, how should one interpret the magnitude of these findings? Most of the reported differences are below 0.05 standard deviations of student achievement and consider only differences covering a single year. An effect of this size would generally be considered small if it were the impact of an educational intervention implemented during a given period (such as a year) like a grade-specific curriculum reform or small-group pullout program. However, unlike a typical intervention, if low-income students are more likely to have less effective teachers year after year, the key questions are how the effect accumulates over time and what contribution it makes to the student achievement gap. In this sense, the relevant comparison might be with a whole-school intervention such as attending a charter school versus a traditional public school, and considering multiple years is important.[2] Previous research has not investigated the longitudinal implications of low-income students having less effective teachers over multiple years. Second, what accounts for the variation in results across studies? Are differences driven by variations across states and districts in district policies, residential segregation, and other real factors? Or are they mainly an artifact of different statistical methods, especially in calculating value-added estimates to measure teacher effectiveness?

Our study adds to the literature on access to effective teachers in three main ways:

1. We develop a method for measuring average differences between low- and high-income students in access to effective teachers—known as the effective teaching gap (ETG)—and show how it can be extended to answer further questions beyond the average gap in 1 year. For example, to put the ETG in perspective, for a given distribution of students across schools and a given distribution of teacher value added, we calculate the maximum possible difference in teacher effectiveness between high- and low-income students. This maximum ETG accounts for how high- and low-income students are distributed within and across schools. We also estimate how large the ETG would have to be to cut the student achievement gap in half, and how much the student achievement gap could be reduced by providing equally effective teachers to high- and low-income students over 5 years. In sum, these extensions of the basic method give a better sense of the magnitude of differences in access to effective teachers than just providing a single number. We also supplement our main approach by using a variety of methods to document differences between high- and low-income students.
2. While most of the past studies in this literature have focused on teachers in a single district or single state,[3] we use data on teachers in 26 districts located in 15 states in all four Census regions, using data from five school years (2008–2009 to 2012–2013). Thus, the estimates from our analysis cover a broad range of districts in different geographic areas and operating under different conditions, and the resulting estimates are less likely to be influenced by idiosyncratic conditions in a single large district or state. Another advantage is that the sample allows us to use a common approach to measure access to effective teachers within each district, which we use to assess regional variations in access. In addition, to reconcile

differences in the literature, we supplement our benchmark value-added model with estimates from alternative models and measure the degree to which the results change.

3. Given the policy emphasis on differences in the proportion of novice teachers across low- and high-poverty schools, we examine how these differences contribute to differences in access to effective teachers by decomposing the difference in teacher effectiveness into (a) differences in the likelihood of being taught by a novice teacher and (b) differences in being taught by a more effective teacher, accounting for differences in experience.

Based on our primary model of teacher value added, we find very little evidence of inequity on average or across the distribution of teachers at various levels of effectiveness. To the contrary, the data show nearly equal access to effective teachers within the study districts, so rearranging teachers in a district to obtain perfect equity would do little to narrow the sizable student achievement gap between low- and high-income students. We find that within our sample, there is meaningful variation in access to effective teachers across regions of the United States but also that methods matter—differences in access to effective teachers are somewhat smaller in our primary model in which teacher effectiveness is measured with a value-added model that includes peer effects, and are larger in a model that excludes these peer effects. Finally, the difference in the likelihood that a low-income student is taught by a novice teacher contributes a negligible amount to differences in access to effective teachers for low-income students.[4]

Data and District Context

Characteristics of Study Districts

We purposely selected 30 medium to large districts from across the country to participate in the study. We recruited districts with a mix of low- and high-income students, as we planned to measure differences in teacher effectiveness between these two groups. We also targeted districts with data linking teachers to the students they taught. After obtaining data from 30 districts, we ultimately included 26 districts in our main analysis.[5]

Although we did not use a nationally representative sample of districts, the districts were chosen to be geographically diverse, with at least three districts from each of the four U.S. Census regions. The districts are large—with a median enrollment of approximately 70,000 students—and have high percentages of low-income and minority students (Table 1). In the study districts, 63% of the students are FRL, 29% are Black, and 42% are Hispanic. These characteristics distinguish study districts from the typical district nationally. The median U.S. district has an enrollment of about 1,000 students. Nationally, 44% of students are FRL, 17% are Black, and 22% are Hispanic. Overall, the achievement levels of students in the study districts lag the average achievement levels of other students in their respective states, with performance levels of the average student in our sample at the 45th percentile in ELA and at the 46th percentile in math. The study districts are similar on most measures to the 100 largest U.S. districts, a group that includes many of the study districts. For example, the largest U.S. districts have the same median enrollment as the study districts (approximately 70,000 students) and the percentage of Black and English learner (EL) students differs by no more than five percentage points from the study districts.

Study districts differ from the 100 largest U.S. districts in two main ways. First, study districts are more urban—69% of the students live in large cities compared with 46% in the 100 largest districts—and have more low-income students, with 63% FRL compared with 53% in the 100 largest districts. Second, study districts agreed to participate in the study and could provide the data needed to estimate value-added models. At the time of district recruitment in spring 2011, districts that could provide such data tended to have more sophisticated and well-organized data systems.

The poverty level of schools within study districts is one important factor in determining the potential for inequitable access to effective teachers. If most schools in a district have the same percentage of FRL students, for example, then between-school access to effective teachers will be equitable by definition. On average, there

TABLE 1

Comparison of Study Districts to All U.S. Districts and the 100 Largest

District characteristic	All U.S. districts	100 largest U.S. districts	Study districts
District enrollment (district median)	1,000	70,000	70,000
Percentage of students in large city	14%	46%	69%
Percentage of students eligible for free or reduced-price lunch	44%	53%	63%
Student race and ethnicity (percentages)			
White	55%	30%	23%
Black	17%	27%	29%
Hispanic	22%	34%	42%
Percentage English learner students	9%	14%	19%
Average student achievement (percentiles)			
English/language arts	N/A	N/A	45
Math	N/A	N/A	46
Number of districts	13,406	100	26

Source. 2008 to 2009 Common Core of Data.
Note. District enrollment is based on the size of the median district; the other characteristics are based on student-weighted averages for all districts. District enrollment is rounded to the nearest 10,000 to maintain confidentiality. N/A = not available.

is substantial variation across schools in the percentage of FRL students. Overall, 38% of the teachers teach in low-poverty schools (those with less than 60% of FRL students), with 39% in medium-poverty schools (60%–90%), and 23% in high-poverty schools (more than 90%). There is substantial variation across districts in this distribution. In eight districts, more than 60% of teachers are in low-poverty schools with most of the remainder in medium-poverty schools. In seven districts at the other extreme, less than 20% of teachers are in low-poverty schools.

Data

The districts in our study provided data on their ELA and math teachers in Grades 4 through 8 and their students. We collected data to measure access to effective teachers over a 5-year period: the 2008–2009 through 2012–2013 school years for 21 districts, and the 2007–2008 through 2011–2012 school years for five districts.[6]

We collected students' scores on state assessments in ELA and math for Grades 3 through 8, student demographic characteristics, and teacher/student course links. We converted all state assessment data to z-scores based on the mean and standard deviation of test takers in the state. We excluded upper elementary teachers (Grades 4 and 5) in 14 of the 26 districts because they provided data that linked elementary students to their homeroom teacher rather than to specific ELA and/or math teachers. Given that students in these grades may not receive ELA and math instruction from their homeroom teacher (Isenberg et al., 2015), we could not ensure that the homeroom teacher instructed students in both subjects. In addition, we collected data on teacher experience. For analyses that required data on teachers' experience, we used information on teachers' total teaching experience, not just their experience in the district, which limited these analyses to the 18 districts that could provide data on total teaching experience.

Time Period

The 2008–2009 through 2012–2013 study years were marked by a severe economic recession followed by a gradual recovery, with high unemployment by historical standards. Study districts had an average unemployment rate of 8.2% during this period, which matched the national average (U.S. Bureau of Labor Statistics, 2016). The average unemployment rate in study districts increased sharply from 5.9% in 2008 to 9.7% by 2010—a consequence of the economic downturn—and then decreased to 7.4% in 2013. Amid these poor economic conditions, which existed in tandem with tight state and local government budgets, 19 of the 26 districts laid off teachers during the study years, usually based on seniority (based on study interviews with district staff). Although layoffs during the study period increased involuntary attrition, high levels of unemployment may have discouraged teachers from leaving their positions voluntarily to seek another teaching position or a job in another field. Overall, however, it is not clear how this context may have influenced low-income students' access to effective teachers, particularly because it is not known whether more effective or less effective teachers were more strongly affected by these trends.

Student Achievement Gaps

Student achievement gaps by family income in the study districts mirror those at the national level. Among eighth-grade students in study districts, the student achievement gap in ELA is 0.68 standard deviations of student achievement. This is equivalent to the typical low-income student performing at the 36th percentile on ELA state achievement tests, whereas the typical high-income student is at the 63rd percentile, a gap of 26 percentile points (difference due to rounding). The achievement gap in the study districts in math is 0.63 standard deviations, equivalent to 24 percentile points. The eighth-grade student achievement gap, based on the 2009 National Assessment of Educational Progress (NAEP), is similar, whether measured based on a national sample or a select group of large U.S. cities. The achievement gap in eighth grade is 27 percentile points in ELA and 28 percentile points in math for all U.S. districts, and 26 percentile points in ELA and 28 percentile points in math in selected large city districts in the United States. In fourth grade, the student achievement gap is slightly larger than that in the eighth grade, at 28 percentile points in ELA and 29 percentile points in math in study districts, compared with 28 percentile points in ELA and 30 percentile points in math in the national sample.

Variation in Teachers' Effectiveness

In general, the greater the variation in teachers' effectiveness, the greater the potential for inequitable access—if all teachers were equally effective, there could not be any difference between the average effectiveness of teachers of low- and high-income students. On average, in our districts, the standard deviation of teacher effects was 0.13 in ELA and 0.20 in math. This suggests substantial variation in teacher effectiveness that is consistent with the existing research on value added (Kane et al., 2008; Koedel & Betts, 2011).

Analytic Strategy

To determine whether low-income students are taught by less effective teachers than high-income students, we calculated the ETG in each district by first estimating teacher value added and then subtracting the average value added of teachers of low-income students from the average value added of teachers of high-income students. We defined FRL students as low income; all other students were defined as high income.

Estimating Teacher Value Added

Our value-added model measures the effectiveness of ELA and math teachers in Grades 4 to 8 in the study districts. We estimated separate models for each combination of district, grade, subject, and year. Because we are interested in measuring the ETG, the value-added model includes teacher fixed effects rather than teacher random effects. This strategy avoids the assumption of a random effects model that there is no correlation between student characteristics and teacher assignments. The extent to which student characteristics and teacher assignments are correlated is, after all, the ultimate research question we are investigating.[7] Conceptually, we estimated the following model:

$$Y_{ij} = \lambda' P_i + \beta' X_i + \gamma' C_{ij} + \theta' T_{ij} + \varepsilon_{ij}. \quad (1)$$

In this equation, Y_{ij} is the post-test score for student i of teacher j, and P_i represents test scores for that student in ELA and math in the prior year. The pretest scores capture prior inputs into student achievement. As a measure of true student ability, standardized tests contain measurement error, causing standard regression techniques to produce potentially biased estimates of effective teaching. This occurs because unadjusted coefficients on pretest scores are likely to be attenuated due to measurement error, which would cause the model to attribute too little of a student achievement to past inputs and thereby give too much credit or blame to the current teacher for a student's performance this year. To address this issue, we implemented a measurement error correction that uses the test/retest reliability of the tests used in our value-added models (Buonaccorsi, 2010). By netting out the known amount of measurement error, the errors-in-variables correction eliminates this source of bias.

Additional control variables for individual student background characteristics were included in X_{it}, while C_{ij} represents the peer effects variables. We controlled for the following student-level demographic characteristics: FRL status; limited English proficiency; special education status; gender; whether a student is Black; whether a student is Hispanic, Native American, multirace, or "other" race; and whether a student transferred across schools during the year. We also controlled for three peer effects: classroom average same-subject test scores from the prior school year, the standard deviation of the scores within a student's classroom, and the proportion of FRL students in the classroom. T_{ij} represents a set of binary indicator variables (one for each teacher in the sample) that indicate whether student i appeared on the roster of teacher j during year t.[8] Finally, ε_{ij} is an error term. The key parameters are those included in the vector θ, which are the value-added regression coefficients for individual teachers.

We include peer effects—most importantly, the average prior achievement of other students in the classroom—for four reasons. First, because the purpose of the control variables is to account for the possibility of student sorting into the classrooms based on their characteristics, we follow the general principle of controlling for these characteristics to as great an extent as possible (Koedel et al., 2015). Second, estimates of the validity of value-added models that include peer effects suggest little to no bias in these models (Bacher-Hicks et al., 2014; Chetty, Friedman, & Rockoff, 2014b; Rothstein, 2015), while one study found nontrivial bias in school effectiveness estimates in a value-added model that excluded peer effects (Angrist et al., 2017). Third, the one study that examined bias in various value-added model specifications found less bias in a model that included peer effects than in one that excluded these variables, although the difference was small (Chetty, Friedman, & Rockoff, 2014b). Fourth, a study that used simulated data found that a model with peer effects outperformed a model without peer effects in terms of correctly rank ordering teachers as long as there were multiple years of data available (Thompson & Guarino, 2017). That said, using within-teacher variation in classroom characteristics to estimate peer effects that are then applied to between-teacher differences could lead us to understate or overstate the influence of these characteristics. Therefore, we test the sensitivity of the results to the inclusion of peer effects.

We estimated this model in three steps because (a) we used multiple years of data to estimate the impact of peer effects on student achievement but were interested in teacher value added from each year separately and (b) we required an extra step to obtain standard errors that are robust to clustering. This process allowed us to account for peer effects in our value-added estimates by using cross-year, within-teacher variation in classrooms, while also allowing us to estimate separate effects for each teacher-year-grade combination. Full details of the multistep estimation strategy are given in the Appendix, available in the online version of this article.

Given that we gain precision when measuring the ETG by averaging together estimates for many teachers (as explained in the next section), we focus on measuring value added with minimal bias. For this reason, we neither combine value-added estimates for teachers across years nor apply empirical Bayes shrinkage. The first adjustment would bias the year-specific estimates, which is especially important for teachers in the first few years of their career. The second adjustment, by minimizing the mean squared error, would introduce a source of bias into the results.

However, for measuring differences in the proportion of low- and high-income students who have teachers at different percentiles of the distribution of value-added estimates, we applied empirical Bayes shrinkage as a final step using a procedure outlined in Morris (1983). This ensured that teachers of low-income students, whose value-added estimates tend to be less precisely estimated, will not be more likely to receive estimates that are in the tails of the distribution by chance (Herrmann et al., 2016).

Measuring ETGs

After generating a value-added estimate for each teacher, we linked each student to his or her teacher's value-added estimate. We then calculated a district's ETG as the average value added of the teachers of high-income (non-FRL) students minus the average value added of teachers of low-income (FRL) students. Teachers who have both types of students in their classrooms counted toward both averages in proportion to the number of FRL and non-FRL students they taught. We computed the district ETG using a simple regression:

$$V_{jk} = \alpha + \delta FRL_{jk} + e_{jk}, \qquad (2)$$

where V_{jk} is the value added of teacher j for a particular group of students (k). Each teacher contributed two observations for a given subject: one for FRL students ($k = 0$) and one for non-FRL students ($k = 1$). We regressed V_{jk} on FRL_{jk}, a binary variable that takes a value of one for a teacher's non-FRL students and zero for a teacher's FRL students. That is, each teacher had two observations, with $V_{j1} = V_{j0}$, $FRL_{j0} = 1$, and $FRL_{j1} = 0$. We "reverse coded" FRL status so that a positive ETG would indicate an ETG favoring non-FRL students. Each observation was weighted according to the total number of students of that type, where students who spent less than a full year with the teacher contribute an amount equal to the proportion of the year they were taught by that teacher (their dosage). The estimated coefficient δ measures the estimated mean difference in effective teaching between non-FRL and FRL students in the district, with a positive δ indicating that non-FRL students have more effective teachers on average. To compute an appropriate standard error that accounts for using two observations per teacher, we estimated the regression using cluster-robust standard errors at the teacher level (Arellano, 1987; Liang & Zeger, 1986). For most analyses, we report statistics using δ, but, when displaying district-level ETG results for a group of districts, we reduce the risk that districts with relatively few teachers and students will receive a very high or very low ETGs by chance by applying an empirical Bayes shrinkage procedure to the ETG estimates (Morris, 1983).

Measuring the Cumulative ETG

We also calculated how the cumulative effect of the single-year ETGs translate into changes in the student achievement gap over multiple years. To measure this, we estimated how student achievement gaps would change if high- and low-income students had equally effective teachers between Grades 4 and 8 (or between Grades 6 and 8). We measure value added beginning in fourth grade (or sixth grade) because that is the first grade level where we have the data needed to measure teacher value added. Although all districts had student achievement data for Grades 3 through 8, some districts provided data linking students, teachers, and subjects beginning in fourth grade and other districts beginning in sixth grade. For districts where low-income students already have more effective teachers, we assumed that the current distribution of teachers would not change. So, this analysis describes how student achievement gaps would change if low-income students had *at least* equally effective teachers for multiple years.

In estimating the multiyear effect, intuitively, we expressed the student achievement gap in eighth grade as (a) the contribution of the student achievement gap from the end of seventh grade; (b) the contribution of family or other environmental factors during eighth grade; and (c) the contribution of eighth-grade teachers. Mathematically, this can be written as an identity equal to the sum of (a) the student achievement gap at the end of seventh grade multiplied by a fade-out factor in the persistence of achievement scores from 1 year to the next; (b) the difference in average student characteristics (between high- and low-income students) multiplied by the contribution of student characteristics in eighth grade to student achievement; and (c) the eighth-grade ETG. The student achievement gap at the end of seventh grade favors high-income students. The fade-out factor in the first part of the equation serves to decrease the student achievement gap. However, high-income students have student characteristics that are correlated with higher achievement, so the second part of the equation counteracts this, tending to increase the student achievement gap from seventh grade to eighth grade. A positive ETG will further exacerbate the student achievement gap, whereas a negative ETG will decrease the student achievement gap. We then replaced the seventh-grade student achievement gap by an equation that expresses it as a function of the sixth-grade student achievement gap, the difference in student characteristics in sixth grade, and the sixth-grade ETG. We worked backward recursively until we expressed the eighth-grade student achievement gap as a function of the third-grade student achievement gap, student characteristics in all grades, a set of relationships for each grade between characteristics and achievement, and the ETG for each year. We obtained the parameters for these equations from the value-added model for each grade.

In this framework, by construction teacher effectiveness fades out over time, which is the key to understanding why a 5-year ETG is not simply 5 times the 1-year ETG. The extent of the fade-out—and resulting calculation of the cumulative ETG—depends on the model specification. We estimate the extent of this fade-out using estimates from the value-added model on how students' test scores from the prior year are related to their test scores in the current year.[9] Because this equation is an identity, using student data and parameters from the value-added model, it precisely generates the student achievement gap in eighth grade.

By artificially setting the ETG in each grade to zero, we used the equation to derive a hypothetical eighth-grade student achievement gap. This is the student achievement gap that would have been obtained if we started with the same student achievement gap at the end of third grade, assumed the same students attended from Grades 4 to 8, and assumed the same relationships between student characteristics and achievement each year, but supposed there were no differences in teacher assignment based on student income (that is, no ETG) from Grades 4 to 8.[10] Further methodological details on the ETG and its extensions are given in the Online Appendix.

Measuring the Effect of Disproportionality of Exposure to Novice Teachers

Due to policymakers' concerns about the disproportionate placement of novice teachers in high-poverty schools—ESSA requires states to document whether there are disparities in access to not only ineffective teachers but also specifically inexperienced teachers—we investigated the extent to which disproportionality in rates of placement with novice teachers could lead to greater inequity for low-income students. We defined novice teachers as those in their first 3 years of teaching; teachers with at least 3 years of experience are veteran teachers.

First, we documented the proportion of novices at high-poverty schools, defined as schools with at least 90% FRL students, and the proportion of novices at low-poverty schools, defined as schools with no more than 60% FRL students. We also examined the average difference in value added between novice and veteran teachers.

Second, we formally decomposed the ETG into (a) differences in the likelihood of being taught by a novice teacher and (b) differences in teacher effectiveness, accounting for compositional differences in the likelihood of being taught by a novice.[11] This decomposition, in turn, depends on four factors: (a) the proportion of high-income students taught by novices; (b) the percentage of low-income students taught by novices; (c) the average value added of novices for low-income students; and (d) the average value added of veteran teachers for low-income students. This is shown in Equation (3):

$$\begin{aligned} ETG &= \left[\left(P_{nov}^{LI} - P_{nov}^{HI}\right)\left(\overline{VA}_{vet}^{LI} - \overline{VA}_{nov}^{LI}\right)\right] \\ &\quad + \left[\begin{array}{l} \left(P_{nov}^{HI}\right)\left(\overline{VA}_{nov}^{HI} - \overline{VA}_{nov}^{LI}\right) \\ +\left(P_{vet}^{HI}\right)\left(\overline{VA}_{vet}^{HI} - \overline{VA}_{vet}^{LI}\right) \end{array}\right] \\ &= \textit{Difference in Likelihood of} \\ &\quad \textit{Being Taught by a Novice Teacher} \\ &\quad + \textit{Difference in Teacher} \\ &\quad \textit{Effectiveness, Accounting} \\ &\quad \textit{for Composition.} \end{aligned} \tag{3}$$

In Equation (3), P represents the probability of a low-income (LI) or high-income (HI) student having a novice (nov) or veteran (vet) teacher, and $\overline{VA}$ represents the average value added of a particular group of teachers.

Results

Average ETG

On average across study districts, high-income students have more effective teachers than low-income students, but the differences are small (Table 2). In ELA, the ETG is 0.005 standard deviations of student achievement and in math, the ETG is 0.004. Both are statistically significant.

As another way of describing the degree of inequity in access to effective teachers, we compared where the average teacher of high- and low-income students falls in the overall distribution of teacher effectiveness. In both subjects, the average teacher of a low-income student is just below the 50th percentile, while the average teacher of a high-income student is at the 51st percentile. We also compared the ETG in the average study district with a scenario in which high-income students have the most effective teachers and low-income students have the least effective teachers (the maximum ETG). To measure the maximum ETG in each district, we assumed that the group of students to which each teacher was assigned stayed intact but teachers were reassigned both within and between schools in a way that most benefited high-income students. This measure accounts for the extent of student separation across schools and classrooms by the income level. In districts with no separation by income (all classes and all schools have the same percentage of FRL students), there

TABLE 2

Effective Teaching Gap (ETG)

Subject	Low- vs. high-income	Black vs. White, non-Hispanic	Hispanic vs. White, non-Hispanic	English learners vs. non-English learners
English/language arts	0.005*	0.002	0.003	−0.006
	(0.001)	(0.002)	(0.002)	(0.003)
Math	0.004*	0.010*	0.005	−0.003
	(0.002)	(0.002)	(0.003)	(0.004)
Total number of districts	26	13	16	9

Note. District-level results are weighted across grades and years by the number of students. Overall results are weighted equally across districts. A positive ETG would favor high-income students (Column 1), White, non-Hispanic students (Columns 2 and 3), and non-English learners (Column 4).
*Differences in the value added of the teachers of high- and low-income students (the ETGs) are statistically significant at the 0.05 level, two-tailed test.

would be no possibility of an inequitable distribution of teachers so the maximum ETG would be 0. This is not the case in our sample of districts, where there is some segregation by income and the actual ETG is substantially less than it could be—much lower than the maximum potential gaps of 0.20 in ELA and 0.25 in math.[12]

Access to Effective Teachers by Race, Ethnicity, or EL Status

We measured access to effective teachers for Black, Hispanic, and EL students using the same approach we used to measure low-income students' access to effective teachers. To avoid including districts that enroll just a few students of a given race or ethnicity, we limited this analysis to districts where at least 15% of the students are Black or Hispanic and at least 15% of students are White. Similarly, we included districts where at least 15% of students are ELs. On average, Black students have math teachers who are less effective than those who teach White students, but this difference—0.01—is small. In ELA, Black and White students have teachers who are similarly effective, as the difference between the two groups is not statistically significant. In both subjects, there are no significant differences between teachers of Hispanic and White students, or between teachers of EL and non-EL students (Table 2).

Distribution of Teacher Effectiveness

Although the small ETGs suggest that there are small differences in the effectiveness of teachers of high- and low-income students on average, it is possible that pockets of inequity in access to effective teachers exist within the average study district.[13] To explore this possibility, we examined the likelihood that high- and low-income students are taught by teachers across the distribution of effectiveness.

In study districts, there are small differences or no differences between high- and low-income students in the probability of having one of the most effective teachers or one of the least effective teachers in the district. In both subjects, 10% of high- and low-income students have one of the most effective teachers, on average. In ELA, 10% of low-income students have one of the least effective teachers compared with 9% of high-income students (Figures 1 and 2). In math, among both groups of students, 10% have one of the least effective teachers. Thus, the small difference in the average effectiveness of high- and low-income students' teachers as measured by the ETG does not appear to be concealing larger differences in students' chances of having the most effective or least effective teachers in the district. In either case, the results indicate fairly equitable access to effective teachers in most study districts.

These figures also show small differences in the overall distribution of teachers for high- and low-income students on average. While the most effective teachers boost student achievement substantially relative to the least effective teachers, high-income students are not consistently taught by more effective teachers than low-income students. Instead, both high- and

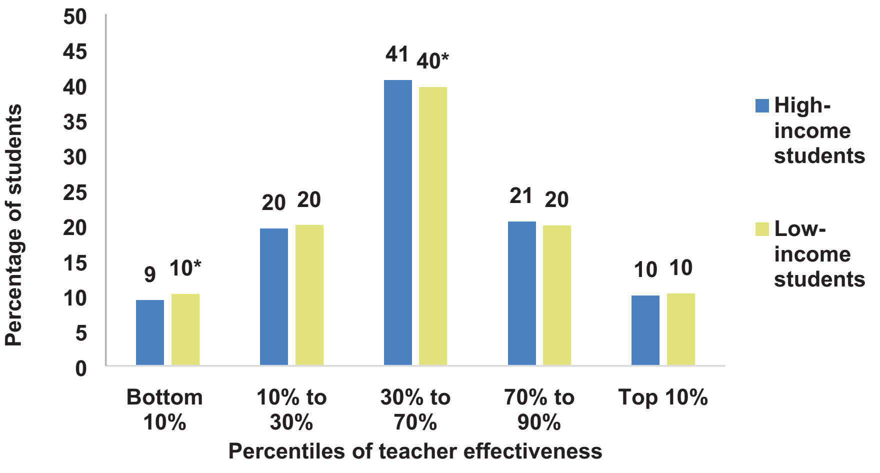

FIGURE 1. *Percentage of low- and high-income students taught by teachers at different levels of effectiveness, English/language arts.*
Source. Authors' calculations based on district administrative data.
Note. Results are based on 26 districts for Years 1 to 5, including Grades 4 to 8 for 12 districts and Grades 6 to 8 for 14 districts. District-level results are weighted across grades and years by the number of students. Overall results are weighted equally across districts.
*Differences in the percentage of low- and high-income students are statistically significant at the 0.05 level, two-tailed test.

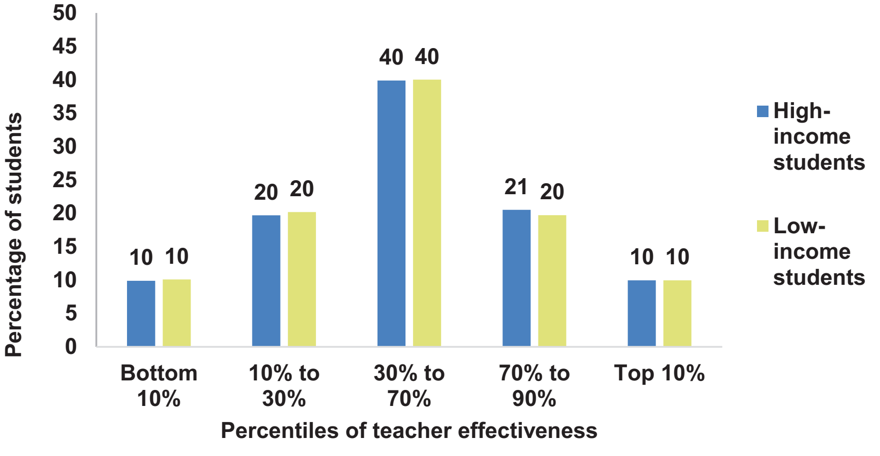

FIGURE 2. *Percentage of low- and high-income students taught by teachers at different levels of effectiveness, math.*
Source. Author's calculations based on district administrative data.
Note. Results are based on 26 districts for Years 1 to 5, including Grades 4 to 8 for 12 districts and Grades 6 to 8 for 14 districts. District-level results are weighted across grades and years by the number of students. Overall results are weighted equally across districts.

low-income students are taught by a mix of more effective and less effective teachers.

Teacher Effectiveness Across Schools

Another way to measure access to effective teachers is to compare the effectiveness of the average teacher across schools with different poverty levels. In study districts, however, the average teacher is similarly effective, no matter the poverty level of the school. When we grouped schools into 10 categories based on the proportion of students in the school who are low income, we found relatively small differences across the categories (Table 3). Average value added ranged from 0.02 to −0.01 across the school poverty

TABLE 3

Average Teacher Value Added, by Poverty Status of Schools

	School poverty rate									
Subject	0%–10%	10%–20%	20%–30%	30%–40%	40%–50%	50%–60%	60%–70%	70%–80%	80%–90%	90%–100%
ELA	0.02	0.00	−0.01	−0.01	−0.01	0.00	0.01	0.00	0.00	0.00
Math	0.03	0.02	−0.01	−0.01	−0.02	−0.01	−0.01	0.01	0.00	0.01

Source. Author's calculations based on district administrative data.
Note. Results are based on 26 districts for Years 1 to 5, including Grades 4 to 8 for 12 districts and Grades 6 to 8 for 14 districts. District-level results are weighted across grades and years by the number of students. Overall results are weighted equally across districts. ELA = English/language arts.

categories for ELA and ranged from 0.03 to −0.02 for math. In addition, there was no pattern of average value added decreasing as school poverty rates increased. Teachers in the lowest poverty schools—0% to 10% in ELA and 0% to 20% in math—have the highest average value added, at 0.02 to 0.03. However, less than 2% of students in study districts attend schools in the 0% to 10% range and 6% attend schools in the 0% to 20% range.

Relationship Between Access to Effective Teachers and the Student Achievement Gap

If low-income students had teachers at least as effective as those of high-income students from fourth through eighth grade, this would have relatively little effect on the student achievement gap. We separately examined results for the 12 districts where we could measure how the ETG accumulates from fourth to eighth grade and the full set of 26 districts where we could measure how it accumulates from sixth to eighth grade.

Among the 12 districts with high-quality teacher–student links from Grades 4 to 8, the typical high-income student has achievement at the 60.5 percentile in ELA and the typical low-income student is at the 35.4 percentile—a student achievement gap of 25.1 percentile points. The gap in math is 24.5 points. Assuming low-income students had teachers at least as effective as those of high-income students over 5 years would reduce the student achievement gap in eighth grade in the average study district from 25.1 to 24.2 percentile points in ELA and from 24.5 to 22.3 percentile points in math. Based on the larger sample of 26 districts, providing low-income students with teachers at least as effective as those of high-income students over 3 years from sixth through eighth grade would reduce the student achievement gap in eighth grade by one percentile point or less in both subjects.

What if low-income students had *more effective* teachers than high-income students? We calculated the ETG that would be needed to cut the student achievement gap in half if implemented from fourth through eighth grade (based on 12 districts). In ELA, the ETG would have to be −0.102 (instead of 0.005 in these 12 districts) to make this amount of progress in reducing the student achievement gap. (A negative ETG means that low-income students have more effective teachers than high-income students). In math, the ETG would need to be −0.080 (instead of 0.004 in these 12 districts) to cut the student achievement gap in half. Given the current placement of teachers, achieving these targets would require a substantial change. In ELA, 30% of teachers would have to switch places with each other to reach an ETG of −0.102, assuming that it were possible for the best teachers in classrooms with mostly high-income students to switch places with the worst teachers in classrooms with mostly low-income students. In math, 11% of teachers would have to switch places to obtain an ETG of −0.080.

Variation Across Study Districts

Even though there is relatively little inequity in students' access to effective teachers on

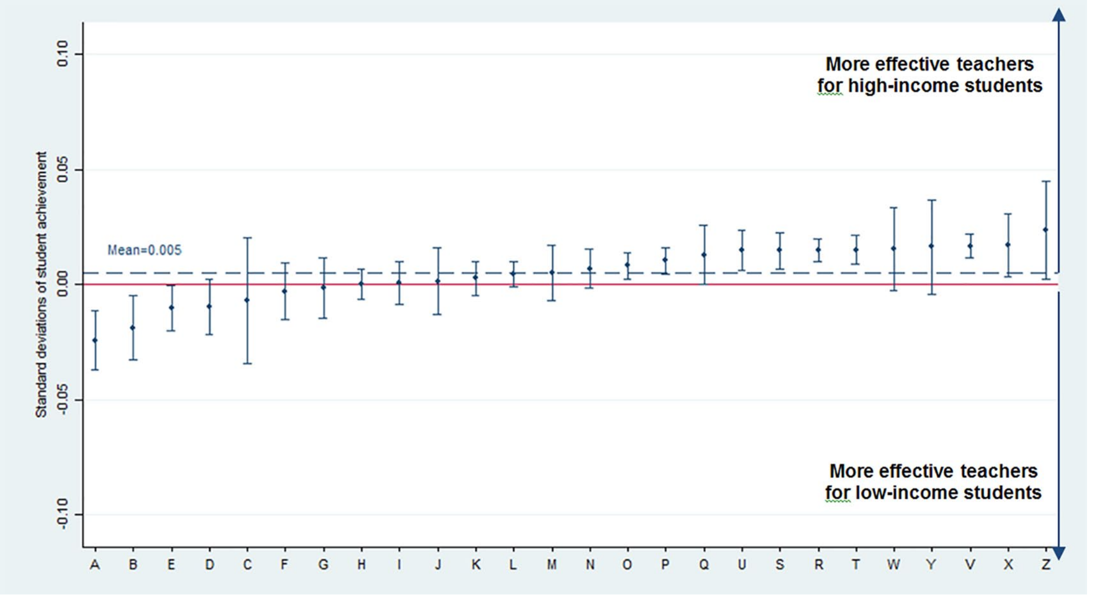

FIGURE 3. *Average ETG in ELA, by district.*
Note. Results are based on 26 districts for Years 1 to 5, including Grades 4 to 8 for 12 districts and Grades 6 to 8 for 14 districts. District identifiers A to Z are assigned according to the size of each district's ETG in ELA (with Z representing the largest positive gap). ETGs are computed within each district-grade-year combination and averaged with equal weight across years within each district. The points represent the district-level ETGs and the vertical lines show the 95% confidence intervals around each point. To reduce the risk that districts, particularly those with relatively few teachers and students, will receive a very high or very low ETGs by chance, we applied an empirical Bayes shrinkage procedure to the estimates. The cross-district average of 0.005 standard deviations (calculated after empirical Bayes shrinkage) is shown by the dashed horizontal line. ETG = effective teaching gap; ELA = English/language arts.

average, there may be individual districts with more inequity in access to effective teachers. In fact, there is modest variation across the study districts in the extent to which low-income students have equal access to effective teachers. The ETG in the 26 study districts ranges from −0.024 to 0.023 in ELA and from −0.050 to 0.040 in math (Figures 3 and 4). This suggests that there are some districts in which low-income students have less effective teachers than high-income students, on average, and other districts in which the opposite is true.[14]

We also see larger inequity in a few districts when we examine the likelihood that students have the most or least effective math teachers. In the three districts in math with the largest inequity, for example, an average of 11% of low-income students have one of the least effective teachers in the district, compared with 8% of high-income students.

Given that low-income students' access to effective teachers varies from district to district, it raises the question of whether certain types of district characteristics are associated with greater inequity in access to effective teachers. To address this question, we examined the relationship between a district's characteristics and the size of its ETG.

There are just two characteristics—district size and region—significantly related to the ETG in both math and ELA (Table 4). Districts that are larger and located in the southern United States tend to have a less equitable distribution of teachers than other districts. These findings are related, as districts in the south tend to be larger than those in other regions. Low-income students' access to effective teachers is not consistently related to the other district characteristics that we examined, including the student achievement gap, the extent to which high- and low-income students are separated across schools, or the percentage of Black, Hispanic, and White students in the district. In ELA, the ETG is significantly larger in districts with a larger percentage of low-income students and those with a larger percentage of minority students, but these relationships are not significant in math.

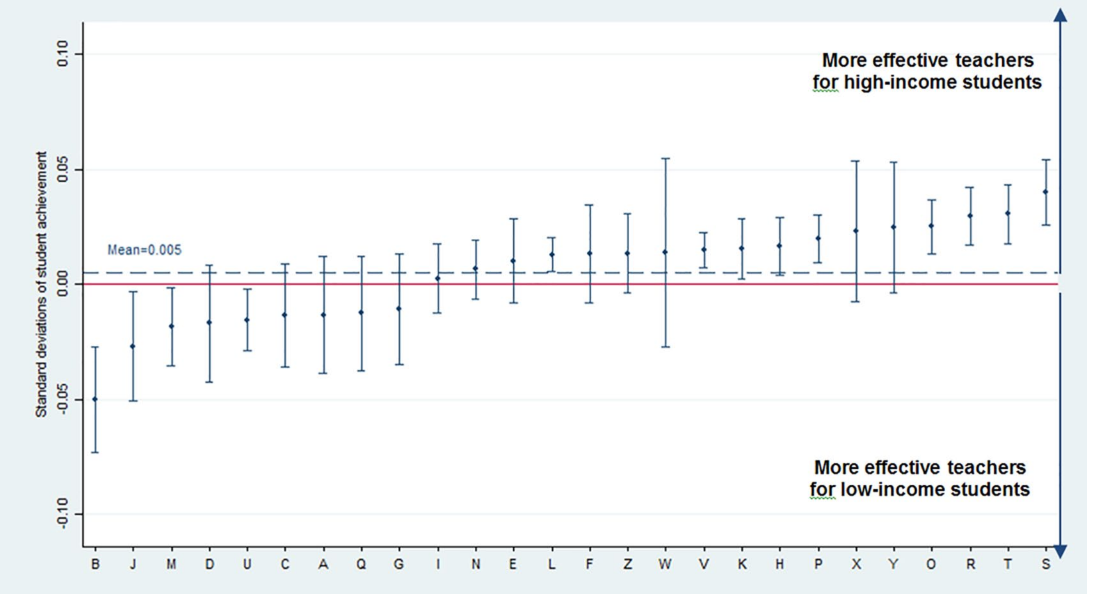

FIGURE 4. *Average ETG in math, by district.*
Note. Results are based on 26 districts for Years 1 to 5, including Grades 4 to 8 for 12 districts and Grades 6 to 8 for 14 districts. District identifiers A to Z are assigned according to the size of each district's ETG in ELA (with Z representing the largest positive gap). ETGs are computed within each district-grade-year combination and averaged with equal weight across years within each district. The points represent the district-level ETGs and the vertical lines show the 95% confidence intervals around each point. To reduce the risk that districts, particularly those with relatively few teachers and students, will receive a very high or very low ETGs by chance, we applied an empirical Bayes shrinkage procedure to the estimates. The cross-district average of 0.005 standard deviations (calculated after empirical Bayes shrinkage) is shown by the dashed horizontal line. ETG = effective teaching gap; ELA = English/language arts.

TABLE 4
Average ETG by District Size and Region

	ETG		
District characteristic	English/language arts	Math	Number of districts
All districts	0.005*	0.004*	26
District size			
Medium districts	0.000	−0.010*	7
Large districts	0.004	0.010*	14
Very large districts	0.014*	0.010	5
Region			
Midwest	−0.003*	−0.014*	6
North	0.007	−0.008*	3
South	0.009*	0.014*	11
West	0.003	0.013	6

Note. Estimates in the table represent the mean ETGs for districts within each group. Results are based on 26 districts, Grades 4 to 8, and Years 1 to 3. Small districts have fewer than 40,000 students, medium districts have 40,000 to 100,000 students, and large districts have more than 100,000 students. Geographic region is based on Census region. ETG = effective teaching gap.
*Whether the ETG in a given category is significantly different from all other districts combined at the 0.05 level, using the standard error of each district's estimate.

TABLE 5

Novice and Veteran Components of the Effective Teaching Gap

Component	Low-income students	High-income students
Percentage of students taught		
By veteran teachers	86.0%	89.8%
By novice teachers	14.0%	10.2%
Average value added		
Veteran teachers	0.000	0.010
Novice teachers	−0.021	−0.023

Note. Results are for teachers in Grades 6 through 8 in 18 districts and in Grades 4 and 5 in 10 of these districts. The sample excludes teachers with missing experience data and five districts that could not provide data on teachers' total teaching experience. Novice teachers have 0 to 2 years of prior experience and veteran teachers have 3 or more years of prior experience. The results are presented as an average across districts weighted by the number of students taught by each teacher in the analysis.

Comparing Results Across Value-Added Models

Our main results for the ETG are based on a value-added model of teacher effectiveness that accounts for a set of three peer effects. However, data limitations sometimes lead researchers to estimate a value-added model that does not account for peer effects. To gauge how sensitive the results are to this feature of the model, we estimated an alternative value-added model without peer effects. However, there is a greater risk using this model that failing to account for peer effects could lead to bias that inflates the estimate of the ETG.

When we measure teacher effectiveness without accounting for peer effects, we find results that are not as small as those in our main model: The typical low-income student has a teacher whose value added is 0.029 lower in ELA than that of the typical high-income student. In math, the ETG is 0.031. In both ELA and math, 11% of low-income students and 7% to 8% of high-income students are taught by one of the least effective teachers in a district. High-income students are 3 percentage points more likely to be taught by one of the most effective teachers in ELA and 1 percentage point more likely in math. Finally, providing high- and low-income students with equally effective teachers over the 5 years between Grades 4 and 8 would reduce the student achievement gap by 3.4 to 3.8 percentile points in ELA and math.

Novice Teachers and the ETG

As is true of many districts, a higher proportion of novice teachers in high-poverty schools suggests that there could be inequitable access to effective teachers. Across the study districts, 18.3% of the teachers in high-poverty schools (90% or more low-income students) are novices, compared with 8.9% of the teachers in low-poverty schools (60% or fewer low-income students). In addition, novices in the study districts are less effective than veteran teachers, by 0.022 in average teacher value added. However, we find that the presence of more novice teachers in high-poverty schools does not create substantial inequity, for two reasons.

First, the substantial difference between high- and low-poverty schools in the prevalence of novice teachers translates into a smaller difference between high- and low-income students in the likelihood of having a novice teacher. Although there are more low-income students in high-poverty schools than in low-poverty schools, both types of students attend each type of school. When calculated at the student level, the difference between the likelihood of being taught by a novice teacher is modest, with 14% of low-income students and 10% of high-income students taught by novices (Table 5). In other words, 86% of low-income students and 90% of high-income students are taught by veteran teachers.

Second, the average difference in the effectiveness of novices and veteran teachers is also modest, at 0.022 across both subjects. Thus, even if all low-income students were taught by novices and all high-income students were taught by veteran teachers, the ETG would be 0.022. The actual difference in the proportion of students taught by a novice teacher is only 4 percentage points. So the component of the ETG resulting from low-income students being taught more frequently by novice teachers is approximately 4% of 0.022, which is slightly less than 0.001. To calculate this precisely, we inserted the values from Table 5 into Equation 3. This produces ([0.140 − 0.102][0.000 − (−0.021)]) = 0.001. That is, the ETG between high- and low-income students resulting from the greater likelihood that low-income students are taught by a novice teacher is 0.001.[15]

Conclusion

Our results show that low-income students have equal or nearly equal access to effective teachers in the great majority of the districts we analyzed. While individual teachers differ substantially in their effectiveness, both high- and low-income students have a mix of the most effective and the least effective teachers. As a result, providing the two groups of students with equally effective teachers—even over a period of 5 years—would not substantially reduce the student achievement gap in most districts. Similarly, the disproportionate number of novice teachers at high-poverty schools contributes almost nothing to the ETG, and, by extension, to the student achievement gap.

Results comparing students in different racial and ethnic groups are similar. Black students have slightly less effective teachers than White students, but there were no statistically significant differences in the effectiveness of their ELA teachers. Similarly, teachers' average effectiveness did not differ significantly when we compared Hispanic students with White students or ELs with non-ELs.

Comparison With Other Studies

Although these results are in line with some of the past literature on access to effective teachers, our findings suggest more equity than those of several other studies. Much of the difference across studies can be explained by four factors: the value-added model used, the grades examined, the scope of the analysis (whether teachers are compared with others in a district or in a state), and the states or districts included in the analysis.

First, as we have documented, the choice of value-added model matters for this analysis. Our primary value-added model aims to eliminate bias from unobservable student characteristics that may be responsible for the way that students are matched to teachers to the extent possible. Otherwise, apparent differences between the effectiveness of teachers of low- and high-income students may be due to a misspecified value-added model rather than to real differences in teacher effectiveness. Specifically, the model not only controls for prior test scores and a set of student characteristics (like special education and EL status) but also (a) accounts for measurement error in pretest scores using an errors-in-variables correction and (b) controls for a set of three peer effects (principally, the average classroom pretest score) by leveraging variation across classes of a given teacher. When we compared our overall results with results that used a value-added model without peer effects, our estimates of the ETGs were higher, at 0.024 in ELA and 0.027 in math. Thus, our results suggest that the inclusion or exclusion of peer effects in the value-added model has a modest effect on the estimates.

Results from the literature support the notion that the specification of the value-added model matters. Two studies that account for peer effects when measuring teacher effectiveness find only small differences in the effectiveness of teachers of high- and low-income students, or that low-income students have more effective teachers than high-income students (Chetty, Friedman, & Rockoff, 2014b; Steele et al., 2014). One exception is Sass et al. (2012), who find that teachers in Grades 4 and 5 in high-poverty schools are less effective than those in low-poverty schools in Florida and North Carolina, with average differences ranging from 0.02 to 0.04. Goldhaber et al. (2016a), who use a value-added model that does not account for peer effects or measurement error in pretest scores, find ETGs for Grades 4 and 5 that range from about 0.025 to 0.035. Steele et al. (2015), whose model also excludes these features, appear to find somewhat larger differences (although Steele et al. [2015] account for up to 12 pretest scores to try to minimize measurement error).[16]

Second, there is evidence that (a) the ETG is higher in upper elementary grades than in middle school grades and (b) a value-added model that accounts for peer effects reduces the ETG in middle school grades but may increase the ETG in upper elementary school grades (Goldhaber et al., 2016b). This may help to explain why Sass et al. (2012) and Goldhaber et al. (2016a) find slightly more inequity than some other papers in the literature, as these papers are based on Grades 4 and 5 only. Our study is based on Grades 4 to 8, but 14 of 26 districts exclude Grades 4 to 5 due to data issues.

Third, Sass et al. (2012) and Goldhaber et al. (2016a) measure not only differences within a district but also differences between districts in a state, which may increase the degree of inequity if states have a greater degree of sorting of highly effective teachers to high-income students than individual districts. Goldhaber et al. (2016b) find that about a third of the measured inequity in North Carolina (i.e., about 0.01) and about half of the measured inequity in Washington state (about 0.015) result from differences between districts.[17] It should be noted, however, that the districts participating in our study were purposefully selected based on having FRL rates between 20% and 80% to ensure a degree of heterogeneity within each district.[18]

Fourth, to the extent that different studies produce different estimates in general, this may also reflect different samples of teachers and students in different districts and states. Within our geographically diverse sample of districts, we find some variation in the degree of access to effective teachers. The districts in our sample with the most inequity have ETGs that are similar to the estimates from other studies. We also document a correlation between higher ETGs and larger districts, and regional variation, with districts in the South having the highest ETGs.

Policy Implications

The findings of our study—based on a cross-section of medium and large districts throughout the United States—suggest that a policy emphasis on correcting for an unequal distribution of "ineffective, out-of-field, or inexperienced teachers" may be misplaced. Across these districts, high- and low-income students are far apart in their achievement by the end of third grade, and the way that students are distributed across schools presents a threat that the student achievement gap could grow considerably by eighth grade if the best teachers were matched to high-income students. But in fact, the achievement gap grows little due to inequitable access to effective teachers. Instead, the value-added estimates show that effective and ineffective teachers teach in all schools. Effective teachers are found in high-poverty schools, even if their accomplishments are often overlooked, as their students—who start out far behind—may still face other obstacles to learning and so have below-average test scores. Conversely, ineffective teachers can be found in "blue ribbon" schools, where subpar performance can be camouflaged by satisfactory outcomes for advantaged students.

It may not be reassuring that public schools are just holding the line on a set of unequal outcomes instead of decreasing them, but public schools are financed and managed within a political system, and our simulation results suggest that it may be difficult to jolt this system in a way that would allow enough of the best teachers to flow to high-poverty schools to bring about a substantial decrease in achievement gaps through teacher mobility alone. This is not to concede that policymakers need to accept the status quo, just that the best policy solutions to diminishing the student achievement gap likely reside outside the realm of focusing on rules by which teachers transfer from school to school or how teachers are hired or retained. Although a well-planned and well-executed set of human capital policies can improve teacher effectiveness overall, it is not likely to diminish the student achievement gap. Rather, our descriptive results might nudge policymakers to consider a broad spectrum of other cost-effective evidence-based policies benefiting disadvantaged students. For example, experimental evidence supports the expansion of tutoring (Nickow et al., 2020) and "No Excuses" charter schools in urban areas (Chabrier et al., 2016; Gleason, 2019). In addition, well-implemented policies that target early learners may disrupt the predictability of student achievement gaps that form as soon as children enter school and stay at a similar level in the early elementary grades (Duncan & Magnuson, 2011). For instance, experimental evidence demonstrates that tutoring has especially strong impacts on literacy for students in early grades (Nickow et al., 2020), as does coaching of teachers (Kraft et al., 2018), and being taught by a teacher from Teach for America (Clark & Isenberg, 2020).

In the study districts and elsewhere in the country, low-income and minority children have inequitable outcomes, but this may have less to do with the educators who staff their school and more to do with other factors, such as differences in the home environment of or resources available to children of different means. In sum, a half-century after the Coleman Report found "that differences between schools account for only a small fraction of differences in pupil achievement" (Coleman et al., 1966), with more sophisticated methods, easier access to data, more computational power, and the ability to take the analysis from the school level to the teacher level, we have concluded much the same thing.

Acknowledgments

We appreciate the valuable efforts of district leaders and central office staff in making this study possible. District staff were critical partners in the study by investing the time and energy in collecting and providing the data for this report. Elizabeth Warner and Lauren Angelo at the Institute of Education Sciences (IES) provided valuable guidance on the study at all stages. This article benefited from the input of the study's technical working group: J.R. Lockwood, Jennifer King Rice, Jonah Rockoff, Andrew Rotherham, Tim Sass, Victoria Van Cleef, and Jacob Vigdor. Steve Glazerman provided expert advice on the study design and analyses, and Brian Gill provided valuable feedback on the analyses. We also thank anonymous reviewers who reviewed and commented on the article at various stages. The report would not have been possible without the valuable contributions of many staff at Mathematica and the American Institutes for Research. A strong team of programmers at Mathematica brought exceptional skill to the preparation and analysis of data for the report. Emma Kopa led the team of programmers, and John Hotchkiss, Kai Filion, Matthew Jacobus, Eric Lundquist, and Raul Torres provided expert programming to implement the analysis for this report. An excellent team of programmers made critical contributions to the preparation and analysis of data for the report: Anna Collins, Anna Comerford, Alena Davidoff-Gore, Dylan Ellis, Nikhil Gahlawat, Serge Lukashanets, Lisa McCusker, Nora Paxton, Elizabeth Potamites, Chelsea Swete, Alma Vigil, and Miles Watkins. The programming team from the American Institutes for Research consisted of Victoria Brady, Alvaro Ballarin Cabrera, Matthew Corritore, and Thomas Gonzalez, with leadership from Zeyu Xu. We also thank Mary Grider who provided valuable guidance to the programming team throughout the study, and Marykate Zukiewicz and Scott Baumgartner who worked closely with programmers to prepare the administrative data for analysis.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was funded by the Institute of Education Sciences (IES) through contract number ED-IES-10-C-0065. It was a collaborative effort that benefited from the contributions of many people.

ORCID iD

Eric Isenberg https://orcid.org/0000-0003-2535-5863

Notes

1. These values are based on an inspection of Figure 18 in Goldhaber et al. (2016a), which does not report the underlying values.

2. For example, studies of charter schools often provide estimates of the impact of attending a charter school for a single year as well as for multiple years. These estimated single-year impacts on math scores (of charter schools located in urban areas and serving disadvantaged students) are as high as 0.20 to 0.40 standard deviations (Abdulkadiroglu et al., 2011; Dobbie & Fryer, 2011; Gleason et al., 2014), so estimates from the literature of the 1-year effective teaching gap (ETG) are small in magnitude by this standard as well.

3. One exception is Glazerman and Max (2011), which examined 10 districts in six states.

4. A companion paper uses teacher personnel data in addition to value-added estimates to explore how hiring of teachers into the district, development of teachers over time, transfer of teachers across schools, and attrition of teachers out of the district are related to the access to effective teachers (Isenberg et al., 2020). We find that although there is greater turnover at schools with more low-income students, there is little difference in the effectiveness of new hires, transfers,

or leavers. As a result, these teacher career transitions do not contribute to large inequities in access to effective teachers for low-income students.

5. We excluded one district that was unable to provide data that reliably distinguished low- from high-income students based on free or reduced-price lunch (FRL) status. We also excluded three districts that could not provide data linking teachers across years, which is necessary for identifying peer effects in the value-added model.

6. Five districts provided data for less than 5 years because they did not have adequate data linking students and teachers: one district had data for 2 years, one district had data for 3 years, and three districts had data for 4 years.

7. Given that we are ultimately interested in comparing the effectiveness of teachers of low-income (FRL) and high-income (non-FRL) students, it may appear problematic that the value-added model controls for FRL status. If less effective teachers are more likely to teach FRL students, one may ask whether this variable would capture this relationship, such that the value-added model would control for the very relationship we are trying to measure. However, the key to the value-added model is the inclusion of teacher fixed effects, which allow us to measure the relationship of student FRL status and student test scores using only differences between FRL and non-FRL students who have the same teacher. The model uses this within-teacher relationship to infer the expected average test scores among teachers who have a larger or a smaller proportion of FRL students. Teachers with students whose actual test scores are higher than expected given their FRL status (and other characteristics) have higher value-added estimates.

8. A student who is taught by multiple teachers during the year—whether through team teaching, supplemental course taking, or transferring across schools—will contribute one observation to the model for each teacher to whom he or she was linked, a technique known as the full roster method (Hock & Isenberg, 2017).

9. The logic underlying this approach is that if a student's previous year teacher caused their test scores to be higher in that year, then the estimated coefficient on prior year test scores in the value-added model should tell us how much the previous year teacher's effect has persisted. This estimate varies across grades and districts, but generally is in the range of 0.7 to 0.8. In other words, our approach implies that 70% to 80% of a teacher's effect persists to the next year.

10. Methodological details on how we calculated the amount by which the student achievement gap would be reduced if the ETG were zero—and we used this result to calculate the annual ETG necessary to cut the student achievement gap in half—are given in the Online Appendix. We also show how we calculated the maximum ETG for a district assuming that teachers could be reassigned across schools and classrooms, conditional on the distribution of teacher value added and the distribution of students across classrooms.

11. A full derivation of this result is shown in the Online Appendix.

12. In a later section, we show that the estimated ETG within a district is not related to the extent of student separation in that district. This provides further evidence that our low ETG estimates cannot be explained by a lack of student separation across schools.

13. We also found that the ETG remained stable over time. Across the 5 years of the study, the ETG in English/language arts (ELA) varied from year to year by 0.01 or less, with no clear trend over time. In math, the ETG varied by less than 0.02 across the 5 years.

14. The variance of unadjusted estimates across districts will generally overestimate the true variance due to sampling error in the estimates of each individual district. However, the distribution of the empirical Bayes (EB) estimates that we present in Figures 3 and 4 will generally underestimate the true variance because it excludes a component that captures the variance of each individual district's EB estimate (Carlin & Louis, 2000). We obtained constrained EB estimates, which correct for this issue, but found that they differed little from the EB estimates.

15. Our estimate of 0.022 for the average difference between novice and veteran teachers may appear to be small relative to estimates in the literature of the return to experience for teachers. Typically studies find gains 3 times as large for teachers as the progress from their first to their fourth year of teaching. We show elsewhere that the return to experience for teachers in the study districts over that interval is 0.068 (Isenberg et al., 2016), similar to other estimates in the research literature (Harris & Sass, 2011). However, in the analysis of novice teachers and the ETG, we are not estimating the return to experience for individual teachers but rather calculating the difference in the average value added of teachers in their first 3 years compared with more experienced teachers. This difference is considerably smaller than the return to experience for two reasons. First, we consider novices to include teachers in their second and third years of teaching, as well as teachers in their first year. Teachers in their second and third years are not likely to have reached their potential. However, for teachers in the study districts (like teachers in other studies), the largest improvement is in the first year of teaching—an average improvement of 0.046. Second, there is a cohort effect in the study districts—teachers entering the profession more recently are more effective at similar levels of experience than teachers who

entered the profession earlier. Moreover, the findings are not sensitive to the difference in effectiveness of novice and veteran teachers. For example, assume that the difference between novices and veterans is 0.088—4 times what we observe in the study districts. Given the proportions of novice teachers for low- and high-income students we found in our study, the contribution to the ETG due to the greater likelihood of low-income students being taught by novices would be 0.004, still a trivial amount.

16. Steele et al. (2015) use a somewhat different metric than the other papers: differences in average value added between schools in the top and bottom quartile based on the proportion of minority students. It is not immediately clear how their results—0.062 in ELA and 0.044 in math—compare with ETGs.

17. These approximations are based on visual inspection of Figure 19 because the authors do not report the underlying values.

18. While our study was designed to focus on the district level (as this is the locus of control for school policy), we examined the issue of whether measuring inequity only within districts may have affected the results by measuring the ETG in the urban core of five county-wide districts and comparing that with the ETG for the district as a whole. The results of this exercise indicated little difference between those measures, but there may be larger differences between districts than between urban and nonurban areas of a single district. See Isenberg et al. (2016) for details.

References

Abdulkadiroglu, A., Angrist, J. D., Dynarski, S. M., Kane, T. J., & Pathak, P. (2011). Accountability and flexibility in public schools: Evidence from Boston's charters and pilots. *The Quarterly Journal of Economics*, *126*(2), 699–748.

Angrist, J. A., Hull, P. D., Pathak, P. A., & Walters, C. R. (2017). Leveraging lotteries for school value-added: Testing and estimation. *The Quarterly Journal of Economics*, *132*(2), 871–919.

Arellano, M. (1987). Computing robust standard errors for within-groups estimators. *Oxford Bulletin of Economics and Statistics*, *49*(4), 431–434.

Bacher-Hicks, A., Kane, T., & Staiger, D. (2014, October). *Validating teacher effect estimates using changes in teacher assignments in Los Angeles* [NBER working paper]. National Bureau of Economic Research.

Boyd, D., Lankford, H., Loeb, S., Rockoff, J., & Wyckoff, J. (2008, June). *The narrowing gap in New York City teacher qualifications and its implications for student achievement in high poverty schools* (NBER Working Paper No. 14021). National Bureau of Economic Research.

Buonaccorsi, J. P. (2010). *Measurement error: Models, methods, and applications*. Chapman and Hall/CRC.

Carlin, B. P., & Louis, T. A. (2000). *Bayes and empirical Bayes methods for data analysis*. Chapman and Hall/CRC.

Chabrier, J., Cohodes, S., & Oreopoulos, P. (2016). What can we learn from charter school lotteries? *Journal of Economic Perspectives*, *30*(3), 57–84.

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

Chetty, R., Friedman, J., & Rockoff, J. (2014b). Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. *American Economic Review*, *104*(9), 2593–2632.

Chetty, R., Hendren, N., Kline, P., & Saez, E. (2014). Where is the land of opportunity? The geography of intergenerational mobility in the United States. *The Quarterly Journal of Economics*, *129*(4), 1553–1623.

Clark, M., & Isenberg, E. (2020). Do teacher for America corps members still improve student achievement? Evidence from a randomized controlled trial of teach for America's scale-up effort. *Education Finance and Policy*, *15*(4), 736–760.

Clotfelter, C., Ladd, H., Vigdor, J., & Wheeler, J. (2007). High poverty schools and the distribution of teachers and principals. *North Carolina Law Review*, *85*, 1345–1379.

Coleman, J., Campbell, E., Hobson, C., McPartland, J., Mood, A., Weinfeld, F., & York, R. (1966). *Equality of educational opportunity*. U.S. Department of Health, Education, and Welfare, Office of Education.

Constantine, J., Player, D., Silva, T., Hallgren, K., Grider, M., & Deke, J. (2009, February). *An evaluation of teachers trained through different routes to certification: Final report* (NCEE 2009-4043). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Dobbie, W., & Fryer, R. G., Jr. (2011). Are high-quality schools enough to increase achievement among the poor? Evidence from the Harlem Children's Zone. *American Economic Journal: Applied Economics*, *3*(3), 158–187.

Duncan, G., & Magnuson, K. (2011). The nature and impact of early achievement skills, attention skills, and behavior problems. In G. Duncan & R. Murnane (Eds.), *Whither opportunity: Rising inequality, schools, and children's life chances* (pp. 47–69). Russell Sage Foundation.

Glazerman, S., & Max, J. (2011). *Do low-income students have equal access to the highest-performing* *teachers?* [NCEE Evaluation Brief]. National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Gleason, P. M. (2019). Searching for policies and practices that make charter schools successful: What can the research tell us?. In M. R. Berends, R. J. Waddington, & J. Schoenig (Eds.), *School choice at the crossroads* (pp. 186–204). Routledge.

Gleason, P. M., Tuttle, C. C., Gill, B., Nichols-Barrer, I., & Teh, B.-R. (2014). Do KIPP schools boost student achievement? *Education Finance and Policy*, *9*(1), 36–58.

Goldhaber, D., Lavery, L., & 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. (2016a, December). *Has it always been this way? Tracing the evolution of teacher quality gaps in U.S. public schools* (CALDER Working Paper No. 171). National Center for Analysis of Longitudinal Data in Education Research.

Goldhaber, D., Quince, V., & Theobald, R. (2016b, December). *Reconciling different estimates of teacher quality based on value added* (CALDER Brief 14). National Center for Analysis of Longitudinal Data in Education Research.

Hanselman, P. (2018). Do school learning opportunities compound or compensate for background inequalities? Evidence from the case of assignment to effective teachers. *Sociology of Education*, *91*(2), 132–158.

Hanushek, E., & Rivkin, S. (2006). Teacher quality. In E. Hanushek & F. Welch (Eds.), *Handbook of the economics of education* (Vol. 1, pp. 1051–1078). North-Holland.

Harris, D., & Sass, T. (2011). Teacher training, teacher quality and student achievement. *Journal of Public Economics*, *95*(7–8), 798–812.

Herrmann, M., Walsh, E., & Isenberg, E. (2016). Shrinkage of value-added estimates and characteristics of students with hard-to-predict achievement levels. *Statistics and Public Policy*, *3*(1), 1–10. https://doi.org/10.1080/2330443X.2016.1182878

Hock, H., & Isenberg, E. (2017). Methods for accounting for co-teaching in value-added models. *Statistics and Public Policy*, *4*(1), 1–11.

Isenberg, E., Max, J., Gleason, P., & Johnson, M. (2020). The relationship between teacher career transitions and access to effective teachers [Unpublished manuscript].

Isenberg, E., Max, J., Gleason, P., Johnson, M., Deutsch, J., & Hansen, M. (2016, October). *Do low-income students have equal access to effective teachers? Evidence from 26 districts* (NCEE 2017-4007). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Isenberg, E., Teh, B.-R., & Walsh, E. (2015). Elementary school data issues for value-added models: Implications for research. *Journal of Research on Educational Effectiveness*, *8*(1), 120–129.

Kane, T., Rockoff, J., & Staiger, D. (2008). What does certification tell us about teacher effectiveness? Evidence from New York City. *Economics of Education Review*, *27*(6), 615–631.

Koedel, C., & Betts, J. (2011). Does student sorting invalidate value-added models of teacher effectiveness? An extended analysis of the Rothstein critique. *Education Finance and Policy*, *6*(1), 18–42.

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

Kraft, M. A., Blazar, D., & Hogan, D. (2018). The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence. *Review of Educational Research*, *88*(4), 547–588.

Liang, K.-Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, *73*(1), 13–22.

Mansfield, R. (2015). Teacher quality and student inequality. *Journal of Labor Economics*, *33*(3), 751–788.

Morris, C. N. (1983). Parametric empirical Bayes inference: Theory and applications. *Journal of American Statistical Association*, *78*(381), 47–55.

Nickow, A., Oreopoulos, P., & Quan, V. (2020, July). *The impressive effects of tutoring on PreK-12 learning: A systematic review and meta-analysis of the experimental evidence* (NBER Working Paper No. 27476). National Bureau of Economic Research.

Reardon, S. (2011). The widening academic achievement gap between rich and poor: New evidence and possible explanations. In G. Duncan & R. Murnane (Eds.), *Whither opportunity: Rising inequality, schools, and children's life chances* (pp. 91–115). Russell Sage Foundation.

Rothstein, J. (2015, October). *Revisiting the impacts of teachers*. University of California, Berkeley.

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

Schultz, L. M. (2014). Inequitable dispersion: Mapping the distribution of highly qualified teachers in St. Louis metropolitan elementary schools. *Education Policy Analysis Archives*, *22*(90), 1–20.

Steele, J. L., Baird, M., Engberg, J., & Hunter, G. P. (2014, March). *Trends in the distribution of teacher effectiveness in the intensive partnerships for effective teaching*. RAND Corporation.

Steele, J. L., Pepper, M. J., Springer, M. G., & Lockwood, J. R. (2015). The distribution and mobility of effective teachers: Evidence from a large, urban school district. *Economics of Education Review*, *48*(5), 86–101.

Thompson, P., & Guarino, C. (2017). *An evaluation of teacher value-added estimation with peer effects* [Working paper]. Oregon State University.

U.S. Bureau of Labor Statistics. (2016). *Civilian unemployment rate*. https://fred.stlouisfed.org/series/UNRATE

U.S. Department of Education, National Center for Education Statistics. (2014). *Education Longitudinal Study of 2002 (ELS: 2002), base year and third follow-up, table 104.91*. Institute of Education Sciences, Digest of Education Statistics.

Authors

ERIC ISENBERG is a senior study director at Westat. His research interests include teacher effectiveness and experimental evaluations of educational interventions.

JEFFREY MAX is a principal researcher at Mathematica. His research focuses on teacher professional development and the use of continuous improvement in education.

PHILIP GLEASON is a senior fellow at Mathematica. His research focuses on school choice, teacher effectiveness, and research methodology.

JONAH DEUTSCH is a senior researcher at Mathematica. His research areas include school and teacher measures of effectiveness, employment and training programs, and program evaluation methods.

Manuscript received November 22, 2019
First revision received October 5, 2020
Second revision received June 3, 2021
Third revision received June 13, 2021
Accepted June 15, 2021