Access to Effective Teaching for Disadvantaged Students

November 2013

NATIONAL CENTER FOR EDUCATION EVALUATION AND REGIONAL ASSISTANCE
Institute of Education Sciences

U.S. Department of Education

This page has been left blank for double-sided copying.

Access to Effective Teaching for Disadvantaged Students

November 2013

Mathematica Policy Research
Eric Isenberg
Jeffrey Max
Philip Gleason
Liz Potamites
Robert Santillano
Heinrich Hock

American Institutes for Research
Michael Hansen

Lauren Angelo
Project Officer
Institute of Education Sciences

NCEE 2014-4001
U.S. DEPARTMENT OF EDUCATION

NATIONAL CENTER FOR EDUCATION EVALUATION AND REGIONAL ASSISTANCE
Institute of Education Sciences

This page has been left blank for double-sided copying.

U.S. Department of Education
Arne Duncan
Secretary

Institute of Education Sciences
John Q. Easton
Director

National Center for Education Evaluation and Regional Assistance
Ruth Curran Neild
Commissioner

November 2013

The report was prepared for the Institute of Education Sciences under Contract No. ED-IES-10-C-0065. The project officer is Lauren Angelo in the National Center for Education Evaluation and Regional Assistance.

IES evaluation reports present objective information on the conditions of implementation and impacts of the programs being evaluated. IES evaluation reports do not include conclusions or recommendations or views with regard to actions policymakers or practitioners should take in light of the findings in the reports.

Isenberg, Eric, Jeffrey Max, Philip Gleason, Liz Potamites, Robert Santillano, Heinrich Hock, and Michael Hansen (2013). *Access to Effective Teaching for Disadvantaged Students* (NCEE 2014-4001). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

This report is available on the IES website at http://ies.ed.gov/ncee.

Upon request, this report is available in alternate formats such as Braille, large print, audiotape, or computer diskette. For more information, please contact the Department's Alternate Format Center at 202-260-9895 or 202-205-8113.

This page has been left blank for double-sided copying.

ACKNOWLEDGMENTS

This study was a collaborative effort that benefited from the contributions of many people. 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 who invested the time and energy in collecting and providing the data for this report.

This report benefited from the input of the study's technical working group: Brian Jacob, J.R. Lockwood, Jennifer King Rice, Jonah Rockoff, Andrew Rotherham, Tim Sass, Petra Todd, Victoria Van Cleef, and Jacob Vigdor. Steve Glazerman provided valuable feedback on the study design and Brian Gill provided a technical review and comments for the report. We also thank two anonymous reviewers who reviewed and commented on the report.

This report benefited from the contributions of many staff at Mathematica and our research partner, the American Institutes for Research. A strong team of programmers brought exceptional skill to the preparation and analysis of data for the report. Mary Grider led the team of programmers, and Emma Kopa and Matthew Jacobus provided expert programming to implement the analysis for this report. Will Crumbley, Kai Filion, Serge Lukashanets, Andrew McGuirk, and Ama Takyi-Laryea made critical contributions to the data cleaning and analysis for this report. The programming team from the American Institutes for Research consisted of Victoria Brady, Matthew Corritore, and Thomas Gonzalez. We thank the team of researchers and systems analysts who provided valuable guidance to programmers in preparing the data for analysis: Russell Cole, Matthew Johnson, Joanne Lee, Albert Liu, Tim Novak, Ali Protik, and Chris Rodger. We also thank the team of staff who secured the participation of districts in the study and collected data from the study districts for their valuable contributions: Julie Bruch, Patty Cloud, Adam Dunn, Brittany English, Celina Kamler, Rebecca Kleinman, Virginia Knechtel, Bryce Onaran, and Claire Smither Wulsin.

This report was edited by Lisa Ferraro Parmelee and produced by Jackie McGee.

This page has been left blank for double-sided copying.

CONTENTS

TABLES

FIGURES

EXECUTIVE SUMMARY

In this report, we describe disadvantaged students' access to effective teaching in grades 4 through 8 in 29 diverse school districts, using value-added analysis to measure effective teaching. Recent federal initiatives emphasize measuring teacher effectiveness and ensuring that disadvantaged students have equal access to effective teachers. These include Race to the Top, the Teacher Incentive Fund, and the flexibility policy for the Elementary and Secondary Education Act, which allows states to waive a number of provisions in exchange for a commitment to key reform principles (U.S. Department of Education 2009, 2012a).

Federal efforts to promote the equitable distribution of effective teachers arise from concerns that disadvantaged students may have less access to effective teachers, thereby contributing to sizable achievement gaps for disadvantaged students (Reardon 2011; U.S. Department of Education 2012b). A growing body of research uses value-added analysis to measure teacher effectiveness and examine the extent to which disadvantaged students have access to effective teachers. Value added measures a teacher's contribution to student learning, accounting for a student's previous achievement level and background characteristics. Studies consistently find considerable variation in teacher effectiveness based on value-added measures (Nye et al. 2004; Rockoff 2004; Rivkin et al. 2005; Kane et al. 2006; Aaronson et al. 2007; Koedel and Betts 2009). In addition, there is evidence of better long-run outcomes for students taught by more effective teachers as measured by value added, including lower rates of teen pregnancy, increased likelihood of college attendance, and higher wages (Chetty et al. 2011).

Given the importance of teachers in improving student achievement and concerns about unequal access to effective teachers (Jerald et al. 2009; Brown and Haycock 2011), more evidence on access to effective teaching is needed. This report focuses on access to effective teaching in 29 school districts over the 2008-2009 to 2010-2011 school years.

The main findings are:

- **On average, disadvantaged students had less access to effective teaching in the 29 study districts in grades 4 through 8. The magnitude of differences in effective teaching for disadvantaged and nondisadvantaged students in a given year was equivalent to a shift of two percentile points in the student achievement gap.** Students eligible for a free or reduced-price lunch (FRL) experienced less effective teaching than non-FRL students on average within districts, with statistically significant differences of 0.034 standard deviations of student test scores in English/language arts (ELA) and 0.024 standard deviations in math. Providing equal access to effective teaching for FRL and non-FRL students would reduce the student achievement gap from 28 percentile points to 26 percentile points in ELA and from 26 percentile points to 24 percentile points in math in a given year. In one alternative model specification, however, access to effective teaching for disadvantaged students and nondisadvantaged students was not statistically different.[1]

[1] We conducted this alternative model in 9 districts for the upper elementary grades and 23 districts for the middle school grades where the necessary data were available.

- **Access to effective teaching for disadvantaged students did not change over time in the study districts.** Average differences in effective teaching between FRL and non-FRL students did not differ over the three study years for either ELA or math.
- **Disadvantaged students' access to effective teaching varied across school districts.** Access to effective teaching varied across study districts, ranging from districts with equal access to districts with differences in effective teaching for FRL and non-FRL students as large as 0.106 standard deviations of student test scores in ELA and 0.081 standard deviations of student test scores in math. Disadvantaged students did not have greater access to effective teaching in any school district in the sample.
- **Unequal access to effective teaching was most related to the school assignment of teachers and students rather than to the way that teachers were assigned to students within schools.** The average between-school measure of access to effective teaching was significantly greater than the average within-school measure in both the upper elementary and middle school grades. Differences in effective teaching between schools for FRL and non-FRL students were larger than differences within schools by 0.020 standard deviations of student test scores in ELA and by 0.008 standard deviations in math. In other words, unequal access to effective teaching depended more on FRL students attending *schools* with less effective teaching than on FRL students being assigned to *classrooms* (within schools) with less effective teaching.

Research Questions and Study Overview

To address the need for evidence on access to effective teaching, the U.S. Department of Education's Institute of Education Sciences (IES) contracted with Mathematica Policy Research to study the issue in a diverse set of school districts over the five-year period from the 2008–2009 to the 2012–2013 school years. The study's primary research questions are:

1. To what extent do disadvantaged students have equal access to effective teaching within school districts, and how does this change over time?
2. Is access to effective teaching related to different patterns of teacher hiring, retention, and mobility for high- and low-poverty schools?

This study builds on the current evidence base in three ways. First, it documents access to effective teaching in districts that are diverse in terms of geography and size, with 29 districts in 16 states and all 4 U.S. Census regions. Second, it examines whether access to effective teaching changes over time. In this report, we measure access to effective teaching over a three-year period. Ultimately, we will measure changes over a five-year period. Third, we measure the extent of inequities between as well as within schools, allowing us to incorporate the effects of both between-school sorting of students and teachers to schools and within-school matching of teachers to students.

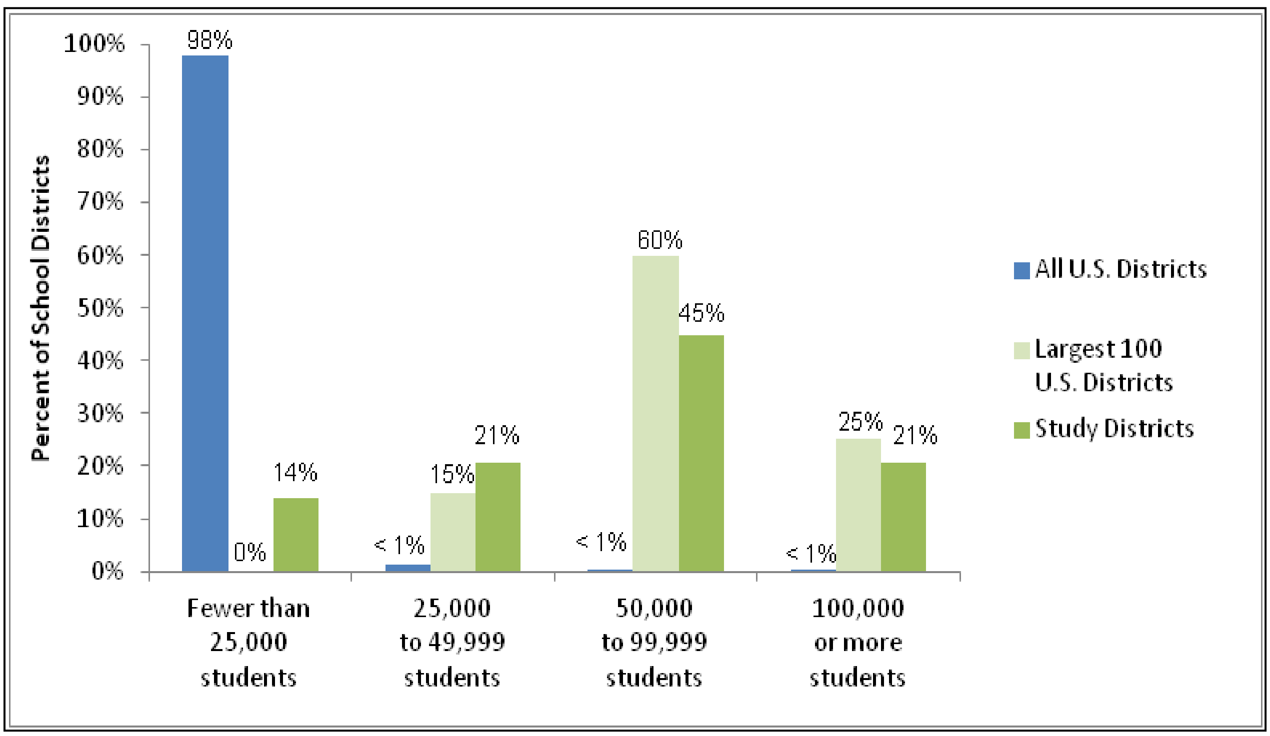
In this report, the first of three, we provide results that answer the first research question based on the first three years of the study (2008–2009 through 2010–2011 school years). The second report will address the second research question for the same school years, and the final report will update the results for both research questions to cover an additional two years (through the 2012–2013 school year).

Participating Districts

To document access to effective teaching in a diverse set of districts, the recruitment and selection of districts focused on obtaining a geographically diverse sample that could provide the data needed for a value-added analysis of teachers. We sought districts with a mix of free or reduced-price lunch (FRL) and non-FRL students—because we measure differences in effective teaching between these two groups of students—and districts that implemented different types of policies.

The 29 study districts are geographically diverse, with at least 4 districts from each region of the country. In these study districts, the percentage of students from the South and Midwest is similar to the national distribution, students from the North are underrepresented, and students from the West are overrepresented. The study districts more closely resemble the 100 largest districts than they resemble all districts in the United States. The study districts are large, on average, with a median enrollment of 60,000 students, and are located in medium-sized or large cities (Figure ES.1). Sixteen of the 29 study districts have more than 75 percent of students in a large city. The average study district has an FRL rate of 63 percent, with a range of 34 to 78 percent. Thirty-one percent of students in study districts are Black, 40 percent are Hispanic, and 18 percent are English-language learners.

Figure ES.1. Distribution of Study Districts, U.S. Districts, and Largest 100 U.S. Districts by Size



Source: 2008–2009 Common Core of Data.

Most study districts regard equitable access to effective teaching as a policy priority, but there is variation across districts in the types of policies they are implementing that may affect access to effective teaching. According to interviews with district staff that we conducted, a majority of our districts (17 of the 29) described equitable access to effective teaching as a policy priority. However, most districts (22 of 29) reported they had not used data on teacher effectiveness, such as value-added measures or teacher evaluation ratings, to assess access to effective teaching. Of the 12 policies potentially related to disadvantaged students' access to effective teaching that we asked about, the most commonly reported were in the areas of school improvement and teacher development policies. At least half of the study districts reported using these policies.

Measuring Access to Effective Teaching

To measure whether disadvantaged students have equal access to effective teaching, we calculated what we refer to as the Effective Teaching Gap (ETG). The ETG is a measure that compares the average effectiveness of teaching experienced by nondisadvantaged students with the average effectiveness of teaching received by disadvantaged students. A positive ETG means that the typical disadvantaged student experiences or has access to less effective teaching than the typical nondisadvantaged student, while a negative ETG means that disadvantaged students experience more effective teaching. An ETG of zero indicates that disadvantaged students have equal access to effective teaching.

To further understand access to effective teaching between and within schools in a district, we separated each district's ETG into between-school and within-school ETGs. The district ETG is the sum of the between- and within-school ETGs. Access to effective teaching can differ between schools if disadvantaged students attend schools that have less effective teaching on average than those attended by nondisadvantaged students. These between-school differences are related to how families select schools and how teachers come to be employed—and remain employed—in those schools. Access to effective teaching can also differ within a given school. Within-school differences can occur if teacher-student assignment within schools differs systematically for disadvantaged versus nondisadvantaged students.

Study Design

Student Sample. We examined access to effective teaching in English/language arts (ELA) and math among students in 29 study districts in grades 4 through 8. These are the subjects and grades for which test score data are available from the end of the current and prior school years.

Data. We collected administrative data to estimate teacher-level value-added models and measure access to effective teaching in study districts. In particular, we collected four years of standardized student test scores from state assessments in grades 3 through 8, a set of student characteristics (FRL status, limited English proficiency, special education status, gender, race, and ethnicity), school enrollment data for students, and teacher-student-course links indicating the teacher responsible for teaching ELA and/or math to each student. We report results from the 2008–2009 through 2010–2011 school years for 24 districts, and results from the 2007–2008 through 2009–2010 school years for the other 5 districts where we gathered data from state databases that were lagged by one year. We also collected information on district policies by conducting a document search and then using information from the documents to inform interviews with district staff.

Measuring Effective Teaching. To measure effective teaching, we used value-added analysis, a statistical approach to isolate a teacher's contribution to student achievement. It measures the achievement levels of a teacher's students after accounting for students' prior achievement levels and other characteristics, such as special education or English language learner status that may be related to student achievement during the year. A value-added model predicts the test score each student would have achieved with the average teacher in a district or state, and then compares the average actual performance of a given teacher's students to the average of these students' predicted scores. The difference between the two scores is attributed to the teacher as his or her value-added estimate. One critique of value added is that unmeasured differences between students could bias value-added estimates, but some evidence suggests that unmeasured student characteristics do not play a large role in determining teacher value added (Kane and Staiger 2008; Chetty et al. 2011).

Measuring Access to Effective Teaching. We measured access to effective teaching within each district using the Effective Teaching Gap (ETG). We calculated the district ETG in four steps:

Step 1: Use value-added analysis to measure the effectiveness of each teacher in the district.

Step 2: Assign each student in the district the value added of his or her teacher in the relevant subject. This value-added estimate represents the effectiveness of teaching experienced by the student for a given subject.

Step 3: Using students' free or reduced-price lunch (FRL) status as the measure of disadvantage, we calculate the mean value-added estimate among all nondisadvantaged students in the district and conduct the same calculation among all disadvantaged students.

Step 4: Calculate the district ETG by subtracting the mean value-added estimate for disadvantaged students from the mean value-added estimate for nondisadvantaged students.

Measuring Between- and Within-School Access to Effective Teaching. We calculated the between-school ETG following the same steps described above for the district ETG, but we replaced teacher value added with the average value added of the teachers within each school, grade, subject, and year. For simplicity, we refer to this as the "school value-added estimate" (even though it is computed separately for each grade within a school). It measures the effectiveness of the average teacher at the school. By assigning each student (FRL or non-FRL) the school value-added estimate, we eliminated any differences in access to effective teaching that can arise from the assignment of teachers to students within a school. Thus, the between-school ETG can be calculated by taking the difference in average school value added between non-FRL and FRL students. Since the sum of the between- and within-school ETGs is the district ETG, we first calculated the between-school ETG and then subtracted it from the district ETG to determine the within-school ETG.

We focus solely on value added in this study for two reasons. First, comparisons of teacher effectiveness based on multiple measures are not feasible in this study because other measures of teacher effectiveness, such as structured observations of classroom practices, are not conducted in all study districts or not measured consistently across districts. This study takes advantage of the existing student achievement data available from districts to measure access to effective teaching in a consistent way across districts. Second, value added is a policy-relevant measure because current federal policy encourages the use of student achievement growth (of which value added is an example) as a significant factor in assessing teacher effectiveness.

Access to Effective Teaching

We examined whether disadvantaged students have equal access to effective teaching within districts by documenting the size of the ETG in the 29 study districts for grades 4 through 8. We present evidence on the ETG separately for ELA and math.

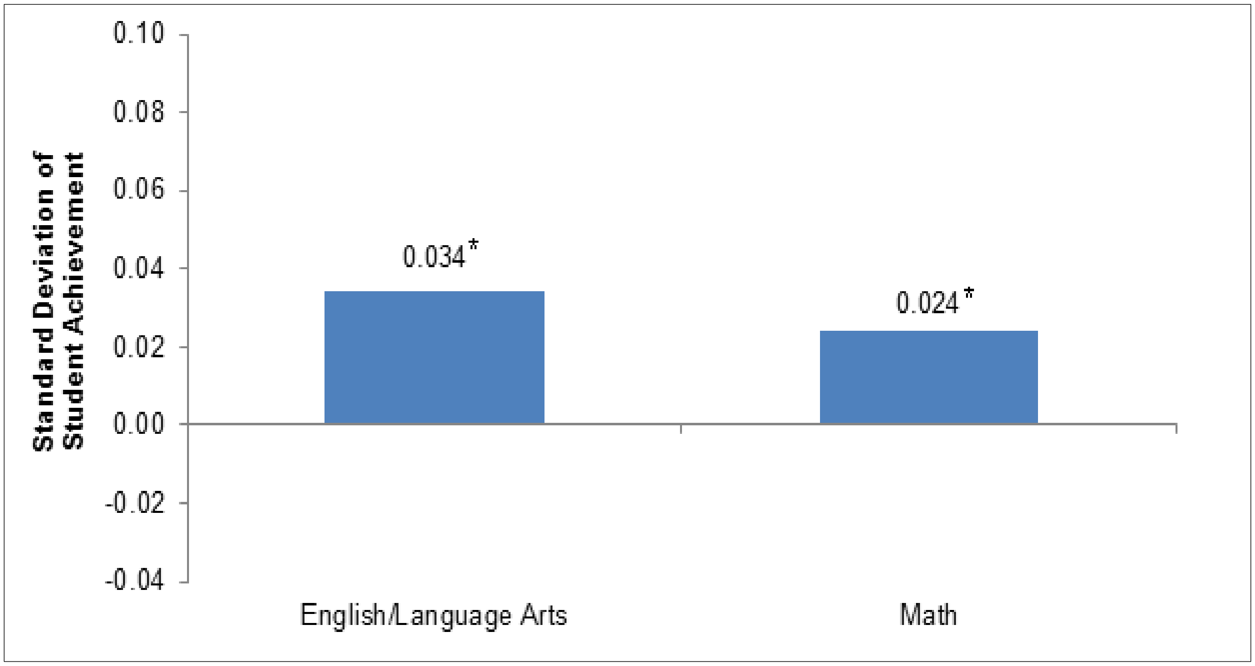
On average in the 29 study districts and across the three school years, disadvantaged students did not have equal access to effective teaching. The differences in effective teaching for FRL and non-FRL students in a given year were equivalent to a shift of two percentile points in the student achievement gap. Teachers of non-FRL students had higher value added than teachers of FRL students on average, with statistically significant differences of 0.034 standard deviations of student test scores in ELA and 0.024 standard deviations in math (Figure ES.2). The results imply that the typical FRL student experiences less effective teaching than the typical non-FRL student within a district. In addition, the average ETG did not significantly differ over the three years of the study for either subject.

Reducing the ETG from its current level to zero for one year—in other words, providing equal access to effective teaching for FRL and non-FRL students—would reduce the student achievement gap in the average study district from 28 to 26 percentile points in ELA and from 26 to 24 percentile points in math. We also calculated how the average teacher of a nondisadvantaged student compares to the average teacher of a disadvantaged student in terms of percentiles of the teacher distribution. For ELA, the average teacher of nondisadvantaged students is at the 56th percentile in the teacher distribution, compared to the average teacher of disadvantaged students at the 47th percentile. In math, it is the difference between a teacher at the 53rd percentile and a teacher at the 48th percentile.

The main findings were not sensitive to calculating the ETG based on a comparison of students in different racial/ethnic groups rather than different FRL status. The Black/White ETG and Hispanic/White ETG differ from the FRL ETG by no more than 0.005 standard deviations of student test scores in either subject.

We also calculated the ETG when effective teaching is based on two alternative value-added models, a value-added model that used an additional year of baseline test scores of students and a separate model that incorporated characteristics of students' classroom peers to capture peer effects. In the alternative model with two years of baseline scores, disadvantaged students had less access to effective teaching. However, in the alternative model incorporating peer effects, access to effective teaching for disadvantaged and nondisadvantaged students was not statistically different.

Figure ES.2. Effective Teaching Gap in Study Districts, 29-District Average, Years 1 to 3



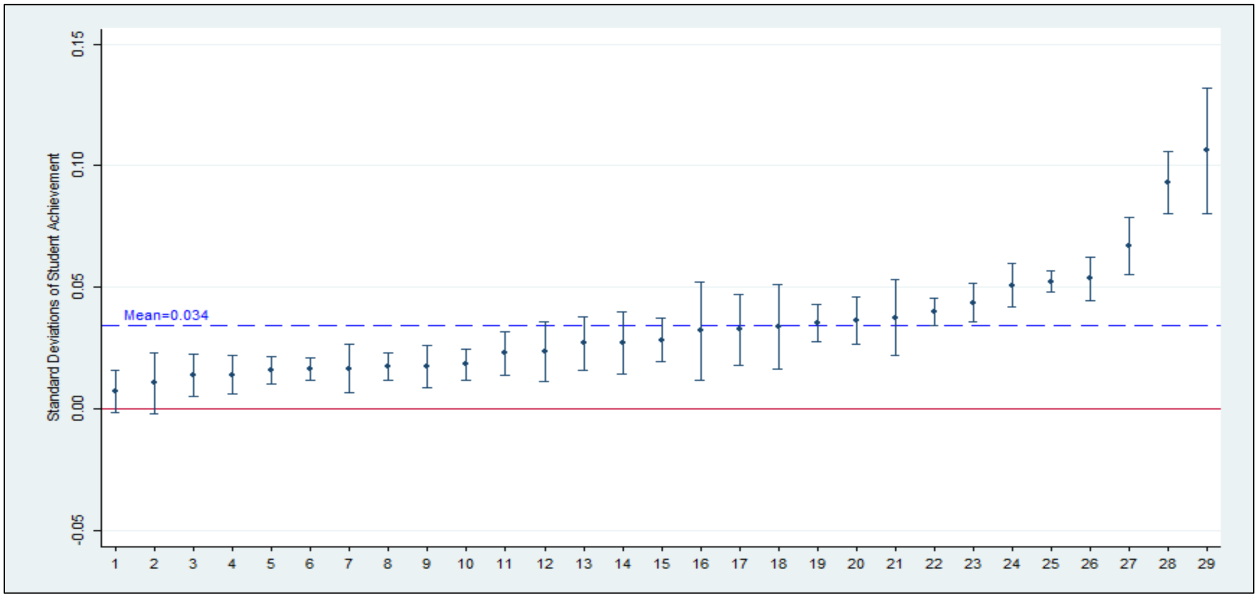
Source: District administrative data

Note: Results are for 29 districts, grades 4 through 8, and years 1 to 3. District-level results are weighted across grades and years by the number of teachers. Overall results are weighted equally across districts.

*Indicates statistical significance at the 0.05 level, two-tailed test. Statistical significance is based on variation across districts.

Access to effective teaching for FRL and non-FRL students varied across study districts, with equal access to effective teaching in some districts and unequal access favoring non-FRL students in other districts. The ETG ranges from districts with equal access (ETGs not significantly different from zero) to districts with ETGs as large as 0.106 in ELA and 0.081 in math (Figures ES.3 and ES.4). Variation in the ETGs across study districts is greater than would be expected to occur by chance. We found unequal access to effective teaching in 27 of the 29 districts in ELA and in 19 of the 29 districts in math. In the remaining districts, disadvantaged students have equal access to effective teaching, as shown by ETGs that are not significantly different from zero. None of the study districts has a statistically significant ETG favoring FRL students.

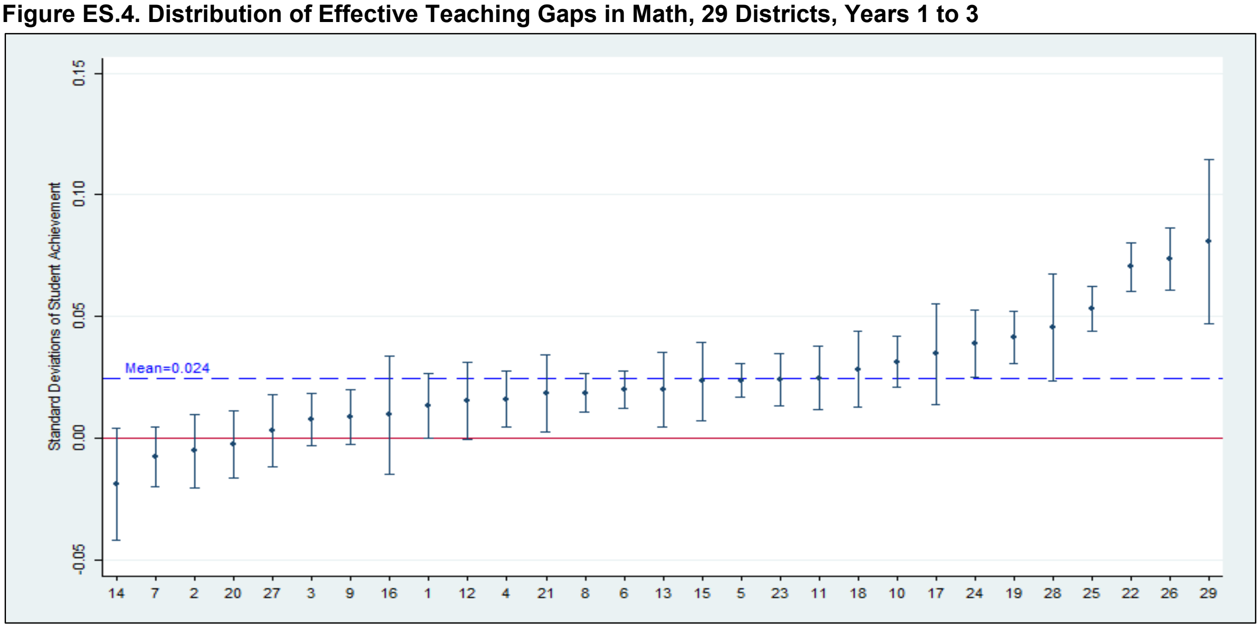
Figure ES.3. Distribution of Effective Teaching Gaps in English/Language Arts, 29 Districts, Years 1 to 3



Source: District administrative data.

Note: Results are for 29 districts, grades 4 through 8, and years 1 to 3. ETGs are computed within each district-grade-year combination and averaged with equal weight across years within each district. Circles represent the district-level ETGs and the vertical lines show the 95 percent confidence intervals around each point. The cross-district average of 0.034 standard deviations is shown by the dashed horizontal line. Districts are ordered by the size of the ETG. District codes 1 to 29 are assigned in order of the size of the district ETG in English/language arts and are consistent across figures. The ETG is expressed in terms of standard deviations of student test scores.

Figure ES.4. Distribution of Effective Teaching Gaps in Math, 29 Districts, Years 1 to 3



Source: District administrative data.

Note: Results are for 29 districts, grades 4 through 8, and years 1 to 3. ETGs are computed within each district-grade-year combination and averaged with equal weight across years within each district. Circles represent the district-level ETGs and the vertical lines show the 95 percent confidence intervals around each point. The cross-district average of 0.024 standard deviations is shown by the dashed horizontal line. Districts are ordered by the size of the ETG. District codes 1 to 29 are assigned in order of the size of the district ETG in English/language arts and are consistent across figures. The ETG is expressed in terms of standard deviations of student test scores.

Differences in access to effective teaching are larger between schools than within schools. States and districts often focus on access to effective teaching between schools but do not capture inequities within schools. The ETG allows us to separately measure the degree to which differences in effective teaching occur between schools and within schools. We find that most of the district ETG is accounted for by between-school differences, but some differences in access to effective teaching arise within schools as well.

The between-school ETG is larger than the within-school ETG, especially in ELA at the elementary grades. For ELA, the between-school ETG is 0.029 standard deviations of student test scores larger than the within-school ETG in the upper elementary grades (grades 4 and 5) and 0.014 standard deviations larger in the middle school grades (Table ES.1). In addition, the difference in the between- and within-school ETGs is significantly larger for upper elementary grades than for middle school grades. District-level results for ELA are shown in Figures ES.5 and ES.6.

The between-school ETG for math is also significantly larger than the within-school ETG. However, the between-school differences in effective teaching do not explain unequal access to the same degree that it does in ELA. The between-school ETG is larger than the within-school ETG by 0.011 in the upper elementary grades and 0.005 in the middle school grades. District-level results for math are shown in Figures ES.7 and ES.8.

The patterns of between- and within-school ETGs may be related to the tendency for elementary schools to be smaller than middle schools, resulting in a more homogenous student population due to less diversity in household income within smaller attendance areas.

Table ES.1. Average Between-School and Within-School Effective Teaching Gaps, 29-District Average, Years 1 to 3

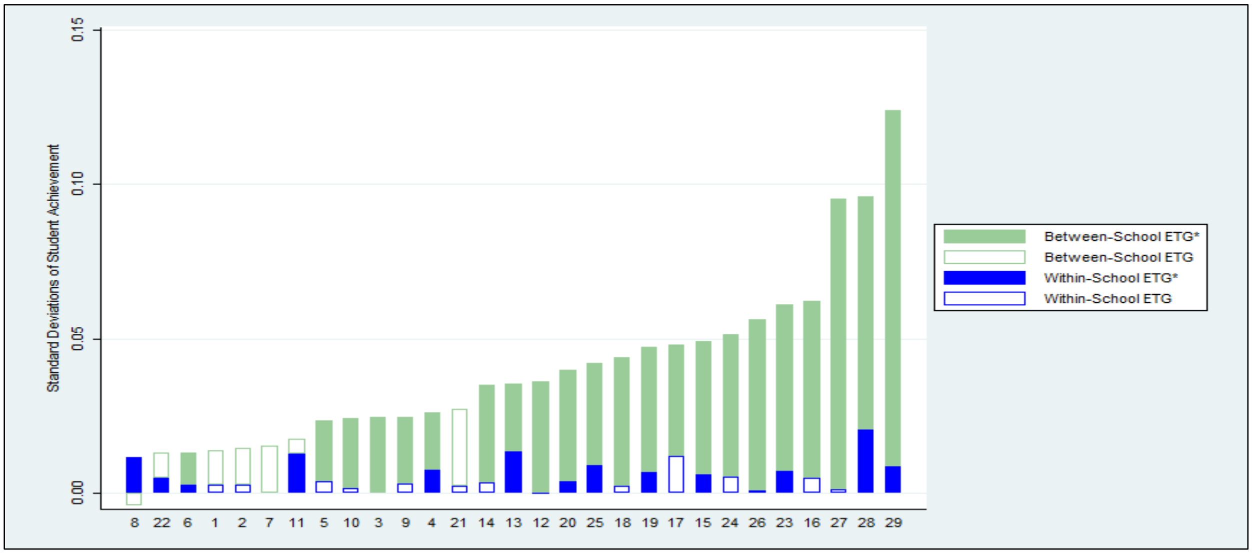
Subject	Between-School	Within-School	Difference	P-Value
English/Language Arts				
Upper Elementary	0.035	0.005	0.029*	0.00
Middle	0.022	0.008	0.014*	0.00
All Grades	0.027	0.007	0.020*	0.00
Math				
Upper Elementary	0.016	0.005	0.011*	0.00
Middle	0.016	0.011	0.005*	0.04
All Grades	0.016	0.008	0.008*	0.00

Source: District administrative data

Note: Results are for 29 districts, grades 4 through 8, and years 1 to 3. Upper elementary is grades 4 and 5; middle school is grades 6 through 8. District-level results are weighted across grades and years by the number of teachers. Overall results are weighted equally across districts. The difference in the between- and within-school ETG is the average of the differences for individual districts. The ETG is expressed in terms of standard deviations of student test scores.

*Indicates statistical significance at the 0.05 level, two-tailed test. We test whether differences in the between- and within-school ETGs are statistically significant by using variation within districts.

Figure ES.5. Between-School and Within-School Effective Teaching Gaps in English/Language Arts by District for Upper Elementary Grades, 29 Districts, Years 1 to 3

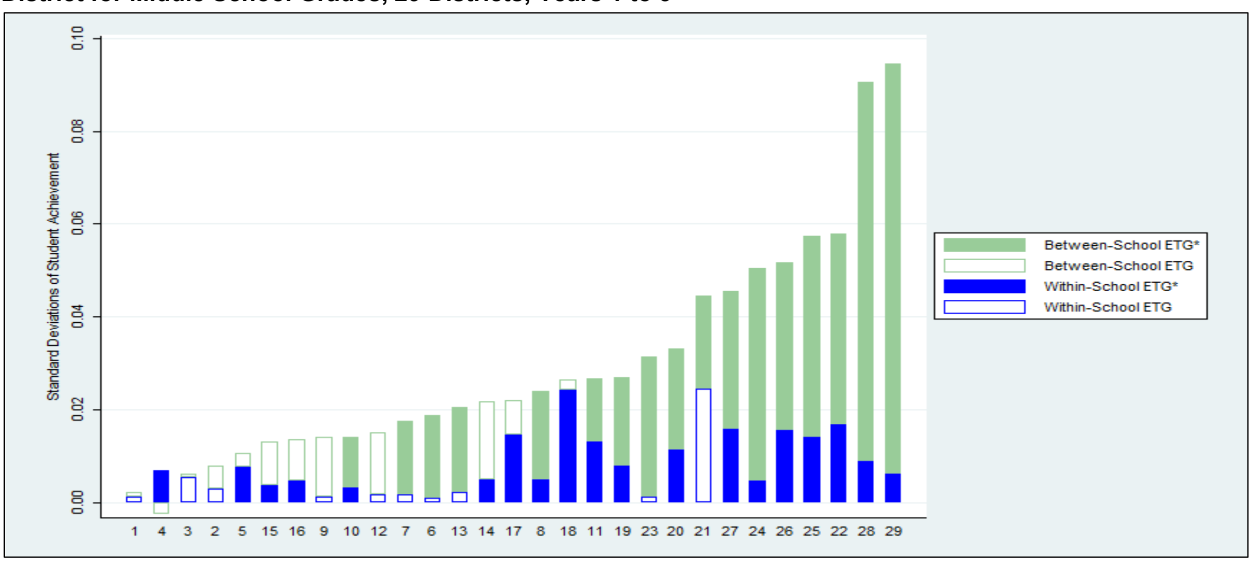


Source: District administrative data.

Note: Results are for 29 districts, grades 4 and 5, and years 1 to 3. The ETGs are computed within each district-grade-year combination and then averaged with equal weight across years within each district. Districts are ordered by the size of the district ETG. District codes 1 to 29 are assigned in order of the size of the district ETG in English/language arts and are consistent across figures. The solid bars show between- or within-school ETGs that are significantly different from zero at the 0.05 level; the hollow bars indicate between- or within-school ETGs that are not significantly different from zero.

*Indicates statistical significance at the 0.05 level, two-tailed test.

Figure ES.6. Between-School and Within-School Effective Teaching Gaps in English/Language Arts by District for Middle School Grades, 29 Districts, Years 1 to 3

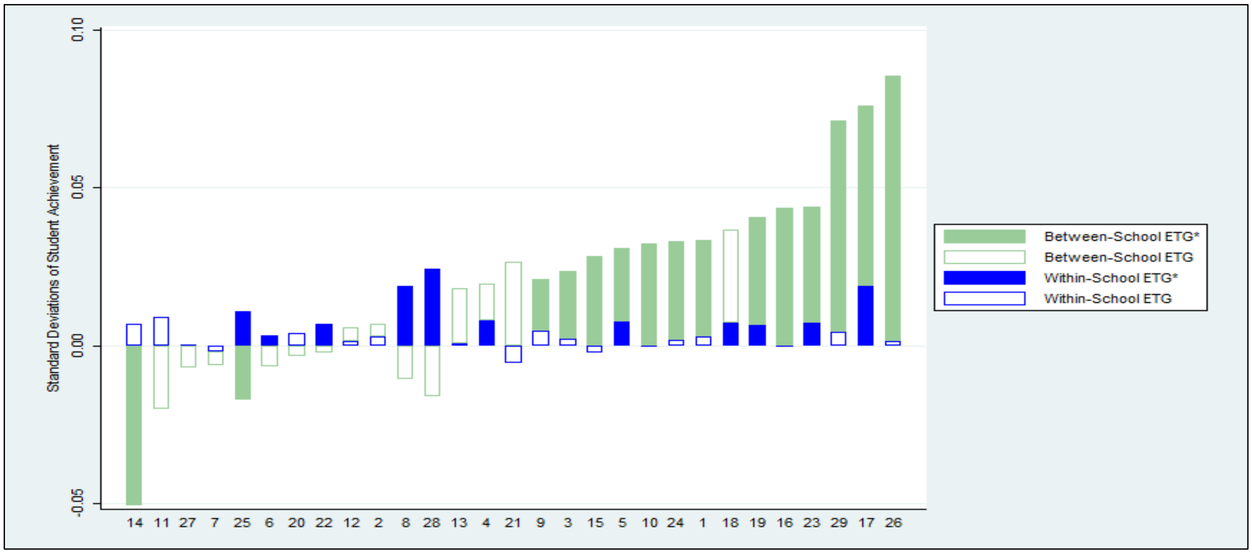


Source: District administrative data.

Note: Results are for 29 districts, grades 6 to 8, and years 1 to 3. The ETGs are computed within each district-grade-year combination and then averaged with equal weight across years within each district. Districts are ordered by the size of the district ETG. District codes 1 to 29 are assigned in order of the size of the district ETG in English/language arts and are consistent across figures. The solid bars show between- or within-school ETGs that are significantly different from zero at the 0.05 level; the hollow bars indicate between- or within-school ETGs that are not significantly different from zero.

*Indicates statistical significance at the 0.05 level, two-tailed test.

Figure ES.7. Between-School and Within-School Effective Teaching Gaps in Math by District for Upper Elementary Grades, 29 Districts, Years 1 to 3

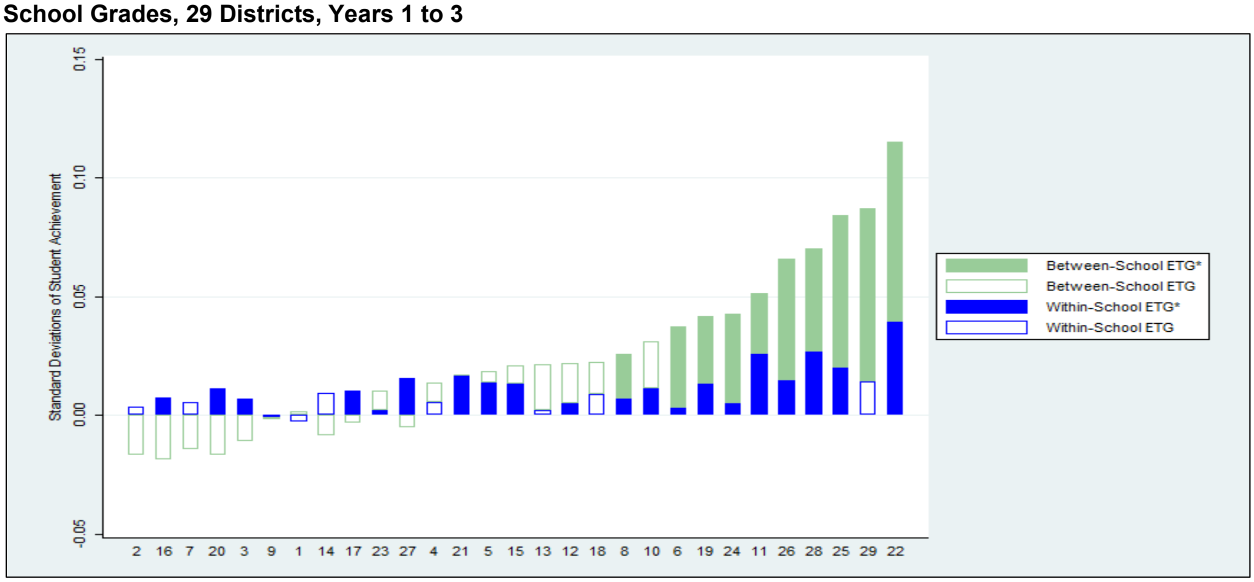


Source: District administrative data.

Note: Results are for 29 districts, grades 4 and 5, and years 1 to 3. The ETGs are computed within each district-grade-year combination and then averaged with equal weight across years within each district. Districts are ordered by the size of the district ETG. District codes 1 to 29 are assigned in order of the size of the district ETG in English/language arts and are consistent across figures. The solid bars show between- or within-school ETGs that are significantly different from zero at the 0.05 level; the hollow bars indicate between- or within-school ETGs that are not significantly different from zero.

*Indicates statistical significance at the 0.05 level, two-tailed test.

Figure ES.8. Between-School and Within-School Effective Teaching Gaps in Math by District for Middle School Grades, 29 Districts, Years 1 to 3



Source: District administrative data.

Note: Results are for 29 districts, grades 6 to 8, and years 1 to 3. The ETGs are computed within each district-grade-year combination and then averaged with equal weight across years within each district. Districts are ordered by the size of the district ETG. District codes 1 to 29 are assigned in order of the size of the district ETG in English/language arts and are consistent across figures. The solid bars show between- or within-school ETGs that are significantly different from zero at the 0.05 level; the hollow bars indicate between- or within-school ETGs that are not significantly different from zero.

*Indicates statistical significance at the 0.05 level, two-tailed test.

This page has been left blank for double-sided copying.

I. INTRODUCTION

In this report, we describe disadvantaged students' access to effective teaching in grades 4 through 8 in 29 diverse school districts, using value-added analysis to measure effective teaching. Recent federal initiatives emphasize measuring teacher effectiveness and ensuring that disadvantaged students have equal access to effective teachers. These include Race to the Top, the Teacher Incentive Fund, and the flexibility policy for the Elementary and Secondary Education Act, which allows states to waive a number of provisions in exchange for a commitment to key reform principles (U.S. Department of Education 2009, 2012a).

Federal efforts to promote the equitable distribution of effective teachers arise from concerns that disadvantaged students may have less access to effective teachers, thereby contributing to sizable achievement gaps for disadvantaged students (Reardon 2011; U.S. Department of Education 2012b). There is some evidence that teachers prefer to teach in schools with fewer disadvantaged students. Schools that serve more disadvantaged students have more teacher turnover, and teachers who move tend to transfer into schools with fewer disadvantaged students (Lankford et al. 2002, Hanushek et al. 2005, Scafidi et al. 2007, Boyd et al. 2008, Jackson 2011, Feng and Sass 2012). There is also some evidence that disadvantaged students are more likely to be assigned to less qualified teachers within schools (Kalogrides et al. 2013).

Some studies have used teacher qualifications as a measure of effective teaching, including years of teaching experience, teacher test scores, and credentials such as teacher certification or attaining a master's degree. These studies show that disadvantaged students are more likely to be taught by less qualified teachers as measured in these ways (Carroll et al. 2000; Lankford et al. 2002; Presley et al. 2005; Clotfelter et al. 2006; Education Trust 2008). However, most research has found no consistent empirical link between student learning and these teacher qualifications, with the exception of whether a teacher is relatively new to teaching (Rivkin et al. 2005; Kane et al. 2006; Gordon et al. 2006; Aaronson et al. 2007; Koedel and Betts 2007).

A growing body of research uses a different measure—value-added analysis—to measure teacher effectiveness and examine the extent to which disadvantaged students have access to effective teachers. Value added measures a teacher's contribution to student learning, accounting for the student's previous achievement level and background characteristics. Studies using this approach consistently find considerable variation in teacher effectiveness (Nye et al. 2004; Rockoff 2004; Rivkin et al. 2005; Kane et al. 2006; Aaronson et al. 2007; Koedel and Betts 2009). In addition, there is evidence of better long-run outcomes for students taught by more effective teachers as measured by value added, including lower rates of teen pregnancy, increased likelihood of college attendance, and higher wages (Chetty et al. 2011).

Given the importance of teachers in improving student achievement and concerns about unequal access to effective teachers (Jerald et al. 2009; Brown and Haycock 2011), more evidence on access to effective teaching is needed. To address this need, the U.S. Department of Education's Institute of Education Sciences (IES) contracted with Mathematica Policy Research to examine access to effective teaching in a diverse set of school districts over a five-year period from the 2008–2009 to the 2012–2013 school years. The study's primary research questions are:

1. To what extent do disadvantaged students have equal access to effective teaching within school districts, and how does this change over time?
2. Is access to effective teaching related to different patterns of teacher hiring, retention, and mobility for high- and low-poverty schools?

In this report, the first of three, we provide results that answer the first research question based on the first three years of the study (2008–2009 through 2010–2011 school years). A second report will address the second research question for the same school years, and the final report will update the results for both research questions to cover an additional two years (through the 2012–2013 school year).

To better understand the contribution of this study, we summarize findings from the existing research and then describe how this study builds on the current evidence. The existing research has several limitations. First, it focuses on a narrow range of districts. Second, it has concentrated largely on access to effective teaching between schools but not on differences that might arise within schools. Finally, the existing studies do not examine changes in access to effective teaching over time. Key findings from this research include:

- A study in Tennessee showed that schools with higher percentages of low-income and minority students had fewer of the most effective math teachers. In particular, these highly effective teachers made up 17 percent of all teachers in the more disadvantaged schools compared to 21 percent in schools with lower percentages of low-income and minority students (Tennessee Department of Education 2007).
- A study of 10 large districts found that effective teachers were distributed across schools differently in elementary grades compared to middle grades. Specifically, high value-added teachers in English/language arts (ELA) and math were underrepresented in the highest-poverty middle schools within districts (Glazerman and Max 2011). In middle school math, for example, 15 percent of teachers in the highest poverty schools were the highest performing compared to 29 percent in the lowest poverty schools. Highly effective elementary school teachers, in contrast, were evenly distributed among high-poverty and low-poverty schools. Access to effective teachers varied across the 10 districts: some districts were characterized by unequal access favoring low-poverty schools for both elementary and middle schools, some districts were characterized by unequal access favoring low-poverty schools just for middle schools, and two districts were characterized by unequal access favoring high-poverty elementary schools.
- In Los Angeles, students eligible for a free or reduced-price lunch (FRL) in upper elementary and middle school grades were less likely than non-FRL students to be taught by a highly effective teacher (Hahnel and Jackson 2012).
- A study of high school teachers in all subject areas in North Carolina found that teachers in the highest poverty schools had average value added that was 0.03 standard deviations of student achievement lower than the average value added of teachers in the lowest poverty schools (Mansfield 2012). The study compared teacher value added for the 25 percent of schools with the highest proportion of FRL students to the 25 percent of schools with the lowest proportion of FRL students.

- In North Carolina and Florida, elementary teachers in higher-poverty schools were less effective on average than teachers in lower-poverty schools for ELA in both states. The results were mixed for math, with teachers in higher-poverty schools on average less effective in North Carolina but more effective in Florida. The differences ranged from 0.01 to 0.04 standard deviations of student achievement (Sass et al. 2010).

As in earlier studies, we focus on ELA and math outcomes in grades 4 through 8, the subjects and grades for which test score data are available from the end of the current and prior school years. However, this study builds on the current evidence base in three ways.

First, it includes districts that are diverse in terms of geography and size. Earlier studies documented access to effective teaching in 3 southern states and for samples of one to 10 large or very large districts. In the current study, we document access to effective teaching in 29 medium to very large districts in 16 states and all four U.S. Census regions.

Second, this study will ultimately examine changes in access to effective teaching over a five-year period. Although earlier studies used multiple years of value-added data, none examined whether access to effective teaching changed over time.

Third, we measure the extent of inequities between as well as within schools. Most of the earlier research focuses on access to effective teachers between schools, ignoring potential within-school differences between disadvantaged and nondisadvantaged students in access to effective teachers. In this study, the measures of access to effective teaching incorporate the effects of both between-school sorting of students and teachers to schools and within-school assignment of teachers to students. Understanding the extent to which unequal access to effective teaching occurs within schools versus between schools can help policymakers identify relative sources of inequity and better target policies or programs to address any inequity. Policies that affect teacher hiring and retention, for example, are most likely to affect differences in access to effective teaching between schools, while policies that influence the way schools assign students to teachers can affect access within schools.

This page has been left blank for double-sided copying.

II. METHODS AND DATA

In this chapter, we describe the data and methods used to measure whether disadvantaged students have equal access to effective teaching. First, we outline our approach to measuring access to effective teaching. Then, we describe the data used for this analysis. We provide a more detailed description of the analytic methods in Appendix A.

A. Measuring Access to Effective Teaching

To examine disadvantaged students' access to effective teaching, we must first measure teacher effectiveness. In this study, we do so using value-added analysis—a statistical approach that isolates a teacher's contribution to student achievement. In the current context of teacher evaluations, federal policy emphasizes the use of multiple measures to assess teacher effectiveness (U.S. Department of Education 2012a). However, we focus solely on value added in this study for two reasons. First, comparisons of teacher effectiveness based on multiple measures are not feasible in this study because other measures of teacher effectiveness, such as structured observations of classroom practices, are not conducted in all study districts or not measured consistently across districts. This study takes advantage of the existing student achievement data available from districts to measure access to effective teaching in a consistent way across districts. Second, value added is a policy-relevant measure because current federal policy encourages the use of student achievement growth (of which value added is an example) as a significant factor in assessing teacher effectiveness. An overview of value added and its use by policymakers and researchers and a discussion of a few key concerns about the use of value-added are provided in the box on the next page.

To measure whether disadvantaged students have equal access to effective teaching, we calculated what we refer to as the Effective Teaching Gap (ETG). The ETG is a measure that compares the average effectiveness of teaching experienced by nondisadvantaged students with the average effectiveness of teaching received by disadvantaged students. A positive ETG means that the typical disadvantaged student experiences less effective teaching than the typical nondisadvantaged student, on average, while a negative ETG means that the disadvantaged student experiences more effective teaching. An ETG of zero indicates that disadvantaged students have equal access to effective teaching (Figure II.1).[1]

Figure II.1. Interpreting the Effective Teaching Gap for Disadvantaged Students

Disadvantaged Students Have Greater Access to Effective Teaching	Disadvantaged Students Have Equal Access to Effective Teaching	Disadvantaged Students Have Less Access to Effective Teaching
• ETG < 0 • Disadvantaged students receive *more effective* teaching, on average.	• ETG = 0 • Disadvantaged students receive *equally effective* teaching, on average.	• ETG > 0 • Disadvantaged students receive *less effective* teaching, on average.

[1] We refer throughout the report to "access to effective teaching" synonymously with the ETG. As described in Figure II.1, the ETG refers to whether or not disadvantaged students receive equally effective teaching. Although the term "access to effective teaching" suggests that the measure accounts for the full set of schools potentially available to households with disadvantaged and nondisadvantaged students, the ETG measures differences in effective teaching based on the choices households make rather than the choices they could have made.

Overview of Using Value Added to Measure Teacher Effectiveness

As the federal government's focus shifts from teacher qualifications to teacher effectiveness, state and district policymakers are seeking new ways to define and measure teacher effectiveness. Partially in response to ESEA Flexibility, many states are poised to develop teacher evaluation systems that combine observations of classroom practices with measures of a teacher's contribution to student achievement, such as value added (Mead 2012). Value-added estimates have been used by districts to evaluate teacher effectiveness (Value Added Research Center 2010; Isenberg and Hock 2012; Johnson et al. 2012), and by researchers to evaluate the role of teachers in student learning (Nye et al. 2004; Rockoff 2004; Rivkin et al. 2005; Kane et al. 2006; Aaronson et al. 2007; Koedel and Betts 2009; Chetty et al. 2011).

A value-added analysis attempts to isolate a teacher's contribution to student achievement using statistical methods. It measures the achievement levels of a teacher's students after accounting for students' prior achievement levels and other characteristics, such as special education or English language learner (ELL) status, that may be related to student achievement during the year. A value-added model predicts the test score each student would have achieved if taught by the average teacher in a district or state—that is, a teacher of average effectiveness—and then compares the average actual performance of a given teacher's students to the average of these students' predicted scores. The difference between the two scores is attributed to the teacher as his or her value-added estimate. Although value added does not measure every aspect of effective teaching, it is positively correlated with other methods of measuring effective teaching (Kane et al. 2012).

The implicit assumption of a value-added model is that if two classrooms contain students with identical measured baseline characteristics, those students will not differ systematically in ways that affect their achievement, such as by having different levels of motivation to succeed. Given that families select schools and that principals match students to teachers in ways that are typically not random, unmeasured differences between students in different schools and classrooms could bias value-added estimates. For example, especially motivated parents may choose schools for their children based on factors other than the student characteristics accounted for by the value-added model, or principals may assign students to teachers based on information about the teachers or their students that is not accounted for by these characteristics in the model, such as matching hard-to-teach students with teachers who fare relatively well with such students. Value-added researchers have debated the theoretical importance of systematic differences in unmeasured characteristics (Rothstein 2010; Goldhaber and Chaplin 2012) although empirical work in experimental settings (Kane and Staiger 2008) and quasi-experimental settings (Chetty et al. 2011) provides some evidence that differences in unobservable student characteristics do not play a large role in determining teacher value added.

A second concern is whether a value-added estimate isolates the effectiveness of each *teacher*—based on his or her knowledge, ability, and skills—or whether it measures the *teaching* received by students, including the teacher's instructional effectiveness and the efficacy of school inputs. For example, either a principal's leadership or school policies may affect student achievement in the classroom. Technically, value-added estimates measure effective teaching, as a value-added model cannot distinguish school inputs from teacher inputs. From a student's perspective, what matters is the effectiveness of the teaching he or she experiences, whatever the source, which is why we use value-added estimates to measure the Effective *Teaching* Gap. However, some evidence suggests that a teacher's value-added estimate is not unduly affected by school-level factors—the value added of teachers who change schools persists in the teachers' new settings (Chetty et al. 2011).

We calculated the district ETG by following four steps described in the following box.

Calculating the Effective Teaching Gap for a District

Step 1: Use value-added analysis to measure the effectiveness of each teacher in the district.

Step 2: Assign each student in the district the value-added estimate of his or her teacher in the relevant subject. This value-added estimate represents the effectiveness of teaching experienced by the student for a given subject.

Step 3: Using students' FRL status as the measure of disadvantage, calculate the mean value-added estimate among all nondisadvantaged students in the district and conduct the same calculation among all disadvantaged students.

Step 4: Calculate the district ETG by subtracting the mean value-added estimate for disadvantaged students from the mean value-added estimate for nondisadvantaged students.

To further learn about access to effective teaching within and between schools in a district, we separated each district's ETG into between-school and within-school ETGs (Figure II.2). The district ETG is the sum of the within- and between-school ETGs.

Figure II.2. Effective Teaching Gaps Between and Within Schools

District ETG		Within-school ETG		Between-school ETG
Compares the effectiveness of teaching experienced by disadvantaged and nondisadvantaged students	=	Compares the effectiveness of teaching experienced by disadvantaged and nondisadvantaged students *within the same schools*	+	Compares effectiveness of teaching experienced by disadvantaged and nondisadvantaged students *between schools*

Less access to effective teaching for disadvantaged students can occur between schools if they attend *schools* that have, on average, less effective teaching than those attended by nondisadvantaged students. These between-school differences, measured by the between-school ETG, are related to how families select schools and how teachers come to be employed—and remain employed—in those schools. Access to effective teaching can also differ within a given school. Within-school differences, measured by the within-school ETG, can occur if teacher-student assignment within schools differs systematically for disadvantaged versus nondisadvantaged students.

In the remainder of this section, we describe in further detail our methodology for measuring the ETG. We begin by describing the value-added model used to measure effective teaching (Step 1). We then describe the calculations that use these value added estimates to produce the district ETG (Steps 2 through 4) and the between-school and within-school ETG measures. We also briefly discuss alternative approaches to measuring value-added and student disadvantage.

1. Estimating Teacher Value Added

The first step in measuring access to effective teaching was to estimate the value added of participating districts' teachers in math and ELA for grades 4 through 8. We measured value added in the two subjects separately.

We designed a value-added model that would measure effective teaching for the purpose of subsequently measuring ETGs and comparing them across districts and years. Thus, we used the same data and methods in each district even when more data were available in certain districts. In this section, we describe the value-added model and highlight key decisions we made in developing the model.

Our basic approach was to rely on a regression model that accounted for a set of baseline student characteristics potentially related to academic achievement and that may otherwise be confounded with the assignment of students to teachers. Accounting for student background characteristics was important to ensure that certain teachers did not obtain high value-added estimates simply because they were assigned students who would likely have demonstrated strong achievement during the year regardless of the teacher. Specifically, we assumed that a student's post-test score depended on prior achievement, background characteristics, the student's current teachers, and additional unmeasured factors unrelated to teaching assignments.

In our value-added model we accounted for a common set of student characteristics in each study district to ensure that any differences in effective teaching across districts did not result from estimating different statistical models in different districts.[2] To allow the relationships between background characteristics and student achievement to vary by grade within a district, we estimated value-added models of effective teaching for each grade separately. The common value-added model included the following student characteristics, which we obtained from district administrative records:

- Math and ELA scores from the prior school year (we account for prior-year scores in both math and ELA regardless of the post-test subject)
- FRL status
- Limited English proficiency
- Special education status
- Gender
- Whether a student is African American or Black
- Whether a student is Hispanic, Native American, multi-race, or "other" race[3]

[2] Chetty et al. (2011) find that value-added estimates are robust to excluding characteristics such as household income or parents' marital status from a value-added model that includes pre-tests and other student characteristics commonly available from district administrative data.

[3] We combined multiple race and ethnicity groups into three categories for the value added model. Given that we estimate value-added models for each grade and subject separately, we wanted to avoid having race and ethnicity categories that lacked a sufficient number of students to precisely estimate the relationship between student

- Whether a student transferred across schools during the year

We included these characteristics because they may be correlated with factors that affect student achievement. For example, a student who transfers across schools during the year may have done so because of a disruptive environment at home, which could in turn affect achievement at school. As a second example, racial and ethnic characteristics may be correlated with the amount of resources a family has to support students in reaching higher levels of achievement. Although the model also includes FRL status, which captures family income, there may be differences in family resources even among families at the same income level. We also accounted for the amount of time a teacher spent with each student, weighting students in the teacher's value-added estimate in proportion to the amount of time they spent with the teacher from the beginning of the school year to the beginning of the testing window for the state test.

We based the value-added model on a single year of teacher performance because we are interested in comparing how ETGs change from year to year. In some contexts, researchers advise using multiple years of data to estimate a teacher's value added in order to more precisely estimate the permanent component of teacher effectiveness (McCaffrey et al. 2004). Multiyear estimates, however, capture effective teaching over the multiple years covered by the data, and may be biased estimates of effective teaching in a given year if there are true changes in effective teaching from year to year. Thus, there is a trade-off between obtaining unbiased estimates of effective teaching in a given year and increasing the precision of individual teachers' value-added estimates. Given that we averaged value-added estimates for multiple teachers when measuring ETGs rather than directly using value-added estimates for individual teachers, the precision gained by using multiple years of data is less valuable for this study than in other contexts. We therefore used single-year value-added measures.

Given that measurement error in the pre-test could lead to misleading inferences about the ETG, we included in our value-added model a correction for measurement error. Measurement error in the pre-test can lead to biased estimates of teachers' contributions to student achievement by weakening the association between pre-test and post-test.[4] Consequently, if the value-added model did not account for measurement error, a portion of a student's true prior achievement would be attributed to the teacher. This, in turn, could lead to biased estimates of the ETG, which is based on the relationship between FRL status and teachers' value-added estimates. We guarded against measurement error in student test scores by using an errors-in-variables technique based on published information on the test/re-test reliability of a given pre-test to correct for the bias that would otherwise arise from measurement error (Buonaccorsi 2010).

Ultimately, a value-added analysis produces estimates of the effectiveness of a set of teachers relative to one another. In our analysis, we estimated separate value-added models for

---

(continued)
race/ethnicity and achievement. The three categories provided a consistent and parsimonious approach for coding race and ethnicity that could accommodate the different race and ethnicity classification systems used by districts. The three categories were (1) African American or Black; (2) Hispanic, Native American, multi-race, or "other" race; and (3) Asian, Pacific Islander, and White.

[4] Measurement error in the post-test can lead to less precise value-added estimates but does not introduce bias.

teachers in different districts, grades, and subjects. Thus, the value-added estimate of an individual teacher in our sample was a measure of that teacher's effectiveness relative to other teachers of the same subject and in the same grade and district (Table II.1).

Table II.1. Interpretation of Teachers' Value-Added Estimates

Value-Added (VA) Estimate	Interpretation
VA Greater than 0	Teacher is more effective than the average teacher in the same subject, grade, and district. For a teacher with a value-added estimate of 0.15, for example, the average student of this teacher achieves test scores 0.15 standard deviations higher than would have been achieved under an average teacher in that subject, grade, and district.
VA Equal to 0	Teacher is as effective as the average teacher in the same subject, grade, and district. The average student of this teacher achieves test scores exactly the same as those he or she would have achieved under the average teacher.
VA Less than 0	Teacher is less effective than the average teacher in the same subject, grade, and district. For a teacher with a value-added estimate of -0.12, for example, the average student of this teacher achieves test scores 0.12 standard deviations lower than would have been achieved under an average teacher.

2. Measuring Effective Teaching Gaps

After generating a value-added estimate for each teacher, we linked each student to his or her teacher's value-added estimate and then calculated the district ETG for a given district as follows:

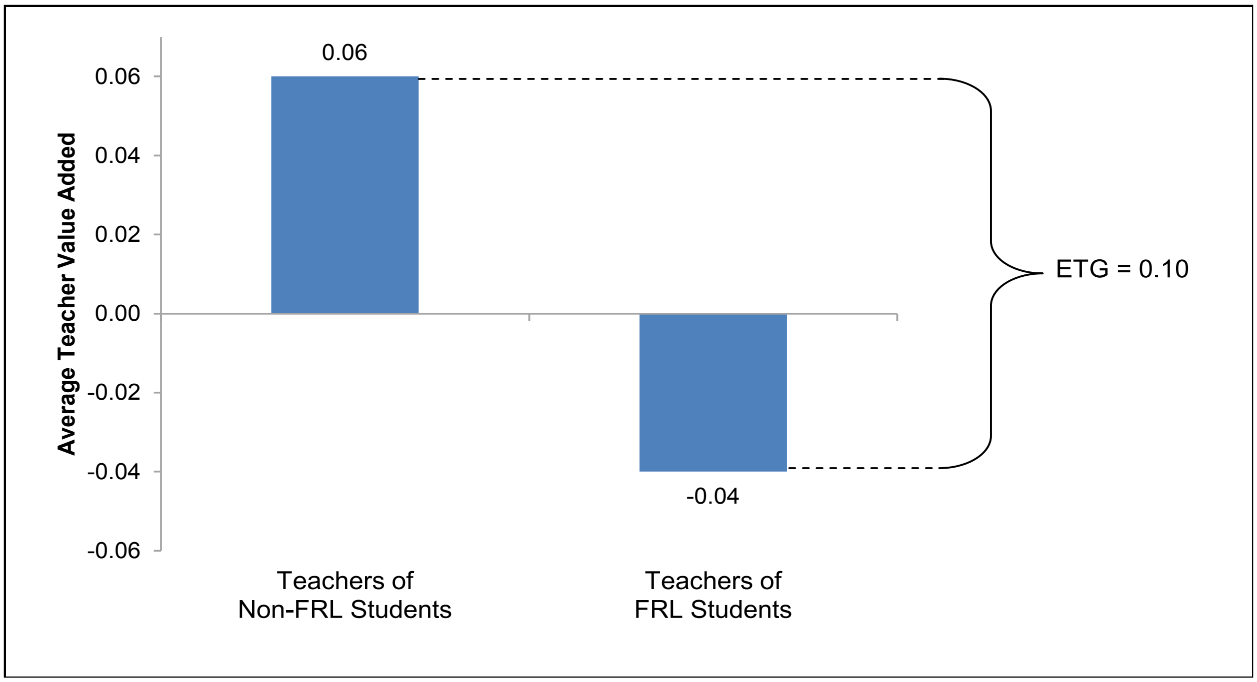
- We calculated a weighted average of value-added estimates for the teachers of all nondisadvantaged (non-FRL) students in the district. We weighted teachers by the number of non-FRL students whom they teach. For example, a teacher with 20 non-FRL students would be weighted twice as heavily as a teacher with 10 non-FRL students in the calculation. In the end, the weighted average represents the value added we would expect for the teacher of a non-FRL student selected at random from the district.
- We repeated this process and calculated a weighted average of the value-added estimates for teachers of disadvantaged (FRL) students in the district.[5] The estimate represents the value added we would expect for the teacher of an FRL student selected at random from the district.
- To calculate the district ETG, we subtracted the average teacher value added for FRL students from the average teacher value added for non-FRL students. Hence, the district ETG is the amount by which we would expect the value added of a typical

[5] Many teachers in a given district would have both FRL and non-FRL students in their classroom. Thus, many teachers were included both in calculating the weighted mean for FRL students and for non-FRL students, although they would likely be weighted differently in the two calculations, based on the number of FRL and non-FRL students that they taught.

non-FRL student's teacher to exceed (or be less than) that of a typical FRL student's teacher. If disadvantaged students have less access to effective teaching, the non-FRL student's teacher would have a higher value-added estimate; thus, the district ETG would be greater than zero.

In Figure II.3, we provide a simple example. In this hypothetical case, teachers of non-FRL students are relatively effective, on average, with a mean value added of 0.06. This indicates that the typical non-FRL student has a teacher who helps that student achieve a test score that is 0.06 standard deviations above what he or she would have achieved with an average teacher. By contrast, the teachers of FRL students tend to be below average, with a mean value added of -0.04. As a result, the district ETG is 0.10 in this example, indicating that FRL students have less access to effective teaching on average. FRL students have teachers who lead to test scores that are 0.10 standard deviations lower than the students would have achieved if they had the same teachers as non-FRL students.

Figure II.3. Less Access for Disadvantaged Students (Hypothetical Example)



The district ETG takes into account both between- and within-school variation in the value added of students' teachers. To better understand the sources of access to effective teaching, we also measure the degree to which differences in access to effective teaching occur between schools (between-school ETG) or within schools (within-school ETG).

To measure only the between-school component of the ETG, we followed three steps to compare effective teaching for non-FRL and FRL students across schools:

- First, we calculated the *average* value added for all teachers within each school, subject, grade, and year, without regard to the FRL status of their students. For simplicity, we refer to this as the "school value-added estimate" (even though it is computed separately for each grade within a school.)

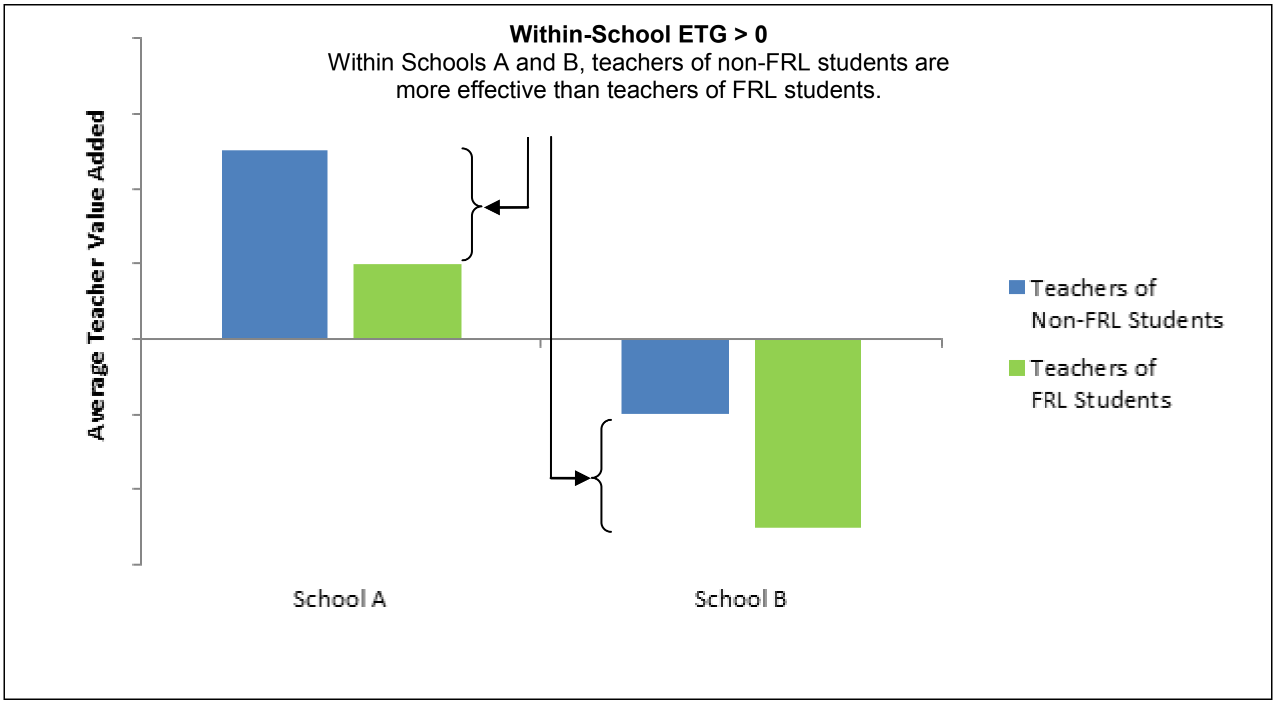
- Second, we assigned each student the average value-added estimate of the teachers in his or her school and grade, linking the non-FRL and FRL students in the same school and grade to the same value-added estimate. By using these school value-added estimates, we measured only the between-school differences in effective teaching.
- Third, we calculated the average school value added for disadvantaged students and the average school value added for nondisadvantaged students.
- Fourth, we subtracted the average school value added of disadvantaged students from the average school value added of nondisadvantaged students. This produced the between-school ETG, which measures only how the sorting of teachers to schools and the selection of schools by students' families relate to gaps in effective teaching.

The within-school ETG is the difference between the district ETG (based on teacher value added) and the between-school ETG (based on school-level value added). Given that the district ETG measures both between- and within-school differences in effective teaching, subtracting the between-school component from the district ETG captures the portion of the ETG attributable to differences in effective teaching within schools.

In Figure II.4, we present a hypothetical example of unequal access to effective teaching within schools. In the figure, non-FRL students have teachers with higher average value added than teachers of FRL students within each school. A situation like this could arise, for example, if FRL students were assigned to classrooms with less effective teachers within schools. In this case, there is unequal access to effective teaching within schools, and the within-school component of the ETG is greater than zero.

Now, as described above, suppose that we assigned FRL and non-FRL students in a school the same school value added, ignoring any differences in teacher value added within schools. This is illustrated in Figure II.5. In this example, non-FRL students attend schools with more effective teaching than FRL students, as would be the case if school A from the previous example (in Figure II.4) consisted primarily of non-FRL students and school B consisted primarily of FRL students. As a result, there is a positive between-school ETG. In other words, FRL students receive less effective teaching because they are disproportionately in schools with lower average teacher value added.

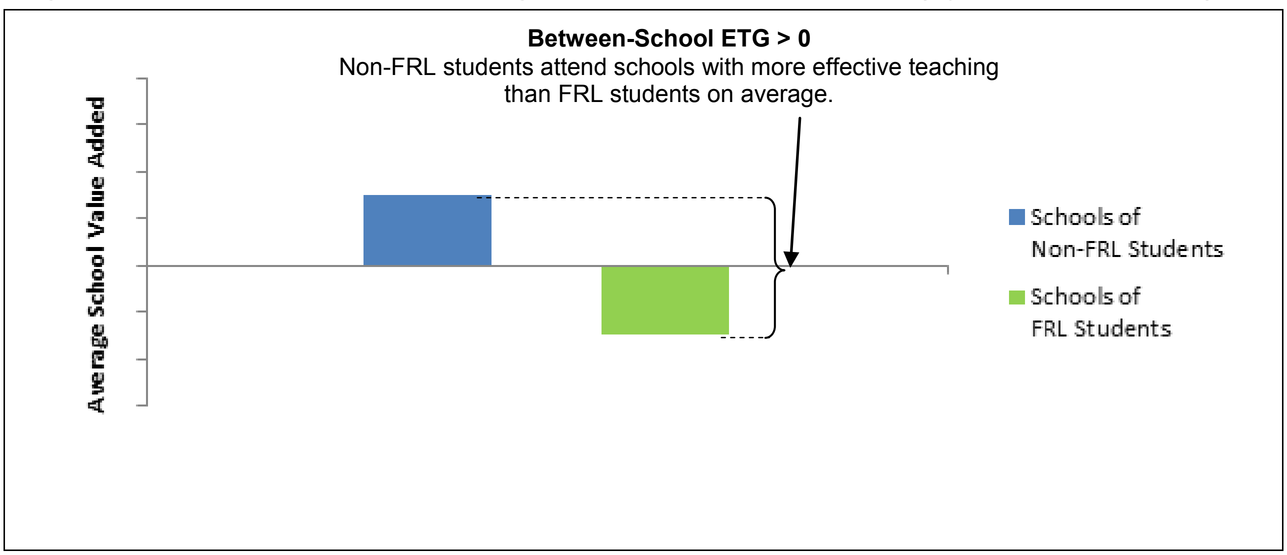
Figure II.4. Less Access for Disadvantaged Students Within Schools (Hypothetical Example)



Note: The figure shows a positive within-school ETG because average teacher value added for non-FRL students is higher than average teacher value added for FRL students within each school.

The between- and within-school ETGs can reinforce or offset one another. For example, they reinforce one another if disadvantaged students attend schools with less effective teaching (a positive between-school ETG, as shown in Figure II.5) and are disproportionately assigned within schools to classrooms with less effective teaching (a positive within-school ETG, as shown in Figure II.4). The between- and within-school ETGs would offset one another, for example, if disadvantaged students attend schools with less effective teaching but are assigned to the best teachers within those schools.

Figure II.5. Less Access for Disadvantaged Students Between Schools (Hypothetical Example)



Note: The figure shows a positive between-school ETG because across schools, the average school value added for non-FRL students is higher than the average value added for FRL students.

3. Measuring Student Disadvantage

We use students' free or reduced-price lunch (FRL) status as our primary measure of student disadvantage. FRL students are defined as disadvantaged, and non-FRL students are defined as nondisadvantaged. FRL eligibility is based on living in a household with an income equal to 185 percent or less of the official poverty line. FRL status is a common indicator of students' socioeconomic status because FRL status is generally available from district administrative data.

Although FRL status provides a way to distinguish between disadvantaged and nondisadvantaged students, it is an imperfect measure of student income for two reasons.[6] First, FRL status is measured with error in both directions. Some eligible students do not apply for or are incorrectly denied the benefit, and some ineligible students receive it. By one estimate, 9.1 percent of FRL students are misclassified in that they are not eligible for the benefit (Ponza et al. 2007). Second, the variation in students' circumstances may be large on either side of the FRL threshold. In particular, for students whose family incomes exceed 185 percent of the poverty line, some may be just above the threshold while others may be substantially above the poverty line. Because FRL status is the key indicator of student disadvantage used in calculating the ETG, misclassification of students could lead to an underestimate of the ETG. In Appendix C, we consider how estimates would change when statistically adjusting for measurement error in FRL status. These sensitivity tests show that accounting for the possibility of misclassification of students' income status with the FRL measure leads to an increase of 0.014 or less in the estimated ETG.

Because of these limitations of FRL status as a measure of student disadvantage, we also estimated an alternative version of the ETG in which we replaced FRL status with indicators for students' race and ethnicity. For districts in which at least 15 percent of the students are White and 15 percent Black, we replaced FRL with indicators of race to measure Black-White differences in access to effective teaching. We excluded Hispanic students from the calculation of this alternative ETG. Likewise, in districts with at least 15 percent White and 15 percent Hispanic students, we calculated Hispanic-White differences in access to effective teaching, excluding Black students from this ETG calculation. We excluded districts with fewer than 15 percent of students from the relevant racial or ethnic categories to avoid imprecise results based on few students in one group.

4. Sensitivity Analyses: Multiple Pre-Tests and Peer Effects

We chose the value added model described above because we believe it provides the best evidence on disadvantaged students' access to effective teaching. Researchers have explored alternative models, such as those that account for two years of student pre-test scores or those that incorporate peer effects, but our main model has several advantages over these alternatives. The alternative model with two years of pre-test scores has more stringent data requirements than our main model. We could estimate this alternative model only using students for whom we have two prior years of test scores. This meant that we could not estimate the model for 4th grade

[6] In addition to the two issues described here, federal regulations that allow some schools to serve free meals to all students complicate the measurement of FRL status in some districts. This issue, along with our strategy for dealing with it, is described in Appendix A.

teachers or for teachers in the first study year. This limits our ability to study access to effective teaching in elementary school grades as well as to examine how access to effective teaching changes from year to year. The main model controls for students' baseline achievement using ELA and math pre-test scores, and allows us to study a broader set of teachers and years.

In the case of the peer effects model, we have greater confidence in the quality of the administrative data required for the main model than for this alternative model. A peer effects model, for example, requires a "classroom identifier" that identifies each separate classroom taught by each teacher in the sample. The accuracy of administrative data on classroom identifiers is hard to verify, and inaccuracies that may be present in the data can lead to inaccurate estimates from a peer effects model. The main model does not rely on data with classroom identifiers. Second, estimates from the main model are more stable than those from a peer effects model, in the sense that estimates from a peer effects model have been shown to vary greatly depending on what model specification is used (Ballou 2004; Hoxby and Weingarth 2006). Moreover, there is no consensus in the literature on the correct peer effects model specification. The main model is thus based on more reliable data and is less dependent on the model specification. Because its data requirements are not as stringent, the main model can also be estimated in a broader sample of districts and grades.

We explored the sensitivity of our findings to these two alternative specifications, a model with an additional year of baseline data and a peer effects model, in a subset of districts and grades where it was possible to estimate these models.

The first alternative model used two years rather than just one year of pre-test data to better account for pre-existing differences in true levels of student achievement (Rothstein 2009). For example, if students in a 6th-grade teacher's classroom were tracked into this teacher's class from a particularly effective 5th-grade teacher, their pre-test scores from the end of 5th grade might reflect an anomalous single-year spike, leading the value-added model to predict higher scores for the students at the end of 6th grade and placing downward pressure on the 6th-grade teacher's value-added estimate. Accounting for the students' 4th-grade test scores may help mitigate such potential bias.

The second alternative model accounted for peer effects of other students in the classroom by including classroom-level variables such as the mean pre-test score of a student's classmates. These classroom-level variables served two purposes. First, they may have reduced measurement error in pre-test scores and other characteristics (to the extent that students' true background characteristics are related to their peers' average characteristics). Second, the peer variables may have captured the effect that other students in a classroom exert on the achievement of individual students during the year (Hoxby and Weingarth 2006; Sacerdote 2011).

We specified the peer effects model to allow for the possibility that (1) a classroom of students with higher levels of prior achievement affects the performance of a given student, (2) fewer disadvantaged students in a classroom affects that student's performance, and (3) a narrower range of achievement in a given classroom affects the performance of students in that classroom if it is more difficult for a teacher to target instruction in classes with a greater diversity of skill levels. Thus, we estimated a peer effects model that included the set of variables in the main value-added model along with the following classroom-level variables:

- Average classroom pre-test score

- Proportion of FRL students in the classroom
- Standard deviation of the student pre-test scores in the classroom

Unlike estimation of the relationship between individual student characteristics and student achievement, which was based on differences in student achievement across different types of students assigned to the same teacher, estimation of peer effects required multiple classrooms per teacher within a year and/or across years. Thus, the model required a classroom identifier that allowed us to identify multiple classes for individual teachers within a year and/or a unique and consistent teacher identifier to identify classrooms for the same teacher over different years.

Incorporating extra variables to estimate the alternative value-added models limited the grades and/or years of data that we could use. For the model with multiple years of pre-test data, we could not estimate results for grade 4 or for any grades from the first year of the study, because only one year of pre-test data was available for 4th graders and all students in year 1. Therefore, we limited comparisons between models with a single year of pre-test data and multiple years of pre-test data to grades 5 through 8 in years 2 and 3 of the study. For the peer effects model, the need to link teacher identifiers either within or across years also prevented us from estimating the model in several districts, especially for elementary school grades. We discuss these limitations in greater detail in the next section.

B. Data to Measure Value Added, the Effective Teaching Gap, and District Context

We asked study districts to provide administrative data that would allow us to estimate value-added models at the teacher level for the 2008–2009 through 2010–2011 school years. Each district provided standardized student test scores, student background characteristics, school enrollment data, teacher-student-course links, and unique teacher and student identifiers that could be used to link students and teachers across years or different types of data. We used the teacher-student links and student demographic data to link disadvantaged and nondisadvantaged students to their teachers in order to calculate the ETG. Given that some districts could not provide these data for all years, we describe below how some districts were excluded from the analysis for certain years when they lacked sufficient data. We also interviewed district staff to better understand relevant district contextual information.

Standardized test scores. We obtained state assessment results for grades 3 through 8 for four consecutive school years. Student test score data from the first school year (usually the 2007–08 school year) served as a pre-test for the first year for which we have data linking teachers to students (usually the 2008–2009 school year) and thus the first year we measure teacher value added and the ETG. Because there were no pre-test scores for 3rd-grade students, 4th grade is the first grade level at which we were able to calculate teacher value added.

We collected data directly from 24 of the districts and from state longitudinal databases for the other 5 districts. Given that annually updated data from the state databases were not available in time to include data from the 2010–2011 school year, we included data that lag by one year for these 5 districts. Thus, for 24 of the districts, we report results from the 2008–2009 through 2010–2011 school years, and, for the other 5 districts, we report results from the 2007–2008 through 2009–2010 school years. Throughout this report, we refer to years 1 through 3 rather than referring to the actual school years.

Student demographic characteristics. Districts provided background data on the following student characteristics in years 1 through 3: FRL status, limited English proficiency status, special education status, gender, race, and ethnicity.

Teacher-student-course links. A teacher-level value-added model requires data linking each student to the teacher responsible for teaching him or her ELA and/or math. Districts provided teacher-student linked data that identified the teacher who taught each student ELA and/or math. When possible, we also used these data to determine the amount of time a teacher spent with each student.

When districts could not provide teacher-student links for certain grade levels or school years, we excluded those grades or years from the study. All districts provided teacher-student links for the 2009–2010 and 2010–2011 school years for grades 4 and 5 and/or grades 6 to 8.[7] Four districts were not able to provide teacher-student links for the 2008–2009 school year, and one of these districts also did not have teacher-student links for middle school teachers in 2009-10 (see the Main Model in Table II.2). In 15 districts, the teacher-student links for upper elementary grades (grades 4 and 5) captured self-contained classrooms in which one teacher was responsible for ELA and math instruction, but they did not account for the possibility that students in these grades received ELA and math instruction from different teachers. As a result, it is possible that the teacher linked to a student in the data may not have taught both ELA and math.[8]

There are two additional data requirements for a peer effects model: (1) classroom identifiers that allow us to construct information about classroom peers for each student and (2) unique teacher identifiers that can link teachers across school years. We excluded three districts from the peer effects model for elementary and middle school grades because they could not provide a reliable classroom identifier and three additional districts because the state assessment includes end-of-course rather than end-of-grade math tests, which prevented us from calculating a pre-test average measured consistently across all classrooms. This resulted in a total of 23 districts for the peer effects model in middle school grades (although the sample size is 21 in year 1 because two of these districts did not have teacher-student links for year 1). For the elementary grades, we also excluded from the elementary school peer effects models the districts that provided only students' homeroom links at the elementary school level because homeroom links do not ensure that the homeroom teacher instructs students in both subjects. This excluded 14 of the remaining 23 districts for which we could estimate the peer effects model at the middle

[7] Students in grades 6-8 are always included with middle school grades even if some of these students were taught in elementary school settings, such as a school that includes grades kindergarten through 8. According to the 2008-2009 Common Core of Data, on average across the study districts, 14 percent of students in grades 4-8 attended a school that included all five of these grades. Among sixth graders, 20 percent attended a school that included grades 4 and/or 5 (but not 7 and 8), 56 percent attended a school that included grades 7 and 8 (but not 4 or 5), and 23 percent attended a school that included at least one upper elementary grade and at least one middle school grade.

[8] In this case, we can identify the group of students that take ELA and math together in the same class but we do not know the identity of the teacher. Despite this, we can still estimate value added for this unknown teacher. If a teacher has multiple departmentalized sections, we will generate multiple value-added estimates for this teacher. This is not ideal, as each estimate will be less precise than a combined estimate across several classes. However, using multiple classroom-level estimates of a teacher's value added will not cause the calculation of the effective teaching gap to be biased.

school level, leaving 9 districts in which we could estimate the peer effects model at the elementary school level.

Taking into account the fact that some districts lacked data of one type or another, we show in Table II.2 the number of districts that we could include in each value-added analysis for the upper elementary and middle school grades.

Table II.2. Number of Districts Included in Each Value-Added Model

	Year 1		Year 2		Year 3	
	Elementary	Middle	Elementary	Middle	Elementary	Middle
Main Model	25	25	29	28	29	29
Multiyear Pre-test Model	0	0	28	27	29	29
Peer Effects Model	9	21	9	23	9	23

School enrollment data. We collected data on students' enrollment patterns in schools. Districts typically stored school enrollment data separately from the teacher-student links, and these data provided information about each school a student attended during the school year and the length of time the student was enrolled in the school. We used the school enrollment data to define the length of time a student was assigned to a teacher when the teacher-student links did not provide this information.

District interview data. We conducted telephone interviews with district staff between September 2011 and January 2012. We asked district staff about their perspectives on the extent to which they viewed access to effective teaching as a policy priority. In addition, we asked them about policies and practices that might improve effective teaching by (1) recruiting more effective teachers into high-need schools, (2) attracting more effective teachers to transfer into these schools, or (3) improving the effectiveness of teachers who are retained in these schools. Specifically, we obtained information on the following 12 policies: targeted teacher recruitment activities for high-need schools, highly selective teacher recruitment programs, bonuses for teaching in high-need schools, performance pay in high-need schools, principal choice in selecting involuntary transfer teachers, early teacher hiring timelines for high need schools, whether teacher performance is considered when laying off teachers, teacher performance for tenure decisions, principal quality initiatives in high-need schools, comprehensive teacher induction, teacher professional development in high-need schools, and school turnaround activities.

III. STUDY SAMPLE

To document access to effective teaching in a diverse set of districts, the recruitment and selection of districts focused on obtaining a geographically diverse sample that could provide the data needed for a value-added analysis of teachers. We also sought variation in the types of policies implemented. In this chapter, we describe the selection of districts, the criteria used to identify districts asked to participate in the study, and the characteristics of the resulting sample.

A. Selection of Study Districts

The selection of districts took place in three stages: (1) selecting an initial group of school districts for recruitment, (2) recruiting the districts to participate in the study, and (3) selecting a final list of study districts from among those recruited. We selected the initial group of districts for recruitment by identifying the largest districts, in terms of student population, within each of the four U.S. Census regions. We focused on the largest districts because they were more likely to have the quality and quantity of data needed for a value-added analysis and for examining access to effective teaching.

We sought geographic diversity, but a nationally representative sample was not feasible because we required districts to provide data that (1) linked students to the teachers responsible for their math and ELA instruction and that (2) included unique identifiers that remained consistent over time for teachers and students. We focused on districts that had developed this capability by 2008–2009 so that we could estimate teacher value-added measures beginning in that school year.

Since the study measures access to effective teaching by comparing effective teaching for FRL and non-FRL students, we sought districts with a mix of both types of students. The ETG is less relevant if nearly all students are FRL or non-FRL, and it is more difficult to obtain a statistically reliable estimate of the ETG if the sample size of one group is small. To ensure meaningful comparisons, we prioritized districts with at least 20 percent and less than 80 percent of students who receive FRL, according to the 2008–2009 Common Core of Data. We focused on FRL status because it is a measure of student disadvantage that is consistently available across districts nationally.

We also sought variation across districts in the types of policies that could influence ETGs, focusing on the following types of policies: (1) teacher compensation programs that offer additional pay for performance or for teaching in a high-need school; (2) recruitment of teachers for high-need schools through programs such as Teach For America, Teaching Fellows, or teacher residency programs; and (3) teacher transfer policies that provide principals with choice in selecting a voluntary or involuntary transfer.

The initial recruitment list consisted of 160 districts representing the largest districts within each region of the country. We actively recruited 104 of these districts, and 36 districts both appeared to have the data systems appropriate for the study and agreed to participate. The most common reasons provided by district contacts for not agreeing to participate were (1) inadequate staff or resources to provide data, often due to budget constraints, and (2) "too much else going on in the district." We selected 30 districts that were spread across states and geographic regions and that were missing the least amount of data needed for the study. We eventually excluded one additional district due to the quality of its FRL data, resulting in a final sample of 29 districts.

B. Characteristics of Study Districts

We examine the characteristics of study districts and compare them to all school districts nationally and to the 100 largest districts across the country. Comparisons are based on data from the Common Core of Data (CCD) for the 2008–2009 school year, the first year for which we are calculating ETGs.

Geographically diverse sample of districts. In school districts across the four U.S. Census regions, different historical and political environments may affect the size of the ETG. For example, in the South, teacher unions are less likely to be allowed to engage in collective bargaining, which can affect teacher compensation, tenure, transfer, and layoff policies. In addition, districts in the South and West tend to be larger than those in the North or Midwest. As mentioned in Chapter I, earlier studies have used data from three Southern states (Florida, North Carolina, and Tennessee), one Western district (Los Angeles), and an undisclosed sample of 10 districts. By recruiting districts from all regions, we are able to measure access to effective teaching in a broader set of policy environments.

The selection protocol resulted in a geographically diverse sample, with at least four districts from each region of the country (Table III.1). The geographic distribution of students in the study sample is within 2 percentage points of the national distribution for the South and Midwest. Northern students are underrepresented in the sample (8 percent of students in the study compared to 16 percent nationally), and students from the West are overrepresented (34 percent compared to 24 percent nationally). Students from districts in the South make up the largest portion of the sample, and students in the North represent the smallest proportion, consistent with the national distribution of students. The geographic spread of study districts is more comparable to the national distribution than to the 100 largest districts, which are primarily located in the South and West.

Table III.1. Regional Distribution of Students Enrolled in All Districts, 100 Largest Districts, and Study Districts

	All Districts in the United States	100 Largest Districts in the United States	Participating Districts
Midwest	22%	8%	21%
North	16%	11%	8%
South	38%	56%	36%
West	24%	26%	34%

Source: 2008–2009 Common Core of Data.

Study districts comparable in size to 100 largest U.S. districts. The median study district is larger than the median district across the country. The median U.S. district enrolled approximately 1,000 students compared to the median study district that enrolled about 60,000 students (Table III.2).[9] Ninety percent of study districts had 25,000 students or more, compared to 2 percent of districts nationally (although this 2 percent enrolled 35 percent of the country's students) (Figure III.1). Study districts are comparable in size to the 100 largest districts in the country, which had a median enrollment of about 70,000. Twenty-one percent of study districts had 100,000 or more students, compared to 25 percent of the 100 largest districts. The size of study districts varied from a minimum of just over 20,000 students to more than 100,000 students. This distribution of district size mirrors that from some of the prior studies based on state data, although our study includes 10 districts with fewer than 50,000 students, which is smaller than any of the 10 districts examined in Glazerman and Max (2010).

Table III.2. Comparison of Study Districts to All Districts and Largest Districts in the U.S.

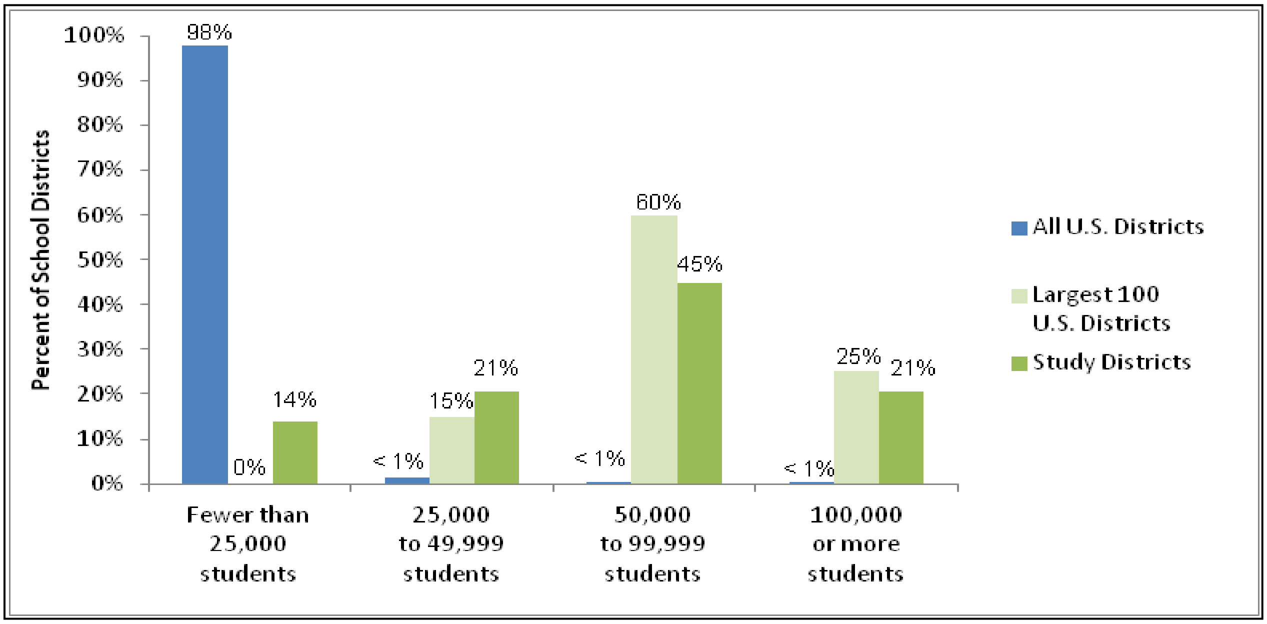
	All Districts in the U.S.	100 Largest Districts in the U.S.	Study Districts
District Enrollment (district median)	1,000	70,000	60,000
Percent of students in large city (percent of students)	14%	46%	69%
FRL (percent of students)	44%	53%	63%
Student Race and Ethnicity (percent of students)			
Percent White	55%	31%	23%
Percent Black	16%	27%	31%
Percent Hispanic	21%	34%	40%
English Language Learners (percent of students)	9%	13%	18%
Separation of Students by FRL Status (D-Index)			
Elementary Schools	0.27	0.40	0.47
Middle Schools	0.17	0.31	0.34
Number of Districts	13,437	100	29

Source: 2008–2009 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. All characteristics in this table are defined by the Common Core of Data, including the definition of "large city," student race and ethnicity, and the determination of whether a school is an elementary school or middle school. We describe the extent to which FRL and non-FRL students are separated in different schools using a measure known as the Index of Dissimilarity (D-Index). This measure can be interpreted as the percentage of students from one group (FRL or non-FRL) who would have to change schools to achieve a perfectly even distribution.

[9] District enrollment is rounded to the nearest 10,000 to maintain district confidentiality.

Figure III.1. Distribution of Study Districts, U.S. Districts, and Largest 100 U.S. Districts by Size



Source: 2008–2009 Common Core of Data.

Study districts predominantly in urban areas. The study districts are a mix of large city, medium-sized city, and countywide districts. We define urbanicity based on the proportion of students in grades 4 through 8 who attend schools in a large city with a population of 250,000 or more. The percentage of students in large cities is 69 percent across the study districts compared to 46 percent of students in the 100 largest districts and 14 percent of students in all U.S. districts. Among study districts, the percentage varies from 0 to 100 percent. Sixteen of the 29 study districts have more than 75 percent of students in large cities.

Study districts high in poverty and proportion minority. To obtain precise estimates of the ETG, we sought to avoid districts with a relatively small proportion of FRL or non-FRL students, and therefore emphasized economic diversity in the selection process. According to the 2008–2009 CCD, the average study district had an FRL rate of 63 percent (with a range of 34 to 78 percent). This is 19 percentage points higher than the average district nationally and 10 percentage points higher than the 100 largest districts in the country (Table III.2).[10]

[10] The average percentage of FRL students in study districts based on the administrative data used for the analysis in this study was larger than the corresponding percentage based on CCD data. The average FRL rate for our analysis sample was 66 percent, which is higher than the CCD's 63 percent. FRL rates were within the original target range of 20 to 80 percent for 23 of the 29 districts; 6 districts had between 82 and 92 percent FRL. The district with 92 percent FRL provided an indicator of what they labeled "economic disadvantage" that was distinct from the indicator of FRL they provided. We concluded that the students who were "economically disadvantaged" were free lunch eligible, even though many students in this group were not coded as being FRL. We assigned these students to be FRL, resulting in a higher FRL rate than reported in the CCD.

Differences in FRL rates between the two data sources may be attributable to the CCD data's inclusion of all students in the district, whereas the data for our analysis contain only students in grades 4 through 8. For example, the FRL rate for study districts is 68 percent using CCD data when we include only elementary and middle schools, compared to the 66 percent in our analysis sample.

Study districts have a higher proportion of minority students and English language learners (ELLs) than districts nationally. Thirty-one percent of students in study districts are Black and 40 percent are Hispanic; these rates are roughly double the percentages for students nationally. The racial and ethnic make-up of study districts is more similar to the 100 largest districts in the country, where 61 percent of students are Black or Hispanic. The proportion of ELLs in study districts is 18 percent compared to 9 percent nationally and 13 percent in the 100 largest districts.

Greater separation of students by FRL status in study districts than average U.S. district. The extent to which disadvantaged students are distributed across schools within districts and across classrooms within schools can affect the ETG. For example, if FRL and non-FRL students were equally distributed between and within schools, unequal access to effective teaching could not exist because both types of students would experience the same level of effective teaching. We examined the extent to which FRL and non-FRL students are separated in different schools using a measure known as the Index of Dissimilarity (D-Index) that can be interpreted as the percentage of students from one group (FRL or non-FRL) who would have to change schools to achieve a perfectly even distribution.[11]

Table III.2 shows that the level of student separation by FRL status in study districts is higher than that of the average district nationally and closer to that of the 100 largest districts. In the average study district, 47 percent of FRL or non-FRL students would need to change elementary schools to achieve a perfectly even distribution compared to 40 percent of students in the largest 100 districts, and 27 percent of students nationally. At the middle school level, 34 percent of FRL or non-FRL students would need to change schools to achieve a perfectly even distribution in study districts, compared to 31 percent in the 100 largest districts, and 17 percent nationally.

Separation of students by FRL status is greater across schools than within schools. We used data from study districts to examine the extent to which students were separated by FRL status within schools—that is, the extent to which FRL students were assigned to teachers with other FRL students, and non-FRL students were assigned to teachers with other non-FRL students. Students may be sorted by FRL status within schools due to tracking policies that assign students to different classrooms based on their prior achievement, although these policies tend to be more common in the middle school grades (Loveless 2009). This type of sorting affects the potential for unequal access to effective teaching within schools. To compare the extent to which students are separated by FRL status within and between schools, we documented how much of the variation in FRL status in a district is due to variation within schools (between teachers) and between schools. The amount of separation by FRL status is greater between schools than within schools: no more than 4 percent of the variation in FRL status occurred between teachers within schools, while between 17 and 25 percent occurred between schools (see Appendix A for details).[12]

---

[11] The D-index is calculated as $D = \frac{1}{2}\sum_{j=1}^{N}\left|p_j^{FRL} - p_j^{nonFRL}\right|$, where N is the number of teachers in the district, p_j^{FRL} is the proportion of the district's FRL-eligible population with teacher j, and p_j^{nonFRL} is the proportion of the district's FRL-ineligible population with teacher j.

[12] The remaining portion of variation is due to the variation in FRL status for individual teachers within a school.

The median family income of non-FRL students is $54,000 higher than that of FRL students on average in study districts. Although we use FRL eligibility to define students as disadvantaged and nondisadvantaged, differences in family income for FRL and non-FRL students may vary across districts. If so, this could contribute to differences in access to effective teaching across districts. For example, one might expect larger differences in effective teaching between FRL and non-FRL students in districts with relatively large differences between the two groups in family income than in those with smaller differences. To understand the extent to which FRL status provided a consistent measure of student disadvantage across districts, we measured differences in median income for families with children that are below the threshold for FRL eligibility (185 percent of the poverty level) and similar families that are above the FRL threshold.[13] Although these estimates include some families with children who do not attend public schools, they provide a measure of the differences in family income between FRL and non-FRL students.

In the average district, the median family income for non-FRL students is $54,000 higher per year than the median family income for FRL students, as median family income is $16,000 for FRL students compared to $70,000 for non-FRL students. While there is some variation across districts in the difference in median income between FRL and non-FRL students, this difference is between $45,000 and $65,000 for 23 of the 29 districts and between $35,000 and $85,000 across all study districts.

Below-average student achievement in study districts. We compared student achievement in study districts to the achievement level for all students in their respective states. Student achievement is measured in percentiles of student performance on the state assessment in ELA and math. The 50th percentile represents average achievement in the respective state for each district. Students in study districts were lower performing on average than the peers in their respective states, with performance at the 45th percentile in ELA and at the 46th percentile in math (Table III.3). However, the level of ELA and math achievement varies widely across study districts, from a minimum of the 25th percentile to a maximum of the 67th percentile.

Table III.3. Average Student Achievement and Student Achievement Gaps in Study Districts, Percentiles of Student Achievement

	Mean	Standard Deviation	Minimum	Maximum
Average Student Achievement				
English/Language Arts	45	11	26	63
Math	46	12	25	67
Average Student Achievement Gap				
English/Language Arts	28	7	10	42
Math	26	7	10	36

Source: District administrative data.

Note: Results are for 29 districts, grades 4 through 8, and years 1 through 3. Student achievement and student achievement gaps within a district are weighted across grades and years by the number of students, and then averaged with equal weight across districts.

[13] To approximate differences in median family income for FRL and non-FRL families, we used family income data for study districts from the American Community Survey for the 2008 through 2010 years. We used income estimates based on samples of families with children under 18 years old, and we defined FRL-eligible families as those with incomes less than $30,000.

Presence of student achievement gaps by FRL status. The gap in student achievement between FRL and non-FRL students documents the magnitude of inequitable outcomes between these two groups, which may result from unequal access to effective teaching in grades 4 through 8, unequal access in the early years of schooling (kindergarten through third grade), and other sources of inequity, including differences in resources at home. Student achievement gaps are reported in terms of percentiles of student achievement. Table III.3 shows that FRL students have lower average achievement than non-FRL students in all of the study districts, with average student achievement gaps of 28 percentile points in ELA and 26 percentile points in math for grades 4 through 8. Student achievement gaps in study districts range from 10 percentile points to 42 percentile points. We also examined the student achievement gap by grade level and found that the gap did not significantly differ across grade levels (see Appendix Table C.1).

Districts implemented policies relevant for access to effective teaching. We provide information from district interviews about the policies implemented in study districts in 2010-11 as context for the ETG findings in this report. A majority of our districts (17 of the 29) described equitable access to effective teaching as a policy priority. This is not surprising given our district selection process. However, most districts (22/29) reported they had not used data on teacher effectiveness, such as value added measures or teacher evaluation ratings, to assess access to effective teaching.[14] Of the 12 policies that we asked about, the most common were in the areas of school improvement and teacher development policies. At least half of the study districts reported using these policies.

[14] Since all states are required by the Elementary and Secondary Education Act to report information on the percentage of highly qualified teachers (HQT) in high- and low-poverty schools, we did not count this when identifying districts that have measured access to effective teaching. National data suggest a high proportion of teachers meet HQT requirements, with minimal differences for high- and low-poverty schools. Data from the U.S. Department of Education (2011) show that, nationally, 97 percent of core courses are taught by teachers meeting the HQT requirements, including 96 percent in high-poverty schools and 97 percent in low-poverty schools.

IV. ACCESS TO EFFECTIVE TEACHING

In this chapter, we analyze whether disadvantaged students have equal access to effective teaching within districts by documenting the size of the ETG. We examine the district ETG in 29 study districts for grades 4 through 8 separately by ELA and math. In addition, we show the extent to which the ETGs occur between and within schools. Finally, we examine how access to effective teaching varies by race and ethnicity, test the sensitivity of the ETG results to alternative models, and show how district characteristics are related to differences in the ETGs across districts.[15]

A. Effective Teaching Gaps

1. Inequitable Access to Effective Teaching Contributes Two Percentile Points to the Difference in Student Achievement Between FRL and Non-FRL Students in a Given Year

Disadvantaged (FRL) students do not have equal access to effective teaching on average in the 29 study districts and three school years. Teachers of FRL students have lower value added than teachers of non-FRL students on average, with statistically significant differences of 0.034 standard deviations of student test scores in ELA and 0.024 standard deviations in math (Table IV.1). In other words, the typical FRL student experiences less effective teaching than the typical non-FRL student within a district.

We considered how eliminating any differences in access to effective teaching might affect the student achievement gap. Suppose, for example, that a district improved teacher effectiveness in high-need schools and/or principals changed how they assign teachers to classrooms. As shown in Chapter III, during the three years we examined, the average difference in student achievement between FRL and non-FRL students in the study districts was 28 percentile points in ELA and 26 percentile points in math in grades 4 through 8. We estimate that reducing the ETG to zero for one year—in other words, providing equal access to effective teaching for FRL and non-FRL students—would decrease this difference from 28 percentile points to 26 percentile points in ELA and from 26 percentile points to 24 percentile points in math.[16] This suggests that unequal access to effective teaching contributes to differences in student achievement between FRL and non-FRL students, but it is not the only factor—pre-existing achievement differences and other disadvantages facing FRL students also contribute.

---

[15] Additional information is available in the appendices. Statistical details of the methods used for calculating value-added estimates, calculating the ETG, and relating district characteristics to ETGs are contained in Appendix A. Diagnostic information on the value-added estimates is in Appendix B. Extra tables related to this chapter and detailed results from the sensitivity analyses are in Appendix C, including the number of districts with significant ETGs across years, ETG results by grade and grade span, and results from the sensitivity analyses.

[16] We calculated the average student achievement gap in terms of standard deviations of student test scores, and then translated the difference between FRL and non-FRL students into percentile terms using a cumulative normal distribution function. To simulate changes in the student achievement gap if the ETG were zero, we assumed that the test scores of FRL and non-FRL students would converge by the amount of the ETG. We then used the same methodology to translate this newly calculated hypothetical student achievement gap under an ETG of zero into percentiles.

Table IV.1. Effective Teaching Gap in Study Districts, 29-District Average, Years 1 to 3

	Effective Teaching Gap	
	English/Language Arts	Math
Average	0.034*	0.024*
Standard Deviation	0.023	0.024
Minimum	0.007	-0.019
Maximum	0.106	0.081

Source: District administrative data.

Note: Results are from the main model for 29 districts, grades 4 through 8, and years 1 to 3. District-level results are weighted across grades and years by the number of teachers. Overall results are weighted equally across districts.

*Indicates statistical significance at the 0.05 level, two-tailed test. Statistical significance is based on variation across districts.

To better understand the magnitude of the district ETG, we consider two additional ways to think about these ETG estimates. One way to assess the magnitude of the average ETG is to compare how the average teacher of a nondisadvantaged student compares to the average teacher of a disadvantaged student in terms of percentiles of the teacher distribution. For ELA, the average teacher of nondisadvantaged students is at the 56th percentile in the teacher distribution, compared to the average teacher of disadvantaged students at the 47th percentile. In math, it is the difference between a teacher at the 53rd percentile and a teacher at the 48th percentile.

Another way to examine the magnitude of the ETG is to compare it to the difference in value added between a novice and a more experienced teacher. Kane et al. (2006) found a difference of 0.035 standard deviations between the value added of ELA teachers in their first and fourth year of teaching, using a sample of grade 4 through 8 teachers in New York City. The difference in value added for math teachers in their first and fourth year was 0.073 standard deviations.[17] So our finding of an ETG of 0.034 in ELA indicates that the average difference between the teaching experienced by FRL and non-FRL students is about the same as the average difference in the effectiveness of a first-year teacher compared to that of a fourth-year teacher. In math, the ETG of 0.024 is only one-third as large as the difference between a first- and fourth-year teacher.[18]

[17] In a follow-up brief, we will examine the relationships between the ETG and teacher mobility in study districts, including how the ETG relates to the likelihood that FRL and non-FRL students are assigned to novice teachers.

[18] We also examined whether the amount of variation in teacher value added differed for disadvantaged and nondisadvantaged students. The standard deviation of value-added estimates for teachers of disadvantaged students is 0.13 in ELA and 0.21 in math, compared to 0.11 in ELA and 0.19 in math for teachers of nondisadvantaged students.

Access to effective teaching varies across districts. Access to effective teaching ranged from districts with approximately equal access (ETGs that are not significantly different from zero) to districts with ETGs as large as 0.106 in ELA and 0.081 in math (Table IV.1). Significance tests confirm that in both subjects, variation in the ETGs across study districts is greater than would be expected to occur by chance. Figures IV.1 and IV.2 show the estimated ETGs by district, with a bar representing the ETG for each district.[19] The length of the bar indicates a 95 percent confidence interval around a point estimate of the ETG that is shown by a diamond in the middle of the bar. ETGs whose confidence intervals include zero are not statistically significant.

Figures IV.1 and IV.2 show that disadvantaged students, on average, experience less effective teaching in 27 of 29 districts in ELA and 19 of 29 districts in math. In the remaining districts, disadvantaged students have equal access to effective teaching, as shown by ETGs that are not significantly different from zero. In none of the 29 districts was the ETG negative and statistically significant. That is, the ETG never indicated greater access to effective teaching for FRL students.

Access to effective teaching is stable over three years. We find no evidence that access to effective teaching varied over the three years of our study. There are no statistically significant differences in the average ETG across the three years of the study for either subject, as shown in Table IV.2.[20] Given the stability in the ETG results across the first three years of the study, we focus in the remainder of this chapter on results that are averaged across the three years.

Table IV.2. Effective Teaching Gap by Study Year, 29-District Average, Years 1 to 3

	Effective Teaching Gap	
Year	English/Language Arts	Math
1	0.034*	0.026*
2	0.033*	0.023*
3	0.033*	0.023*
Three-Year Average	0.034*	0.024*

Source: District administrative data.

Note: Results are for grades 4 through 8 for 25 districts in year 1, 28 districts in year 2, and 29 districts in year 3. District-level results are weighted across grades and years by the number of students. Overall results are weighted equally across districts.

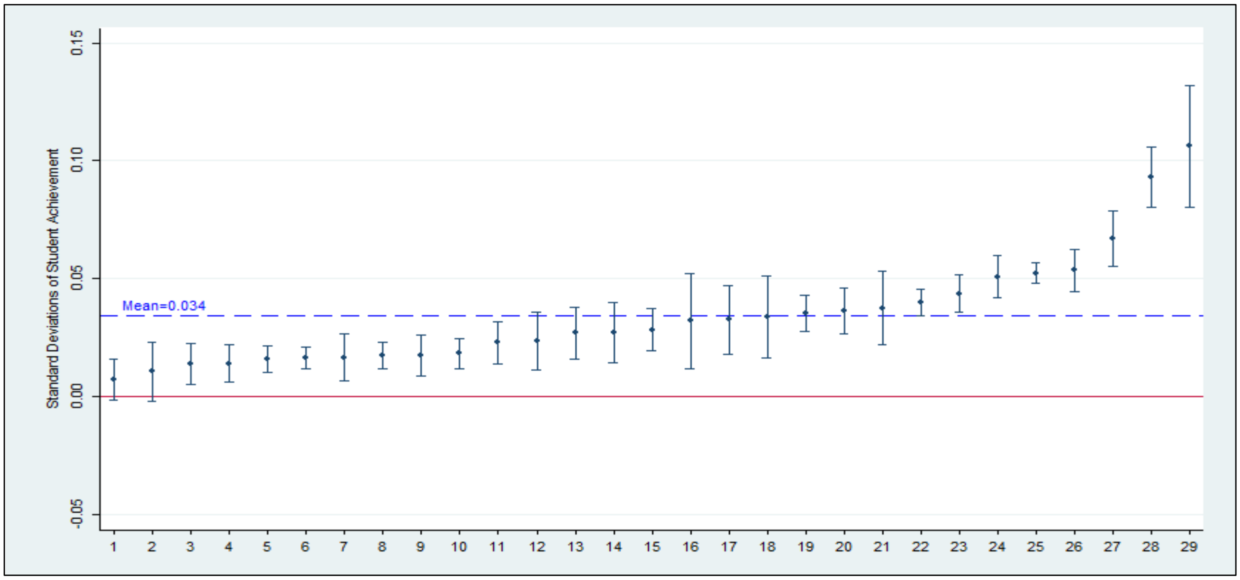
Differences across years are not statistically significant in either subject. We average the year-to-year differences for individual districts and use the estimated variation within districts to test the significance.

*Indicates statistical significance at the 0.05 level, two-tailed test. Statistical significance is based on variation across districts.

[19] The bars in Figures IV.1 and IV.2 are arranged from lowest to highest ETG, and each district is assigned an identification number that is consistent across figures. Identification numbers are assigned based on Figure IV.1, which shows ETGs for ELA. The district with the lowest ELA ETG is assigned a "1," the second lowest "2," and so on.

[20] Another way to gauge year-to-year stability of the results is to examine whether the number of districts that have significantly positive ETGs, statistically insignificant ETGs, and significantly negative ETGs changes from year to year. We find little change over time in the number of study districts in each of these categories. See Table C.1 in Appendix C for more details.

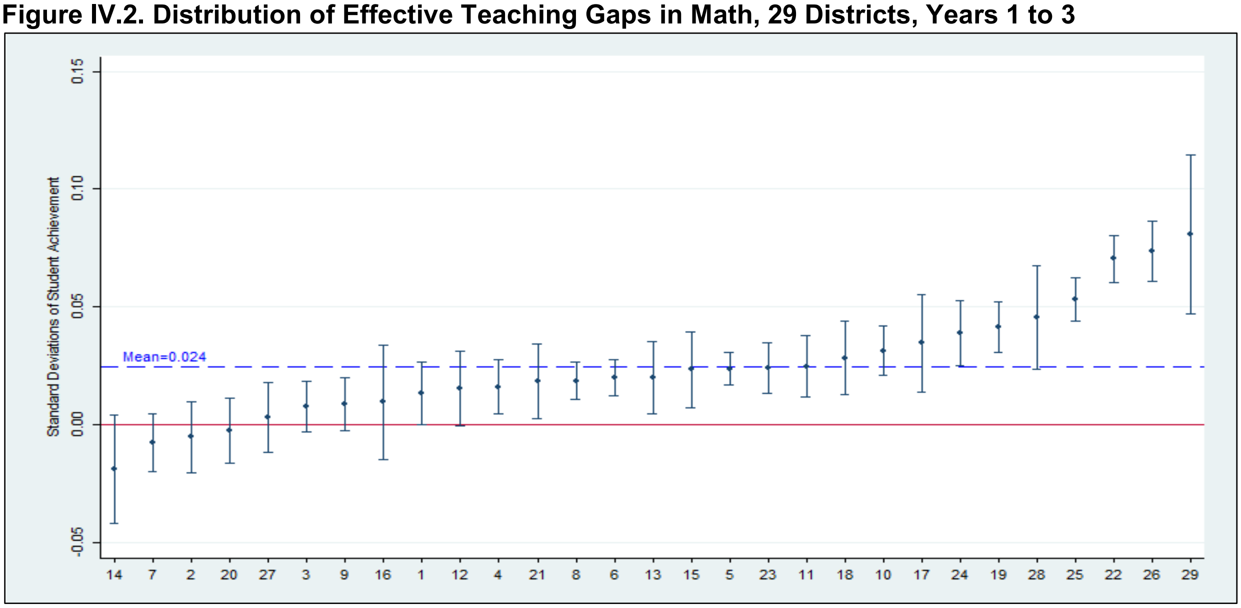
Figure IV.1. Distribution of Effective Teaching Gaps in English/Language Arts, 29 Districts, Years 1 to 3



Source: District administrative data.

Note: Results are for 29 districts, grades 4 through 8, and years 1 to 3. District-level results are weighted across grades and years by the number of students. Circles represent the district-level ETGs and the vertical lines show the 95 percent confidence intervals around each point. ETGs whose confidence intervals include zero are not statistically significant. The cross-district average of 0.034 student standard deviations is shown by the dashed horizontal line. Districts are ordered by the size of the ETG. District codes 1 to 29 are assigned in order of the size of the district ETG in English/language arts and are consistent across figures to permit comparison across content areas. The ETG is expressed in terms of standard deviations of student test scores.

Figure IV.2. Distribution of Effective Teaching Gaps in Math, 29 Districts, Years 1 to 3



Source: District administrative data.

Note: Results are for 29 districts, grades 4 through 8, and years 1 to 3. District-level results are weighted across grades and years by the number of students. Circles represent the district-level ETGs and the vertical lines show the 95 percent confidence intervals around each point. ETGs whose confidence intervals include zero are not statistically significant. The cross-district average of 0.024 student standard deviations is shown by the dashed horizontal line. Districts are ordered by the size of the ETG. District codes 1 to 29 are assigned in order of the size of the district ETG in English/language arts and are consistent across figures to permit comparison across content areas. The ETG is expressed in terms of standard deviations of student test scores.

Access to effective teaching in ELA is related to access in math. Districts with larger ETGs for ELA also have larger ETGs for math. ETGs in ELA and math in the same district have a positive and statistically significant correlation of 0.61.

2. Differences in Access to Effective Teaching Are Larger Between Schools Than Within Schools

We separated the ETG into the between-school and within-school components to understand the extent to which unequal access to effective teaching is due to the sorting of students and teachers across schools or to the assignment of students and teachers within schools. The between-school ETG indicates the extent to which there are differences in access to effective teaching between schools. In other words, it measures the degree to which FRL students attend schools with more or less effective teaching on average than non-FRL students. The within-school ETG indicates the extent to which the two groups have differential access to effective teaching within schools.

In all cases, differences in access to effective teaching between FRL and non-FRL students are larger between schools than within schools. These differences are especially pronounced in ELA in the upper elementary grades (4 and 5). As shown in the top panel of Table IV.3, the average between-school ETG for upper elementary grades is 0.035, significantly larger than the average within-school ETG of 0.005. In other words, unequal access to effective teaching depends more on FRL students attending *schools* with less effective teaching than on FRL students being assigned to *classrooms* (within schools) with less effective teaching. This means that the sorting of teachers and students across schools accounts for more of the district ETG in the elementary school grades in ELA than the assignment of teachers to students within schools.[21]

Between- and within-school ETGs in ELA are shown for elementary and middle schools in Figures IV.3 and IV.4. Districts are ordered from lowest to highest by the district ETG in ELA. Between-school ETGs are shown in green (light shading); within-school ETGs are shown in blue (dark shading). If both are positive, between-school ETGs are shown on top of within-school ETGs. Since the district ETG equals the sum of the two parts, the overall height of both bars stacked on top of each other is the district ETG for a district. Statistically significant results are indicated by solid bars; results that are not significantly different from zero are shown with hollow bars.

Figure IV.3 shows more districts with positive and significant between-school ETGs than those with positive and significant within-school ETGs for upper elementary school grades in ELA. In individual districts, 22 of the 29 have positive and significant between-school ETGs and 13 have positive and significant within-school ETGs (Figure IV.3). None has a negative and significant ETG for either component.

[21] In all cases, between-school differences are also more variable across districts than within-school differences are. Thus, variation across district ETGs results primarily from variation in the between-school component of the ETG.

Table IV.3. Average Between-School and Within-School Effective Teaching Gaps, 29-District Average, Years 1 to 3

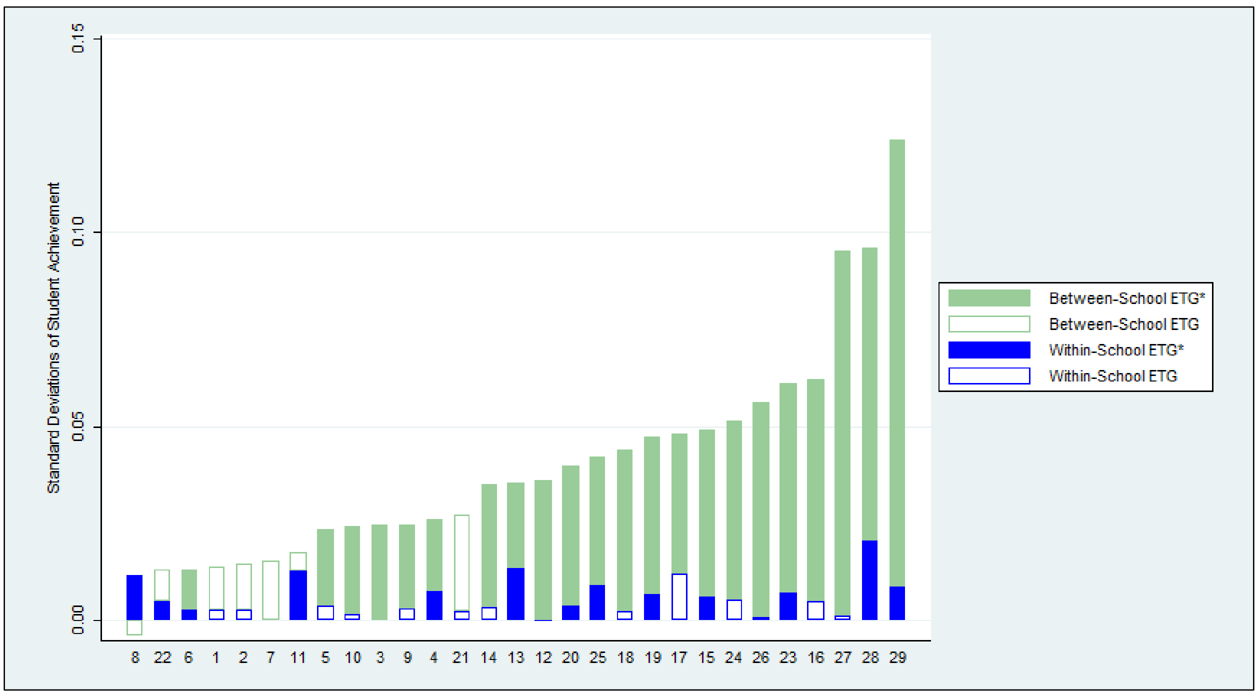
Subject	Between-School	Within-School	Difference	P-Value
English/Language Arts				
Upper Elementary	0.035	0.005	0.029*	0.00
Middle	0.022	0.008	0.014*	0.00
All Grades	0.027	0.007	0.020*	0.00
Math				
Upper Elementary	0.016	0.005	0.011*	0.00
Middle	0.016	0.011	0.005*	0.04
All Grades	0.016	0.008	0.008*	0.00

Source: District administrative data.

Note: Results are for 29 districts, grades 4 through 8, and years 1 to 3. Upper elementary is grades 4 and 5; middle school is grades 6 to 8. District-level results are weighted across grades and years by the number of students. Overall results are weighted equally across districts. The difference in the between- and within-school ETG is the average of the differences for individual districts.

*Indicates statistical significance at the 0.05 level, two-tailed test. We test whether differences in the between- and within-school ETGs are statistically significant using variation within districts.

Figure IV.3. Between-School and Within-School Effective Teaching Gaps in English/Language Arts by District for Upper Elementary Grades, 29 Districts, Years 1 to 3



Source: District administrative data.

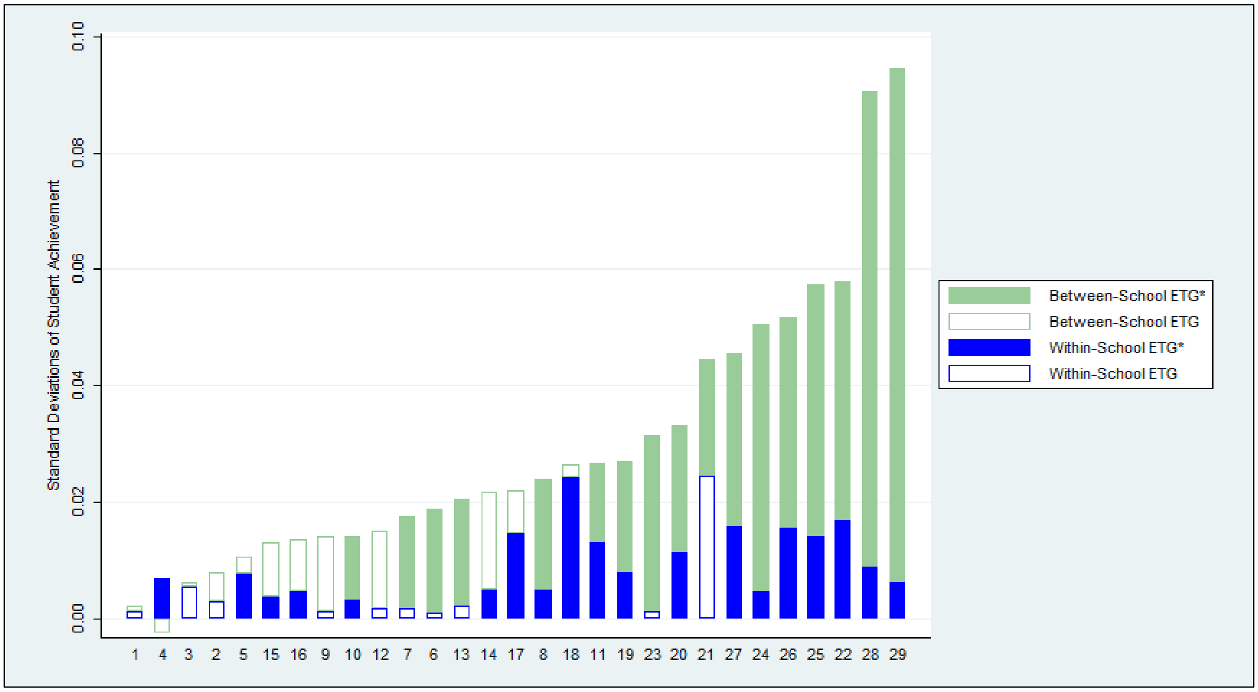
Note: Results are for 29 districts, grades 4 and 5, and years 1 to 3. District-level results are weighted across grades and years by the number of students.

Districts are ordered by the size of the district ETG. District codes 1 to 29 are assigned in order of the size of the district ETG in English/language arts and are consistent across figures.

The solid bars show between- or within-school ETGs that are significantly different from zero at the 0.05 level; the hollow bars indicate between- or within-school ETGs that are not significantly different from zero.

*Indicates statistical significance at the 0.05 level, two-tailed test.

Figure IV.4. Between-School and Within-School Effective Teaching Gaps in English/Language Arts by District for Middle School Grades, 29 Districts, Years 1 to 3



Source: District administrative data.

Note: Results are for 29 districts, grades 6 to 8, and years 1 to 3. District-level results are weighted across grades and years by the number of students.

Districts are ordered by the size of the district ETG. District codes 1 to 29 are assigned in order of the size of the district ETG in English/language arts and are consistent across figures.

The solid bars show between- or within-school ETGs that are significantly different from zero at the 0.05 level; the hollow bars indicate between- or within-school ETGs that are not significantly different from zero.

*Indicates statistical significance at the 0.05 level, two-tailed test.

In middle school grades, differences between FRL and non-FRL students in access to effective teaching in ELA are also larger between schools than within schools, but the difference is smaller than it is for upper elementary school grades. The average between-school ETG for middle schools is 0.022, significantly larger than the average within-school ETG of 0.008 (Table IV.3). This difference of 0.014 in the between- versus within-school ETGs for middle school grades is significantly smaller than the same difference for upper elementary grades. As shown in Figure IV.4, although the within-school ETG is on average smaller than the between-school ETG, statistically significant within-school ETGs are more common in the study districts. The between-school ETGs in ELA are positive and significant for 17 districts, whereas the within-school ETGs are positive and significant for 19 districts. None has a negative and significant ETG for either component. For all grades combined, the average between-school ETG is 0.027 and the average within-school ETG is 0.007, a statistically significant difference of 0.020.

In math, differences in access to effective teaching are larger between schools than within schools as well. In other words, the sorting of teachers and students across schools contributes more to differences in access to effective teaching than the matching of teachers to students within schools. At the upper elementary level, the between-school ETG for math is 0.016 and the within-school ETG is 0.005, a statistically significant difference (Table IV.3). Figure IV.5 shows district-by-district results. The between-school ETG is positive and significant in 13 districts, statistically insignificant in 14 districts, and negative and significant in 2 districts. In addition, 11 districts have positive and significant within-school ETGs; the rest are statistically insignificant.

In middle school grades for math, the between-school ETG is 0.016 and the within-school ETG is 0.011, a statistically significant difference of 0.005 (Table IV.3). However, there are more districts with positive and significant within-school ETGs than between-school ETGs in middle school math. (Figure IV.6). Twenty of the 29 districts have positive and significant within-school math ETGs at the middle school level compared to 10 of the 29 districts with positive and significant between-school math ETGs.

3. Differences in Access to Effective Teaching Between and Within Schools May Be Related to School Size and Student Diversity

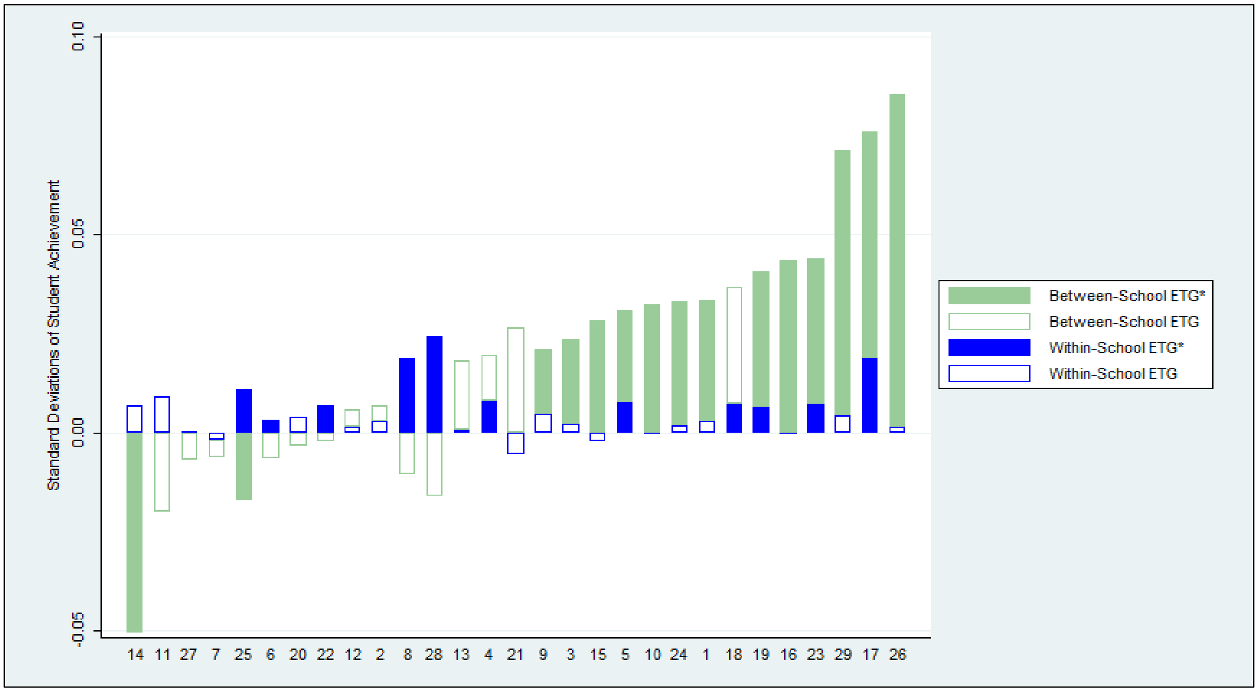
To explore the underlying factors that produce these patterns, we examined differences across grade levels in (1) the separation of FRL and non-FRL students and (2) the grouping of students by achievement level (sometimes referred to as tracking). We found that, within study districts, on average, separation of students by FRL status across schools is greater in the upper elementary school grades than in the middle school grades (see Appendix A). In many study districts, middle schools are larger and more heterogeneous than elementary schools since students from several elementary schools may subsequently attend a single middle school. These differences create an environment in which between-school gaps can be larger in elementary school grades and within-school gaps can be larger in middle school grades.

Tracking policies that assign students to classrooms by achievement level tend to be more common in middle schools (Loveless 2009). A tracking policy could create a larger within-school ETG at the middle school level if achievement level is related to FRL status and if better teachers are assigned to higher-level students. We found that, on average in the study districts, there was greater variation in students' pre-test scores between classrooms in middle schools than in elementary schools, which is consistent with greater tracking in middle schools. However, on average, the evidence suggests far less between-classroom variation in FRL status, suggesting that tracking by ability level may not translate into a comparable level of tracking by FRL status.[22]

In sum, the patterns of between-school and within-school ETGs in both subjects may be related to the tendency for elementary schools to be smaller, resulting in a more homogenous student population due to less diversity in household income within smaller attendance areas. In other words, greater between-school ETGs are more likely to arise in elementary schools, which tend to be smaller and more homogeneous than middle schools.

[22] See Appendix C for more information on the methodology and findings related to the variation in pre-test scores and FRL status between classrooms (as opposed to between students within classrooms) at the elementary and middle school levels.

Figure IV.5. Between-School and Within-School Effective Teaching Gaps in Math by District for Upper Elementary Grades, 29 Districts, Years 1 to 3

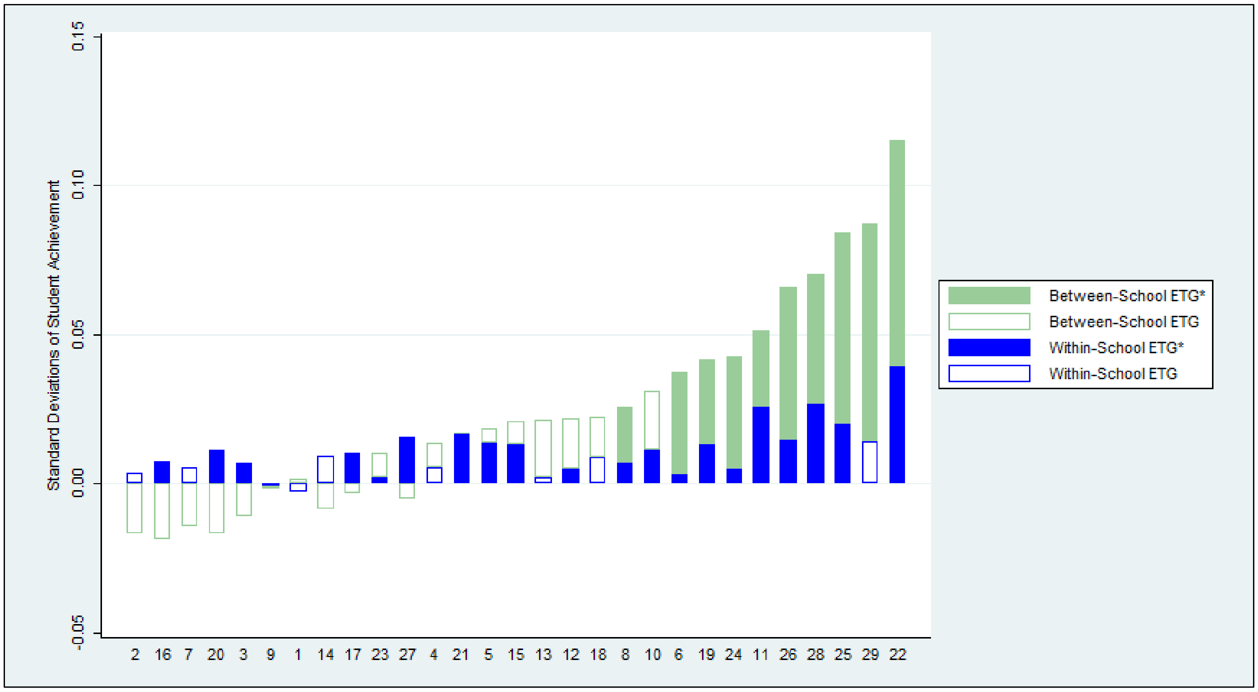


Source: District administrative data.

Note: Results are for 29 districts, grades 4 and 5, and years 1 to 3. District-level results are weighted across grades and years by the number of students. Districts are ordered by the size of the district ETG. District codes 1 to 29 are assigned in order of the size of the district ETG in English/language arts and are consistent across figures. The solid bars show between- or within-school ETGs that are significantly different from zero at the 0.05 level; the hollow bars indicate between- or within-school ETGs that are not significantly different from zero.

*Indicates statistical significance at the 0.05 level, two-tailed test.

Figure IV.6. Between-School and Within-School Effective Teaching Gaps in Math by District for Middle School Grades, 29 Districts, Years 1 to 3



Source: District administrative data.

Note: Results are for 29 districts, grades 6 to 8, and years 1 to 3. District-level results are weighted across grades and years by the number of students. Districts are ordered by the size of the district ETG. District codes 1 to 29 are assigned in order of the size of the district ETG in English/language arts and are consistent across figures. The solid bars show between- or within-school ETGs that are significantly different from zero at the 0.05 level; the hollow bars indicate between- or within-school ETGs that are not significantly different from zero.

*Indicates statistical significance at the 0.05 level, two-tailed test.

4. Access to Effective Teaching by Race and Ethnicity Are Similar to Income-Based Findings

We examined racial and ethnic ETGs to determine whether there are differences in access to effective teaching across students by race and ethnicity.[23] We replaced FRL status with racial and ethnic indicators in the ETG calculation to estimate the differences between Black and White students in one analysis, and between Hispanic and White students in the other. In the analysis of Black/White ETGs, to avoid imprecise results based on few students in one group, we excluded any districts in which fewer than 15 percent of students were Black or fewer than 15 percent were White. This resulted in a sample of 15 districts for the Black/White ETG calculation.[24] Similarly, in the analysis of Hispanic/White ETGs, we excluded districts in which fewer than 15 percent of students were Hispanic or fewer than 15 percent were White; this limited the sample to 18 districts.

Access to effective teaching by race and ethnicity is similar to access to effective teaching by FRL status. In the 15 study districts included in this analysis, Black students experience less effective teaching than White students. The Black/White ETG is 0.019 in ELA and 0.021 in math, compared to FRL ETGs of 0.024 in ELA and 0.015 in math within these same districts. In the 18 study districts in which we conducted the Hispanic/White ETG analysis, the Hispanic/White ETG is 0.033 in ELA and 0.029 in math, compared to FRL ETGs of 0.033 in ELA and 0.024 in math.

In addition to the similarity of the district ETGs by race/ethnicity and by FRL, the district-by-district results are also similar. In other words, districts with larger differences in effective teaching by race/ethnicity also tend to have larger differences in effective teaching by FRL status, as indicated by positive correlation coefficients between the two types of ETGs. Across districts, these correlation coefficients range from 0.52 to 0.81 (Table IV.4). When we remove one outlier district that has much larger ETGs for racial and ethnic gaps than for FRL gaps, the correlations increase, ranging from 0.70 to 0.89 (not shown in table).[25]

---

[23] We also calculated a modified version of the ETG using students' pre-test score as the measure of disadvantage, by regressing teacher value added on individual student pre-test scores. These results were correlated with the FRL ETGs at 0.63 in ELA and 0.80 in math. We focus on presenting results based on FRL status and race/ethnicity because they are predetermined student characteristics that are not affected by the school a student attends. We do not focus on the results using prior student achievement as a measure of disadvantage because of the concern that the same school could be in part responsible for the pre-test level from the prior grade, which would make this an imperfect measure of student disadvantage. If the value-added model captures non-teacher factors in addition to teacher effectiveness (such as school inputs), then the student pre-test scores would be systematically related to the value-added estimates because they are both affected by the school a student attends.

[24] Because the analysis samples differ slightly for math and ELA depending on the availability of student test scores and teacher-student linkages in the two subjects, in one district, there was over 15 percent for both groups in math but slightly under 15 percent for one group in ELA. We included this district in both the ELA and math analyses.

[25] Figures C.7 to C.10 in Appendix C compare results district by district.

Table IV.4. Average Effective Teaching Gaps by FRL and Race (15 Districts) and by FRL and Ethnicity (18 Districts), Years 1 to 3

	English/Language Arts	Math
Black/White Gaps		
FRL ETG	0.024*	0.015*
Black/White ETG	0.019*	0.021*
Correlation	0.52	0.81
Sample Size (districts)	15	15
Hispanic/White Gaps		
FRL ETG	0.033*	0.024*
Hispanic/White ETG	0.033*	0.029*
Correlation	0.52	0.75
Sample Size (districts)	18	18

Source: District administrative data.

Note: Results are for grades 4 through 8, years 1 to 3, and 15 districts for Black/White gaps and 18 districts for Hispanic/White gaps. District-level results are weighted across grades and years by the number of students. Overall results are weighted equally across districts.

*Indicates statistical significance at the 0.05 level, two-tailed test. We test whether the average ETGs are statistically significant from zero using variation across districts.

5. Findings are Robust to an Alternative Value-Added Model That Included an Additional Year of Pre-test Scores, but Sensitive to a Model That Included Peer Effects

To check the robustness of our results, we assessed whether the results depended on (1) our choice of value-added model and (2) whether districts were entirely located within urban areas or spanned urban and suburban areas. Methods and results are discussed here briefly. A fuller description of the results for both sets of sensitivity analyses, including tables and figures, is given in Appendix C.[26] Overall, estimates of disadvantaged students' access to effective teaching are robust to the inclusion of an additional year of math and ELA pre-test scores but sensitive to a value-added model that included the characteristics of a student's classmates. However, estimates based on the peer effects model are only estimated in the subset of districts—9 districts for the upper elementary grades and 23 districts for middle school grades—where we could estimate both the main model and peer effects model.

First, we estimated two additional value-added models to check for the possible influence of test measurement error and peer characteristics on our measures of teacher value added. One value-added model used additional pre-test variables as covariates to account for the possibility that the single year of math and ELA pre-test scores included in our model did not adequately account for students' prior achievement. We also estimated a value-added model that

[26] In Appendix C, we also tested the sensitivity of the results to (1) two alternative calculations of the ETG that consider the possible impact of measurement error in FRL status on the magnitude of the ETG; (2) excluding two districts with at least 20 percent of students in schools affected by Provision 2 or Provision 3, rules that allow schools to participate in federal free- and reduced-price meal programs without collecting individual FRL eligibility data on students every year; and (3) excluding three districts from the middle school math analysis in which students took end-of-course tests rather than end-of-grade tests.

incorporates peer effects, by including as covariates measures of the characteristics of a given student's classroom peers.

The findings from our primary analysis are robust to the inclusion of additional years of pre-test data in the value-added model. The ETG estimates based on the main value-added model differ from the results produced by the alternative model with two years of pretests by no more than 0.007 standard deviations of student test scores. This value-added model produced statistically significant average ETGs of 0.025 in ELA and 0.020 in math, compared with 0.031 and 0.024 for the main model with a single year of pre-tests using a similar sample (Appendix Table C.6).

By contrast, estimates of disadvantaged students' access to effective teaching were sensitive to the inclusion of the characteristics of a student's classmates in the value-added model. The peer effects model produced estimates of the ETGs that were statistically different in ELA and math from those produced by the main model (Appendix Table C.7). This peer effects model produced estimates of average ETGs of 0.006 in ELA and 0.002 in math, with neither estimate statistically significant. ETG estimates from the main model based on the same sample were 0.029 in ELA and 0.024 in math. Despite this difference in the estimated ETG, the value-added estimates from the main model and peer effects model were highly correlated with one another, with a correlation coefficient of 0.96 in ELA and 0.97 in math (Appendix Table B.4).[27]

The final sensitivity analysis was designed to explore whether ETGs arise across urban and suburban schools, as a way of investigating the external validity of our results. Most study districts are defined by city boundaries. Their competing school districts are located in suburban areas for which data are not available. Therefore, the teachers and students in those suburban schools are not included in the calculation of the ETG. However, several study districts cover entire counties that include schools in both the urban core and suburban areas. In this sensitivity analysis, we compared the ETG from an analysis that included only the urban core of five county-level districts with ETG estimates from the whole district.

On average, across five countywide study districts, the ETG for schools in the urban core does not differ from the ETG for all schools by more than 0.005 standard deviations of student test scores in either subject. This suggests that the ETGs in our urban-only districts are not omitting an important dimension of inequity in the distribution of teachers. However, because urban and suburban areas within a countywide district are subject to the same salary schedule and other teacher policies, the results may not fully generalize to settings in which urban districts are surrounded by competing suburban districts with different policy environments.

B. Relating Effective Teaching Gaps to District Characteristics

Figures IV.1 and IV.2 indicate that there is variation from district to district in disadvantaged students' access to effective teaching. In both ELA and math, we found that access to effective teaching is equitable or nearly equitable in some districts, whereas in other districts, there are substantial differences between FRL and non-FRL students in access to

[27] To better understand the influence of peers on disadvantaged students' access to effective teaching, we plan to estimate alternative specifications of the peer effects model in the study's final report.

effective teaching. In this section, we examine the relationship of ETGs to several basic district characteristics—district size, urbanicity, and geographic region. These results help us understand how the results of the study may generalize, and also help us to compare results from this study to those from other studies that have been primarily based on large districts in the South or West. In this analysis, we group districts into three categories by size. Medium-sized districts have fewer than 40,000 students, large districts have 40,000 to 100,000 students, and very large districts have more than 100,000 students. To measure urbanicity, we use data on the percentage of students living in a large city—a distinct concept from district size.

Among study districts in ELA, the ETG is significantly related to urbanicity, whereas associations with district size and region are not statistically significant. Table IV.5 shows that in our study sample, the more urban districts tend to have greater inequities in access to effective teaching (Table IV.5). A district located entirely within a large city has, on average, an ETG that is 0.017 standard deviations of student test scores larger than a district located outside a large city. There are no statistically significant relationships between the ETG and district size or region (Table IV.6). The point estimate for medium-sized districts is smaller than for large or very large districts, but the difference between medium-sized districts and other districts is not statistically significant. By region, the largest point estimate is for southern districts, but this estimate is not significantly different from all other regions combined.

In math, by contrast, urbanicity is not significantly related to the ETG, but district size and region are significantly related to it. In particular, the estimated ETG in medium-sized districts is significantly smaller than that found in large or very large districts. Southern districts have significantly larger ETGs than districts in other regions. This last result suggests that the overall findings of this study on ETGs for math may be smaller than those from prior research in part due to the more diverse mix of districts recruited to participate.

Table IV.5. Relationship Between the Effective Teaching Gap and Urbanicity

	English/ Language Arts	Math
Percent of Students Attending Schools in a Large City	0.017*	0.001
Number of Districts	29	29

Source: District administrative data.

Note: Coefficients from regressions using results for 29 districts, weighting by the inverse variance of the ETG. ETG results were generated for 29 districts, grades 4 through 8, and years 1 to 3.

*Indicates statistical significance of the regression coefficient at the 0.05 level, two-tailed test.

Table IV.6. Relationship Between the Effective Teaching Gap, District Size, and Geographic Region

	Effective Teaching Gap		
	English/ Language Arts	Math	Number of Districts
All Districts	0.034*	0.024*	29
By District Size			
Medium districts	0.022	0.010*	8
Large districts	0.037	0.027	15
Very large districts	0.044	0.037	6
By Geographic Region			
Midwest	0.024	0.014	7
North	0.025	0.013	4
South	0.043	0.037*	12
West	0.034	0.020	6

Source: District administrative data.

Note: Estimates in the table represent the mean ETG for districts within the subgroup represented in the row. Results are based on 29 districts, grades 4 through 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.

*Indicates statistical significance at the 0.05 level, two-tailed test. We test whether the ETG is significantly different from zero for all districts in the first row of the table. In all other rows, the statistical significance is calculated as the difference between districts in a given category and all other districts combined.

C. Summary of Findings

This study documents access to effective teaching over a three-year period in 29 geographically diverse districts. The findings provide information for policymakers on the extent to which disadvantaged students have equal access to effective teaching within districts. In summary, we have these main findings:

- **On average, disadvantaged students did not have equal access to effective teaching in the 29 study districts in grades 4 through 8**. FRL students experienced less effective teaching than non-FRL students on average within districts, with statistically significant differences of 0.034 standard deviations of student test scores in ELA and 0.024 standard deviations in math. The ETG results by race and ethnicity were positively correlated with ETGs by FRL status. This finding was sensitive to one alternative model specification.

- **Providing equal access to effective teaching for FRL and non-FRL students—by reducing the ETG to zero for one year—would reduce differences in student achievement between FRL and non-FRL students by two percentile points**. The difference in student achievement between FRL and non-FRL students would decrease from 28 percentile points to 26 percentile points in ELA and from 26 percentile points to 24 percentile points in math.

- **Access to effective teaching for FRL and non-FRL students varied across study districts, with equal access to effective teaching in some districts and greater access to effective teaching for non-FRL students in other districts**. Across study districts, 27 of 29 districts in ELA and 19 of 29 districts in math had statistically significant ETGs that favored non-FRL students, while the rest had equal access to effective teaching (statistically insignificant ETGs). None of the study districts had a statistically significant ETG that favored FRL students.

- **Access to effective teaching for disadvantaged students in the study districts did not change over time**. Over the three study years, there were no statistically significant differences in the average ETG across years for either ELA or math.

- **The between-school ETG was larger than the within-school ETG**. States and districts often focus on access to effective teaching between schools but do not capture inequities within schools. The ETG allows us to separately measure the degree to which differences in effective teaching occur between schools and within schools. We find that most of the district ETG is accounted for by between-school differences, but some differences in access to effective teaching arise within schools as well.

  For ELA, the between-school ETG was larger than the within-school ETG, especially at the elementary grades. The between-school ETG was 0.029 standard deviations of student test scores larger than the within-school ETG in the upper elementary grades and 0.014 standard deviations larger in the middle school grades. In addition, the difference in the between- and within-school ETGs was significantly larger for upper elementary grades than for middle school grades. In math, these differences were 0.011 in upper elementary grades and 0.005 in middle school grades.

  These patterns of between- and within-school ETGs in both subjects may be related to the tendency for elementary schools to be smaller than middle schools, resulting in a more homogenous student population due to less diversity in household income within smaller attendance areas.

REFERENCES

Aaronson, Daniel, Lisa Barrow, and William Sander. "Teachers and Student Achievement in the Chicago Public High Schools." *Journal of Labor Economics*, vol. 25, no. 1, 2007, pp. 95-135.

Aigner, Dennis. "Regression with a Binary Independent Variable Subject to Errors of Observation." *Journal of Econometrics*, vol. 1, 1973, pp. 49-60.

Arellano, Manuel. "Computing Robust Standard Errors for Within-Groups Estimators." *Oxford Bulletin of Economics and Statistics*, vol. 49, no. 4, November 1987, pp. 431-34.

Ballou, Dale. "Value-Added Assessment: Lessons from Tennessee." In *Value-Added Models in Education: Theory and Application,* edited by Robert Lissitz. Maple Grove, MN: *Journal of Applied Measurement Press*, 2005.

Ballou, Dale, William Sanders, and Paul Wright. "Controlling for Student Background in Value-Added Assessment of Teachers." *Journal of Educational and Behavioral Statistics*, vol. 29, no. 1, 2004, pp. 37-65.

Boyd, Donald, Pam Grossman, Hamilton Lankford, Susanna Loeb, and James Wyckoff. "Who Leaves? Teacher Attrition and Student Achievement." National Bureau of Economic Research Working Paper No. 14022, May 2008.

Briggs, Derek, and Ben Domingue. "Due Diligence and the Evaluation of Teachers: A Review of the Value-Added Analysis Underlying the Effectiveness Rankings of Los Angeles Unified School District Teachers by the Los Angeles Times." Boulder, CO: National Education Policy Center, February 2011.

Brown, Cynthia, and Kati Haycock. "Essential Elements of Teacher Policy in ESEA: Effectiveness, Fairness, and Evaluation." Washington, DC: Center for American Progress and The Education Trust, 2011.

Buonaccorsi, John P. *Measurement Error: Models, Methods, and Applications*. Boca Raton, FL: Chapman and Hall/CRC, 2010.

Carroll, Stephen, Robert Reichardt, and Cassandra Guarino. "The Distribution of Teachers Among California's Districts and Schools." Santa Monica, CA: RAND, October 2000.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. "The Long-Term Impacts of Teachers: Teacher Value-Added and Student Outcomes in Adulthood." Cambridge, MA: National Bureau of Economic Research, 2011.

Clotfelter, Charles, Helen Ladd, Jacob Vigdor, and Justin Wheeler. "High Poverty Schools and the Distribution of Teachers and Principals." Washington, DC: National Center for Analysis of Longitudinal Data in Education Research, December 2006.

Clotfelter, Charles, Helen Ladd, and Jacob Vigdor. "Teacher Credentials and Student Achievement in High School." *The Journal of Human Resources*, vol. 45, no. 3, 2010, pp. 655-681.

Education Trust. "Their FAIR Share: How Texas-Sized Gaps in Teacher Quality Shortchange Low-Income and Minority Students." Washington, DC: EdTrust, February 2008.

Feng, Li, and Tim Sass. "Teacher Quality and Teacher Mobility." Andrew Young School of Policy Studies Research Paper Series, Working Paper No. 12-08, 2012.

Freeman, Richard. "Longitudinal Analysis of the Effects of Trade Unions." *Journal of Labor Economics*, vol. 2, 1984, pp. 1-26.

Glazerman, Steven, and Jeffrey Max. "Do Low-Income Students Have Equal Access to the Highest-Performing Teachers?" NCEE Evaluation Brief. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. Document No. PP11-23a, 2011.

Goldhaber, Dan, and Duncan Chaplin. "Assessing the 'Rothstein Test': Does It Really Show Teacher Value-added Models Are Biased?" Washington, DC: American Institutes for Research, National Center for Analysis of Longitudinal Data in Education Research, 2012.

Gordon, Robert, Thomas Kane, and Douglas Staiger. "Identifying Effective Teachers Using Performance on the Job." Washington, DC: The Brookings Institution, The Hamilton Project, April 2006.

Hahnel, Carrie, and Orville Jackson. "Learning Denied: The Case for Equitable Access to Effective Teaching in California's Largest School District." Oakland, CA: The Education Trust-West, 2012.

Hanushek, Eric, and Steven Rivkin. "Using Value-Added Measures of Teacher Quality." Washington, DC: Center for Analysis of Longitudinal Data in Education Research, The Urban Institute, 2010.

Hanushek, Eric, John Kain, Daniel O'Brien, and Steven Rivkin. "The Market for Teacher Quality." National Bureau of Economic Research Working Paper no. 11154, February 2005.

Hanushek, Eric, John Kain, and Steven Rivkin. "Why Public Schools Lose Teachers." Working Paper No. 8599. Cambridge, MA: National Bureau of Economic Research, 2001.

Hock, Heinrich, and Eric Isenberg. "Methods for Accounting for Co-Teaching in Value-Added Models." Washington, DC: Mathematica Policy Research, June 2012.

Hoxby, Caroline, and Gretchen Weingarth. "Taking Race Out of the Equation: School Reassignment and the Structure of Peer Effects." Working paper. Cambridge, MA: Harvard University, 2006.

Isenberg, Eric, and Heinrich Hock. "Measuring School and Teacher Value Added in DC, 2011–2012 School Year." Washington, DC: Mathematica Policy Research, August 2012.

Jackson, C. Kirabo. “Match Quality, Worker Productivity, and Worker Mobility: Direct Evidence from Teachers.” National Bureau of Economic Research Working Paper No. 15990, May 2010.

Jerald, Craig, Kati Haycock, and Amy Wilkins. “Fighting for Quality and Equality, Too. How State Policymakers Can Ensure the Drive to Improve Teacher Quality Doesn't Just Trickle Down to Poor and Minority Children.” Washington, DC: Education Trust, 2009.

Johnson, Matthew, Stephen Lipscomb, Brian Gill, Kevin Booker, and Julie Bruch. “Value-Added Models for the Pittsburgh Public Schools.” Washington, DC: Mathematica Policy Research, February 2012.

Kalogrides, Demetra, Susanna Loeb, and Tara Beteille. “Systematic Sorting: Teacher Characteristics and Class Assignments." *Sociology of Education,* vol. 86, no. 2, 2013, pp. 103-123.

Kane, Thomas, Jonah Rockoff, and Douglas Staiger. “What Does Certification Tell Us About Teacher Effectiveness? Evidence from New York City.” NBER Working Paper No. 12155. Cambridge, MA: National Bureau of Economic Research, 2006.

Kane, Thomas, and Douglas Staiger. “Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation.” NBER Working Paper No. 14607. Cambridge, MA: National Bureau of Economic Research, 2008.

Kane, Thomas, Douglas Staiger, Steve Cantrell, Jeff Archer, Sarah Buhayar, Kerri Kerr, Todd Kawakita, and David Parker. “Gathering Feedback for Teaching: Combining High-Quality Observations with Student Surveys and Achievement Gains.” Seattle, WA: Bill & Melinda Gates Foundation, 2012.

Koedel, Cory, and Julian Betts. “Re-Examining the Role of Teacher Quality In the Educational Production Function.” Columbia, MI: University of Missouri, Department of Economics, April 2007.

Koedel, Cory, and Julian Betts. “Does Student Sorting Invalidate Value-Added Models of Teacher Effectiveness? An Extended Analysis of the Rothstein Critique.” Cambridge, MA: National Bureau of Economic Research, July 2009.

Lankford, Hamilton, Susanna Loeb, and James H. Wyckoff. “Teacher Sorting and the Plight of Urban Schools: A Descriptive Analysis.” *Educational Evaluation and Policy Analysis*, vol. 24, no. 1, 2002, pp. 37-62.

Liang, Kung-Yee, and Scott L. Zeger. “Longitudinal Data Analysis Using Generalized Linear Models.” *Biometrika*, vol. 73, no. 1, April 1986, pp. 13-22.

Loveless, Tom. “Tracking and Detracking: High Achievers in Massachusetts Middle Schools.” Washington, DC: Fordham Institute, 2009.

Mansfield, Richard. “Teacher Equality and Student Inequality.” Ithaca, NY: Cornell University, 2012.

McCaffrey, Daniel, J. R. Lockwood, Daniel Koretz, and Laura Hamilton. "Evaluating Value-Added Models for Teacher Accountability." Santa Monica, CA: The RAND Corporation, 2004.

McCaffrey, Daniel, Tim Sass, J. R. Lockwood, and Kata Mihaly. "The Intertemporal Variability of Teacher Effect Estimates." *Education Finance and Policy*, vol. 4, no. 4, October 2009, pp. 572-606.

Mead, Sara. "Recent State Action on Teacher Effectiveness: What's in State Laws and Regulations?" Washington, DC: Bellwether Education Partners, 2012.

Morris, Carl N. "Parametric Empirical Bayes Inference: Theory and Applications." *Journal of American Statistical Association*, vol. 78, no. 381, 1983, pp. 47-55.

Nye, Barbara, Spyros Konstantopoulos, and Larry Hedges. "How Large are Teacher Effects?" *Educational Evaluation and Policy Analysis*, vol. 26, no. 3, Fall 2004, pp. 237-257.

Ponza, Michael, Philip Gleason, Lara Hulsey, and Quinn Moore. "Erroneous Payments in the NSLP and SBP. Volume I: Findings. NSLP/SBP Access, Participation, Eligibility, and Certification Study." Washington, DC: Mathematica Policy Research, 2007.

Presley, Jennifer, Bradford White, and Yuqin Gong. "Examining the Distribution and Impact of Teacher Quality in Illinois." Edwardsville, IL: Illinois Education Research Council, 2005.

Reardon, Sean. "The Widening Academic-Achievement Gap between the Rich and the Poor: New Evidence and Possible Explanations." Stanford, CA: Stanford University, January 2011.

Rockoff, Jonah. "The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data." *The American Economic Review*, vol. 94, no. 2, January 2004, pp. 247-252.

Rivkin, Steven, Eric Hanushek, and John Kain. "Teachers, Schools, and Academic Achievement." *Econometrica*, vol. 73, no. 2, March 2005, pp. 417-458.

Rothstein, Jesse. "Student Sorting and Bias in Value-Added Estimation: Selection on Observables and Unobservables." *Education Finance and Policy*, vol. 4, no. 4, 2009, pp. 537-571.

Rothstein, Jesse. "Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement." *Quarterly Journal of Economics*, vol. 125, no. 1, 2010, pp. 175–214.

Rowan, Brian, Fang-Shen Chiang, and Robert Miller. "Using Research on Employees' Performance to Study the Effects of Teachers on Students' Achievement." *Sociology of Education*, vol. 70, no. 4, October 2007, pp. 256-284.

Sacerdote, Bruce. "Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far?" In *Handbook of the Economics of Education*, vol. 3, edited by Eric Hanushek, Stephen Machin, and Ludger Woessmann. Oxford, UK: Elsevier, 2011.

Savoca, Elizabeth. "Measurement Errors in Binary Regressors: An Application to Measuring the Effects of Specific Psychiatric Diseases on Earnings." *Health Services and Outcomes Research Methodology,* vol. 1, no. 2, 2000, pp. 149-164.

Sass, Tim, Jane Hannaway, Zeyu Xu, David Figlio, and Li Feng. "Value Added of Teachers in High-Poverty Schools and Lower-Poverty Schools." CALDER Working Paper No. 52, National Center for Analysis of Longitudinal Data in Education Research. Washington, DC: Urban Institute, 2010.

Scafidi, Benjamin, David Sjoquist, and Todd Stinebrickner. "Race, Poverty, and Teacher Mobility." *Economics of Education Review,* vol. 26, no. 2, 2007, pp. 145-159.

Sullivan, Daniel. "A Note on the Estimation of Linear Regression Models with Heteroskedastic Measurement Errors." Working Paper No. 2001-23. Chicago, IL: Federal Reserve Bank of Chicago, 2001.

Tennessee Department of Education. "Tennessee's Most Effective Teachers: Are They Assigned to the Schools That Need Them Most?" Nashville, TN: Tennessee Department of Education, 2007.

U.S. Department of Education. "ESEA Flexibility." Washington, DC: U.S. Department of Education, June 2012a.

U.S. Department of Education. "NAEP Data Explorer." Available at [http://nces.ed.gov/ nationsreportcard/naepdata/dataset.aspx]. Accessed September 20, 2012b.

U.S. Department of Education. "A Summary of Highly Qualified Teacher Data for School Year 2009-10." Washington, DC: U.S. Department of Education, June 2011.

U.S. Department of Education. "Teacher Incentive Fund; Notice Inviting Applications for New Awards for Fiscal Year (FY) 2010; Rule and Notice." *Federal Register,* vol. 75, no. 98, May 21, 2010.

U.S. Department of Education. "Race to the Top Program: Executive Summary." Washington, DC: U.S. Department of Education, November 2009.

Value-Added Research Center. "NYC Teacher Data Initiative: Technical Report on the NYC Value-Added Model." Madison, WI: Value-Added Research Center, Wisconsin Center for Education Research, University of Wisconsin-Madison, 2010.

This page has been left blank for double-sided copying.

APPENDIX A

ANALYTIC METHODS

This page has been left blank for double-sided copying.

APPENDIX A: ANALYTIC METHODS

This appendix provides technical details of the value-added models and the method by which we calculated the Effective Teaching Gap (ETG) from value-added estimates.

A. Value-Added Models

In this section, we describe our statistical approach to estimating teacher value added. We describe in the first subsection the basic statistical model. We then explain, sequentially, our approach for handling co-teaching, imprecisely measured pre-test scores, students with missing data, and multiple end-of-course tests for the same subject given within a grade. In the final two subsections we present the approach we used to estimate the error-adjusted standard deviation of value-added estimates, and the details of how we estimated alternative specifications of the value-added model that we used to conduct sensitivity analyses.

1. Framework for Estimating Teacher Value Added

Our basic approach for estimating teacher value added was to use a regression model that controlled for a series of baseline student characteristics that could be related to academic achievement or might otherwise be confounded with the assignment of students to teachers. Specifically, we assumed that a student's post-test score depended on prior achievement, background characteristics, the student's current teachers, and additional unmeasured factors unrelated to teaching assignments.[28]

We estimated value added separately for each of the 29 districts, five grade levels (grades 4 through 8), two subjects (math or English/language arts), and three years (2008–09, 2009–10, and 2010-11).[29,30] The regression equation is:

$$\text{(A.1)} \quad Y_{it} = \boldsymbol{\lambda}'\mathbf{L}_{i(t-1)} + \boldsymbol{\eta}'\mathbf{X}_{it} + \boldsymbol{\theta}'\mathbf{M}_{it} + \varepsilon_{it},$$

where Y_{it} is the post-test score for student i in year t, and $\mathbf{L}_{i(t-1)}$ represents test scores for that student in English/language arts (ELA) and math in the prior year. The pre-test scores captured prior inputs into student achievement. Control variables for individual student background characteristics were included in $\mathbf{X}_{it}$. $\mathbf{M}_{it}$ represents a set of binary indicator variables for the teachers. Finally, ε_{it} is an error term that captured unobserved factors that influence student

[28] To avoid assigning value-added estimates to teachers who may be linked erroneously to a few students in the data, we estimated a coefficient for a teacher only if he or she taught at least 10 students in a given year. The students assigned to these teachers were omitted from the analysis.

[29] As explained in Chapter II, for five districts for which data were collected from statewide databases, we used the school years 2007-08, 2008-09, and 2009-10. Of the remaining districts that provided data to us directly, we excluded four districts in year 1 and middle school grades for one district in year 2 due to lack of teacher-student links for these years and grade levels.

[30] From here on, for simplicity, we will use "district-grade combinations" to refer to district-grade-year-subject combinations.

achievement and measurement error in the post-test, and $\boldsymbol{\lambda}$, $\boldsymbol{\eta}$, and $\boldsymbol{\theta}$ are vectors of parameters to be estimated. The key parameters are those included in the vector $\boldsymbol{\theta}$, which are the value-added regression coefficients for individual teachers. They represent the effect of a teacher on the achievement of his or her students, after accounting for student characteristics.

In our model, we controlled for a set of student characteristics, $\mathbf{X}_{it}$, that was common to all study districts. This approach ensured that any differences we document in access to effective teaching across districts are not a result of using different statistical models in different districts. The common value-added model included the following student characteristics, which we obtained from district administrative records:

- Math and ELA scores from the prior school year (we accounted for prior-year scores in math and ELA regardless of the post-test subject)
- Free or reduced-price lunch (FRL) status
- Limited English proficiency
- Special education status
- Gender
- Whether a student is African American or Black
- Whether a student is Hispanic, Native American, multi-race, or "other" race
- Whether a student transferred across schools during the year

We collected test score data on students in the district on state ELA and math tests. All original scale scores were converted to z-scores by subtracting the mean test score of students in the same state, year, and grade who took the same assessment, and dividing by the standard deviation of the test scores of students in the state. Thus, the value-added estimates obtained from the regression are stated in terms of achievement effect size units—that is, standard deviation units within a statewide population of students.

2. Accounting for Multiple Teachers Responsible for the Same Students

Because students may be taught a subject by more than one teacher over the course of a school year, we used a procedure we call the Full Roster Method to estimate value added (Hock and Isenberg 2012). This approach can be used to account for team teaching, supplemental course taking, and students who transfer across schools. The method is based on the assumption that the combined efforts of team teachers constitute a single input into student achievement, with these teachers' joint effectiveness attributed to all teachers on the team. It yields results very similar to a method that would form an extra variable for each set of team teachers, but does not require specifying these team variables explicitly. For teachers who teach some students individually and others as part of a team, the Full Roster Method results in value-added estimates approximately equal to the student-weighted average of their individual estimates and team estimates.

To implement the Full Roster Method, we modified the regression so that the teacher–student link, rather than the student, was the unit of observation. A student contributed one observation to the model for each teacher to whom he or she was linked. For example, students who have a single math class taught by two teachers each contributed two observations to the analysis file, while those whose math class was taught by a single teacher contributed a single observation. The corresponding regression equation for student *i* taught by teacher *j* is expressed as:

$$\text{(A.2)} \quad Y_{ijt} = \boldsymbol{\lambda}'\mathbf{L}_{i(t-1)} + \boldsymbol{\eta}'\mathbf{X}_{it} + \boldsymbol{\theta}'\mathbf{R}_{ijt} + \varepsilon_{ijt}$$

where the notation largely parallels that of Equation 1. The term $\mathbf{R}_{ijt}$ is a vector of binary indicators (one for each teacher in the sample) that indicate whether student *i* appeared on the roster of teacher *j* during year *t*. For teacher–student link *ij*, the *j*th element of $\mathbf{R}_{ijt}$ is one, and the remaining values are zero. If all students in the data are linked to a single teacher for the whole year, Equation 2 reduces to the teacher fixed effects approach described by Equation 1. As in the basic model, $\boldsymbol{\theta}_j$ represents the value-added estimate for teacher *j*.

The Full Roster Method accommodates any pattern of team teaching and shared instructional responsibility. In addition to accounting for multiple teachers, we incorporated information, if available, on the proportion of the year that each student spent with each teacher (the dosage) using weights. We calculated dosage explicitly from enrollment and detailed roster information, if available.

To account for cases in which the student spent only part of the year with a given teacher, we estimated the coefficients using weighted least squares (WLS) rather than ordinary least squares (OLS). In this technique, each teacher–student combination is weighted by the dosage associated with that combination. For a student who split time equally between two math teachers, the weight associated with each of the two observations for the student was 0.5. We addressed the correlation in the error term, e_{ijt}, across multiple observations by using a cluster-robust sandwich variance estimator (Liang and Zeger 1986; Arellano 1987) to obtain standard errors that are consistent in the presence of both heteroskedasticity and clustering at the student level.

3. Addressing Measurement Error

We corrected for measurement error in the pre-tests by using grade-specific reliability data available from test publishers. 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 pre-test scores are likely to be attenuated due to measurement error. 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. By netting out the known amount of measurement error, the errors-in-variables correction eliminates this source of bias.

The specific errors-in-variables method we used is a moment-based correction to the linear regression estimator based on the reliability ratio, which is the proportion of the observed variability in the pre-tests that is not due to measurement error (Buonaccorsi 2010). Focusing first on the case in which each student was linked to only one teacher, the uncorrected OLS estimate of the full set of regression coefficients could be written as

$$(\text{A.3}) \quad \hat{\boldsymbol{\beta}}^{OLS} = \left[\mathbf{Z}_t'\mathbf{Z}_t\right]^{-1}\mathbf{Z}_t'\mathbf{y}_t\,,$$

where $\mathbf{y}_t$ is a stacked vector of post-tests for year t, and matrix $\mathbf{Z}_t = [\mathbf{L}_t, \mathbf{X}_t, \mathbf{M}_t]$ represents a stacked matrix of pre-tests, characteristics, and teacher-student link variables for the same year. The corresponding moment-corrected set of estimates is

$$(\text{A.4}) \quad \hat{\boldsymbol{\beta}}^{EIV} = \left[\mathbf{Z}_t'\mathbf{Z}_t - \mathbf{Q}_t\right]^{-1}\mathbf{Z}_t'\mathbf{y}_t\,,$$

where $\mathbf{Q}_t$ is a diagonal matrix with the kth element equal to $(1 - r_t^k)v_t^k$; the terms r_t^k and v_t^k represent the reliability ratio and the total observed variability for the kth variable contained in $\mathbf{Z}_t$. It can be shown that equation (A.4) yields a consistent estimator of the true regression coefficients under traditional assumptions about the measurement error structure.[31] To apply this correction, we obtained reliability information for the pre-test variables from published information about state tests by year, grade, and subject.[32,33] The reliability ratio for the other variables is assumed to be one, so that only the first two diagonal elements of $\mathbf{Q}$ (corresponding to the pre-tests) are non-zero.

---

[31] The specific assumptions are that: (1) the main regression error term is mean-independent of the measurement error in all of the covariates, (2) the measurement error term for any covariate is mean independent of the level of all covariates, and (3) the measurement error terms are uncorrelated across all covariates.

[32] Reliability information was obtained from technical reports distributed by the test publishers or state education departments, if available. In cases where test information could not be found for a given state, year, grade, and subject, we set the reliability to be 0.9, which was approximately equal to the mean of the reliability measures for cases in which the information was available.

[33] We use a single measure of test/re-test reliability for each test. A more efficient estimator would account for varying reliability across the range of test scores, which tend to be most reliable in the middle of the distribution of student achievement, and less reliable toward the extremes of the test score distribution for a given grade. Sullivan (2001) describes a Heteroskedastic Errors-in-Variables (HEIV) estimator that accounts for varying levels of measurement error for different observations. Sullivan (2001) emphasizes that (1) failing to correct for measurement error will lead to estimates that are biased and inconsistent; and (2) the advantage of using the HEIV approach over the errors-in-variables approach lies in greater asymptotic efficiency. In other words, our choice to use the errors-in-variables approach and not HEIV implies that our approach addresses the potential problems of bias and inconsistency but produces less precise estimates. For estimating value added in this study, implementing HEIV would be very resource-intensive, requiring incorporating information on measurement error for every possible test score for hundreds of pre-tests, while the gains of such an approach would be small, given that the coefficient on pre-test scores tends to be estimated very precisely using the errors-in-variables method that we follow.

Implementing the moment-based errors-in-variables correction in conjunction with the Full Roster Method required two modifications. First, as noted previously, we used a matrix of dosage to estimate a WLS analogue of equation (A.4). Second, we used a two-step procedure to calculate standard errors that account for repeated student observations. In the first step, the errors-in-variables method applied to equation (A.2) obtains unbiased estimates of the set of pre-test coefficients, $\hat{\boldsymbol{\lambda}}$, which allow us to calculate an adjusted gain score: $A_{ijt} = Y_{ijt} - \hat{\boldsymbol{\lambda}}\mathbf{L}_{i(t-1)}$.[34] In the second step, the adjusted gain score was regressed on the student covariates and teacher-student links. The second-step regression yields the same point estimates of teacher value added as the first-step regression, but yields standard errors that are robust to clustering.

The two-step procedure will tend to underestimate the standard errors of the teacher effects. The dependent variable in the second step, A_{ijgt}, is calculated using estimated pre-test coefficients, and the estimates will contain some amount of error. The second-step regression did not account for this common source of error affecting all students in a grade. Nonetheless, large sample sizes of student observations yielded relatively precise estimates of $\hat{\boldsymbol{\lambda}}$, which mitigated this potential concern.

4. Imputation of Missing Data

We imputed values of missing student covariates so that the value-added regression made use of the non-missing data elements for every student.[35] This imputation was done using a regression-based method that estimated the relationships among characteristics for observations with non-missing data. This information was then used to fill in the missing data elements for students with partially missing data based on the values of their non-missing data elements.

Due to the importance of FRL status to the calculation of the ETG, we took extra steps to ensure the completeness of these data. For districts in which there were implausible upward or downward spikes in the percentage of FRL students in one year, we used each student's FRL status from other years in place of the FRL status from the abnormal year.[36] In addition, two federal regulations, known as Provision 2 and Provision 3, posed a particular challenge to identifying individual students' FRL status in some districts. Under these provisions, schools offered free meals to all students, without determining the students' eligibility for the benefits based on their household circumstances. The reimbursement received by a school for each free meal served was based on the pattern of reimbursement from a base year in which the school determined individual students' FRL eligibility status. Schools commonly chose to participate in Provision 2 or Provision 3 if a large proportion of students were eligible for FRL, because the administrative cost of distinguishing between eligible and ineligible students can be higher than the cost of providing free lunches to ineligible students. Therefore, in student-level

[34] We use the eivreg command in Stata.

[35] We did not impute outcome (post-test) data that was missing for students who left the district before the end of the year or were absent on the day of the test.

[36] For example, in a district where there was an implausible change in the percentage of FRL students in Year 2, we replaced students' FRL status in Year 2 with their status in Year 3. If a student was in the data for only two years, we replaced their FRL status in Year 2 with either Year 1 or Year 3.

administrative data sets in districts that included Provision 2 or Provision 3 schools, all students may have appeared as FRL-eligible, even though not all students met the eligibility criteria.

In our data, 3 of 30 districts originally eligible for the study had 20 percent or more of their students enrolled in schools participating in Provision 2 or Provision 3. We included 2 of these 3 districts in our analysis, imputing FRL status for students at these schools. We dropped the third district from the study rather than using the imputation strategy because a majority of students attended Provision 2/Provision 3 schools.

We used the following steps to impute FRL status for all students attending Provision 2 or 3 schools that reported 100 percent of students receiving FRL. First, we used the FRL status for the student from the Provision 2 or 3 "base year" in which the school determined the FRL status of all students, if these data were available. For students for whom these data are not available, we used the regression-based method based on other student characteristics used as control variables in the value-added model, parents' education, and the percentage of FRL students in the school before the school began to participate in this program (according to the Common Core of Data).

We did not impute missing values of pre-test scores, as doing so for this key control variable may introduce unacceptably large errors in the estimates of individual teacher effectiveness. If we had access to statewide databases that allow us to track mobile students across districts, we filled in as many post-test and pre-test score values as possible from the statewide data. Otherwise, student observations with missing pre-test scores were excluded from the value-added regressions. This includes excluding students who skipped or repeated a grade because they had a pre-test score that is from a different grade level than their peers.[37]

5. Multiple Tests for the Same Subject Given Within a Grade

Some states tested middle school students using end-of-course tests rather than end-of-grade tests. For example, seventh-grade students may have taken general math, a lower-level course, or pre-algebra, a higher-level course, and were tested accordingly. This complicates the value-added approach, as we had to calculate value added for teachers in the same grade based on different tests administered to different sets of students within the grade. Continuing the example, there were a set of value-added estimates for teachers of general math and another for teachers of pre-algebra. We ultimately had to create a single set of grade-level estimates to preserve comparability in the measure of the ETG between districts that used multiple tests within a grade and those that do not.

If there was systematic sorting of teachers to different courses, we needed a way to rank teachers of different courses against each other. This cannot be done directly, because their students took different end-of-year tests. For example, a school might have assigned its better teachers to the higher-level courses. In this case, we would not have wanted to simply pool value-added estimates of these two groups of teachers from separate regression models, as that would have presumed that the average teacher of, for example, pre-algebra is of equal

[37] On average across the districts, fewer than one percent of students were excluded because they skipped or repeated a grade. The percentage ranged from 0 to 6.7 percent.

effectiveness as the average teacher of general math when teachers may have been assigned to courses based in part on school principals' knowledge of their effectiveness.

Our approach was to measure value added separately for the different tests that students take and then equate the value-added estimates of teachers across the two tests using teachers who taught both courses and therefore had students who took both tests.[38] The difference in the differences of average value-added estimates between two-course and one-course teachers measured the degree to which teachers are sorted across courses according to their effectiveness. For example, assume that two-course teachers outperformed one-course teachers of general math 7 by half a standard deviation of value-added estimates. To the contrary, assume that two-course teachers lagged one-course teachers of pre-algebra by half a standard deviation. This implies that one-course pre-algebra teachers achieved results that were a full standard deviation above those of one-course general math 7 teachers. We used this gap as a means of comparing the value added for teachers of different courses. For the two-course teachers themselves, we combined their two course-specific scores by first adjusting them and then using a weighted average of the two scores, where the weights are the proportion of students that a teacher had in that course relative to the number of students the teacher had in all courses combined.[39]

6. Calculating the Error-Adjusted Standard Deviation of Teacher Value-Added Estimates

In some statistics, such as the ETG Ratio, we presented the standard deviation of value-added estimates, a measure of the variability of teacher value added for a given district-grade combination. When doing so, we used an adjusted standard deviation that removes estimation error. Because value-added estimates are not known quantities, the unadjusted standard deviation of value-added estimates partly reflects estimation error in each value-added estimate. Therefore, the unadjusted standard deviation of value-added estimates tends to overstate the true variability of teacher value added.

We calculated the error-adjusted variance of teacher value-added estimates by subtracting the mean squared standard error of the value-added estimate from the variance of the unadjusted value-added estimates. Both the calculation of the variance of the unadjusted estimates and the mean squared standard error were weighted. We used an empirical Bayes procedure described by Morris (1983) to derive the weights using an iterative procedure. In general, using this procedure, estimates that have a larger standard error received less weight, and vice versa.

[38] We also equated the standard deviation of the value-added estimates for teachers of different courses using the teachers of both courses as a bridge between teachers who taught only one type of course. After running separate value-added models by course, we multiplied the post-test scores of one group of students by a constant that equalized the standard deviation of value-added estimates for teachers who taught both courses.

[39] Because the standard deviation of general math test scores and standard deviation of pre-algebra scores underestimates the standard deviation of test scores of all students had they taken a common test, we applied a final adjustment to the value-added estimates. We used the teacher-level equating parameters for the mean and standard deviation of teacher value added to translate all student post-tests onto the scale of a general math, and then calculated the ratio of the standard deviation of all test scores to the standard deviation of general math scores. We then multiplied value-added estimates by this ratio.

7. Alternative Value-Added Specifications Used for Sensitivity Analyses

As described in Chapter III, we estimated two alternative value-added models that included additional sets of control variables from those included in the main model. These variables were intended to further help to account for measurement error in pre-test scores and to account for the possibility that a student's classroom peers influence the student's achievement. We tested whether the ETG results depended on the choice of a value-added model. The first alternative model used two years of pre-test data rather than just a single year to better account for pre-existing differences in true levels of student achievement. The second alternative model accounted for peer effects of other students in the classroom by including classroom-level variables such as the mean pre-test score of a student's classmates.

Two years of pre-test scores. This model is based on the following value-added regression:

$$\text{(A.5)} \quad Y_{ijt} = \boldsymbol{\lambda}'\mathbf{L}_{i(t-1)} + \boldsymbol{\omega}'\mathbf{L}_{i(t-2)} + \boldsymbol{\eta}'\mathbf{X}_{it} + \boldsymbol{\theta}'\mathbf{R}_{ijt} + \varepsilon_{ijt},$$

where $\mathbf{L}_{i(t-1)}$ denotes the standard pre-tests included in the model, and $\mathbf{L}_{i(t-2)}$ represents lagged pre-test in math and ELA, that is, student test scores from two years prior to the year in which the post-test is measured. When estimating equation (A.5), lagged pre-tests were treated as regular covariates in the regression. We did not apply an errors-in-variables correction to $\mathbf{L}_{i(t-2)}$. Because only one year of pre-test data was available for 4th graders and all students in year 1, we limited comparisons between models with a single year of pre-test data and multiple years of pre-test data to grades 5 through 8 in years 2 and 3 of the study.

We imputed values of the lagged pre-test to ensure that the analysis sample for the model with two years of pre-test scores was identical to the sample for the model with a single year.[40] This ensured that any differences between results from the main and alternative model were due to differences in the variables included, rather than the students contributing information to the regression. There were several reasons why students may have missed lagged pre-test scores. For example, some students were enrolled in another school district two years prior to the current year. Other students skipped or repeated a grade the year before the current year; these students would not have had lagged pre-test scores from the same grade as other students in their current grade. Imputation of lagged pre-test scores was carried out using a method similar to the one used for imputing other missing student background characteristics.

Peer effects specification. The peer effects model we specified allowed for the possibility that (1) having more high-performing students in a classroom improved the performance of a given student, (2) having fewer disadvantaged students in a classroom improved that student's performance, and (3) having a narrower range of achievement in a given classroom improved the performance of students in that classroom. Thus, we included the following classroom-level variables in the peer effects model: the average classroom pre-test scores, the proportion of students in the classroom who are eligible to receive FRL, and the standard deviation of the

[40] We imputed pre-test values for 7 percent of students on average across the districts, with the percentage ranging from 0 to 12 percent.

student pre-test scores in the classroom.[41] The specification we estimated assumed that the peer characteristics we examined (mean prior achievement, proportion FRL, and variation in classmates' prior achievement) potentially influenced a student's achievement linearly.

Because peer effects were calculated at the classroom level, we used multiple classrooms per teacher to provide variation in the peer effect variables for individual teachers. Otherwise, we might have confounded the characteristics of students in a teacher's classroom with the selection of teachers who work with students like these.[42] Unlike estimation of individual student characteristics, which was based on differences in student achievement across different types of students assigned to the same teacher, estimating peer effects required multiple "observations" of a teacher's classroom. In particular, we used variation in classroom-level characteristics for teachers of multiple sections in a given grade and subject. For teachers with multiple years of data, we also captured year-to-year variation in the composition of a teacher's class.[43]

Estimating the peer effects model required an extra step because we used multiple years of data to estimate peer effects, but were interested in teacher value added from each year separately. First, using data from years 1 to 3, we estimated a pooled regression across years within a district-grade combination:

$$\text{(A.6)} \quad Y_{ijt} = \boldsymbol{\lambda}_*'\mathbf{L}_{i(t-1)} + \boldsymbol{\eta}_*'\mathbf{X}_{it} + \boldsymbol{\theta}_*'\mathbf{R}_{ijt} + \boldsymbol{\psi}_*'\mathbf{C}_{ijt} + \varepsilon_{ijt} .$$

In this equation, $\mathbf{C}_{ijt}$ represents the peer effects variables described previously, which were calculated separately for each teacher-student link in each year *t*. The coefficients are subscripted by an asterisk to indicate that they are constrained to be the same across years. This restriction is necessary to calculate stable estimates of $\boldsymbol{\psi}_*$, which measures the relationships between the post-test and the classroom-level measures. In the absence of this restriction, year-to-year variability in $\mathbf{C}_{ijt}$ needed to estimate peer effects would be fully absorbed by the year-specific teacher effects or could be confounded with changes over time in the other coefficients.

Based on the estimated coefficients on the classroom characteristics, $\hat{\boldsymbol{\psi}}_*$, we then calculated a peer effects-adjusted post-test measure:

[41] We calculated classroom variables individually for each student, excluding that student's contribution to the classroom statistic.

[42] Using the "fixed effects" strategy described here avoided biasing results that could arise from confounding teacher selection with peer effects, but involved two trade-offs. First, we assumed that differences in effects of peer composition between classrooms of the same teacher extrapolated to larger differences in peer composition that might occur across classrooms of different teachers. Second, because relatively few classrooms are taught by the same teacher, if there was measurement error in the classroom characteristics, the estimates of peer effects might have been too small (attenuation bias).

[43] We treated teachers who switched schools across years as separate teachers when calculating peer effects so that peer effects results were not driven by teachers with large variation in peers because they moved to a different teaching environment In other words, we relied only on within-school variation for teachers when measuring peer effects.

$$\text{(A.7)} \quad Y_{ijt}^{adjusted} = Y_{ijt} - \hat{\boldsymbol{\psi}}_*' \mathbf{C}_{ijt} .$$

This measure was calculated individually for each teacher-student link in each year.

Finally, we entered this adjusted post-test measure as the dependent variable in regression equation A.2 to obtain our final peer effect–adjusted value-added measures for teachers for a given district-grade-year combination. These value-added estimates reflected teacher performance in a single year, even though we used multiple years of data to estimate peer effects.

B. Effective Teaching Gap Measure

To document the differences in access to effective teaching between disadvantaged and nondisadvantaged students, we used a measure called the Effective Teaching Gap (ETG).

1. Single-Grade Measures of the ETG

The district ETG is the average value added of the teachers of nondisadvantaged students minus the average value added of teachers of disadvantaged students. Teachers who have both types of students in their classrooms counted toward both averages in proportion to the number of disadvantaged and nondisadvantaged students they taught. We computed the district ETG using a simple regression:

$$\text{(A.8)} \quad V_j = \alpha + \delta FRL_{jk} + e_{jk} ,$$

where V_j is the value added of teacher *j*. Each teacher contributed two observations for a given subject: once for FRL students and once for non-FRL students. Each observation was weighted according to the total dosage for students of that type. For example, a teacher who had 20 FRL-eligible students and 10 FRL-ineligible students would have weights of 20 and 10. We regressed V_j 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. The estimated coefficient δ measures the estimated mean difference in effective teaching between nondisadvantaged and disadvantaged students in the district, with a positive δ indicating an inequitable gap and a negative δ indicating a compensatory gap. 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 (Liang and Zeger 1986; Arellano 1987).

The ETG can measure relative access to effective teaching, even though the value-added model used to generate measures of effective teaching included FRL as a control variable. Because we included teacher fixed effects when estimating value added, the estimates of the coefficients on the covariates, including FRL, were based on within-teacher variation. We distinguished between differences in outcomes for FRL and non-FRL students due to access to effective teaching and differences due to other factors correlated with FRL status because there was always a group of teachers in our study districts who taught both FRL and non-FRL students. This allowed us to estimate the coefficient on FRL status in the value-added model. For example, suppose that FRL students score, on average, 0.1 standard deviations below non-FRL students who have the same teacher and the same other baseline characteristics. The value-added model would assign a coefficient of -0.1 to the FRL indicator. Now, suppose that two students—one FRL and non-FRL—otherwise have the same baseline characteristics but are taught by two different teachers, and the FRL student scores 0.3 standard deviations below the non-FRL student. Since the FRL student scored even lower than what would have been expected based on his FRL status, the model attributes this difference to the FRL student having had a less effective teacher.

To estimate the between-school component of the ETG, we calculated a weighted average of the teachers' value-added estimates at the school-grade level, where each teacher is weighted by the number of student-equivalents linked to that teacher in the analysis file. We then linked this average to every student in that school-grade, and estimated Equation A.8 using school-grade level *j* in place of teacher *j*. This component of the ETG accounted for the fact that some schools in the district have, on average, more effective teaching across the school than do other schools. It does not account for the fact that within individual schools, FRL and non-FRL students may be matched to teachers who exhibited varying levels of effective teaching. That aspect of relative access to effective teaching is captured by the within-school ETG, which is the difference between the district ETG and between-school ETG.

2. Student Separation and Variation in Teacher Effectiveness: District Characteristics that Bound the ETG

There are two key characteristics that can affect the maximum possible ETG in a district, and that therefore could be related to the magnitude of the ETG in a given district: the degree of student segregation and the variation in teacher value added. In this context, a "larger" potential ETG means that the ETG is larger in absolute value. In other words, a larger potential ETG could theoretically lead to bigger gaps that either favor FRL or non-FRL students.

The separation of disadvantaged and nondisadvantaged students between classrooms and across schools. There can be no gap in effective teaching if disadvantaged and nondisadvantaged students were perfectly integrated across teachers: the ETG would be zero regardless of the variation in teacher value added because on average all students would have equal access to effective teaching.[44] For example, if every teacher in a school had classes in which 70 percent of students are FRL, the average value added for FRL and non-FRL students within the school would be the same, regardless of how teachers were distributed. Similarly, if every school had an FRL rate of 70 percent, the average value added across schools for FRL and non-FRL students would be the same. In general, more separation by FRL status between schools leads to higher potential between-school ETGs, and more separation between classrooms within schools leads to higher potential within-school ETGs.

[44] Complete segregation of students would lead to a circumstance in which we could not disentangle the relationship between FRL status and student achievement from the effective teaching gap because there would be no within-teacher variation in student-level FRL status, which allows us to measure the relationship between FRL status and student achievement. This situation does not occur in the study districts in years 1, 2, and 3, however.

We examined the level of student separation using two measures. The first measure, known as the Index of Dissimilarity (D-Index), captures the extent to which FRL and non-FRL students are evenly distributed across schools. The between-school D-Index can be interpreted as the percentage of students from one group (FRL or non-FRL) who would have to change schools to achieve a perfectly even distribution. In other words, all schools would have an equal proportion of FRL students after a given proportion of students from one group changed schools. The second measure captures the proportion of variation in FRL status due to variation between schools and variation between teachers within schools. We distinguished variation that occurs between schools, between teachers within schools, and between students within teachers by using a 3-level unconditional hierarchical linear model in which students are nested within teachers, and teachers are nested within schools. The percentage of variation that occurs between schools measures between-school separation, while the percent of variation within schools captures within-school separation. For example, if the proportion of variation in FRL status between schools were zero, this would indicate that students were perfectly integrated across schools. To examine ability tracking, which may contribute to within-school separation by FRL status, we also measured separation by prior test scores. We examined both measures separately by upper elementary and middle school grade levels.[45]

The level of between-school separation by FRL status was larger in the upper elementary grades than in the middle school grades. The average D-Index is 0.46 in upper elementary grades compared to 0.38 in the middle school grades in both subjects (Table A.1). Similarly, for the average district, 25 percent of the total variation in FRL status occurred between schools for upper elementary grades for both subjects, compared to 18 percent for middle school grades for ELA and 17 percent for math. These differences in separation by grade span also occurred when measuring separation by prior achievement. Twelve percent of the variation in math and ELA test scores occurred between schools for upper elementary grades, while 8 percent occurred between schools for middle school grades. These results are consistent with smaller, more homogeneous attendance areas for elementary schools compared with middle schools.

Consistent with a greater degree of ability tracking at middle school grades (Loveless 2009), within-school separation by prior achievement was greater for middle school grades than for upper elementary grades. In ELA, 30 percent of the variation in pre-test scores occurred between teachers within schools in middle school grades, compared to 13 percent in upper elementary grades. Similarly, more variation in math pre-test scores occurred between teachers within schools in middle school grades than in upper elementary grades: 29 percent compared to 11 percent.

[45] When estimating the hierarchical linear models, it was necessary to exclude schools with only a single teacher or classroom, depending on how students were grouped. For these schools, there was no ability to track students, so we included them in the results by assigning a value of 0 for between teacher/class variation and weighting the model estimates by one minus the share of students who were excluded. The average district in the study had 10 percent of math students and 6 percent of ELA students in schools with a single teacher. For the 7 districts in which we were able to group students by class, the average district had 2 percent of math students and 1 percent of ELA students in schools with a single classroom.

Table A.1. Summary of District Characteristics Potentially Related to the ETG

Measure	English/Language Arts Upper Elementary	English/Language Arts Middle School[a]	English/Language Arts All Grades	Math Upper Elementary	Math Middle School[a]	Math All Grades
FRL Separation Between Schools (D-Index)	0.46	0.38	0.41	0.46	0.38	0.41
Proportion of Variation in FRL						
Between Schools	0.25	0.18	0.21	0.25	0.17	0.20
Within Schools (Between Teachers)	0.03	0.04	0.04	0.02	0.03	0.03
Proportion of Variation in Pre-test Score						
Between Schools	0.12	0.08	0.10	0.12	0.08	0.10
Within Schools (Between Teachers)	0.13	0.30	0.22	0.11	0.29	0.21
Standard Deviation of Teacher Value Added	0.17	0.14	0.16	0.24	0.19	0.22

Source: District administrative data.

Note: Results on the proportion of variation between schools are based on 3-level unconditional hierarchical linear models, in which students are nested within teachers and teachers are nested within schools. The table shows the proportion of variation in FRL status and pre-test scores accounted for by variation between schools and within schools (between teachers). The remaining portion of variation is due to variation in FRL status and pre-test scores for individual teachers within a school. Value-added results are for years 1 to 3. Results are based on 29 districts. Statistics are based on averages across all available years for each district. District-level results for value added are weighted across grades and years by the number of teachers. Overall results are weighted equally across districts. Upper elementary includes grades 4 and 5. Middle school grades include grades 6 to 8. Statistics are evenly weighted across years, districts, and grades. The standard deviation is calculated after removing sampling variance from the value-added estimates through an empirical Bayes shrinkage procedure.

[a]Estimates exclude 4 districts from middle school pre-test estimates.

To test the sensitivity of the results to grouping students by individual classrooms, we repeated our analysis to test the proportion of the variation due to variation between schools and variation between classrooms, using seven districts that provided adequate classroom identifiers for all grades. Table A.2 compares results of decompositions that use the teacher and classroom levels for these seven districts. As shown in the bottom panel of this table, the proportion of variation in prior achievement that is between classrooms in middle school grades is 3 percentage points higher than it is between teachers for ELA (38 compared to 35), and 14 percentage points higher for math (46 compared to 32). For upper elementary grades, the proportion of variation in prior achievement that is between classrooms is 3 percentage points lower than it is between teachers for ELA (13 compared to 16), and the same for math (14 percent in both cases). These results reinforce the initial finding that there is more ability tracking in middle school grades than in elementary grades, particularly in math.

Differences in within-school separation by prior achievement across grade levels do not, however, translate into comparable differences in within-school separation by FRL status. Table A.1 shows that, on average in the study districts, the proportion of variation that occurred between teachers within schools is 0.04 for ELA in middle school grades and 0.03 in upper elementary grades. In math, it was 0.03 for middle school grades and 0.02 for upper elementary grades. Examining variation at the classroom level, Table A.2 shows larger differences in within-classroom variation than within-teacher variation between grade levels, but for the purpose of calculating the ETG, it is within-teacher variation that can contribute to a larger potential within-school ETG.

Table A.2. Proportion of Variation in FRL and Pre-Tests Within Schools for Subset of Districts with Classroom Identifiers for All Grade Levels

	English/Language Arts			Math		
Measure	Upper Elementary	Middle School	All Grades	Upper Elementary	Middle School	All Grades
Proportion of Variation in FRL Within Schools						
Between Teachers	0.03	0.05	0.04	0.03	0.05	0.04
Between Classrooms	0.02	0.06	0.04	0.03	0.07	0.05
Proportion of Variation in Pre-test Score Within Schools						
Between Teachers	0.16	0.35	0.27	0.14	0.32	0.25
Between Classrooms	0.13	0.38	0.28	0.14	0.46	0.33

Source: District administrative data.

Note: Results are from 3-level unconditional hierarchical linear models, in which (1) students are nested within teachers and teachers are nested within schools; and (2) students are nested within classrooms and classrooms are nested within schools. The table shows the proportion of variation in FRL status and pre-test scores accounted for by variation within schools (between teachers or between classrooms). The remaining portion of variation is due to variation in FRL status and pre-test scores (1) between schools and (2) within individual teachers or classrooms. These results are based on seven districts with classroom identifiers in all grades that allowed for students to be grouped by classrooms. Upper elementary includes grades 4 and 5. Middle school grades include grades 6 to 8. Statistics are evenly weighted across years, districts, and grades.

The variation in teacher value added. There cannot be gaps in effective teaching if teacher value added does not vary, even if FRL and non-FRL students are segregated. If there were no variation in teacher value added, the average value added for teachers of FRL and non-FRL students would be the same, regardless of how students were distributed. More generally, the greater the variation in teacher value added, the greater the potential ETG. This is relevant for measuring between- and within-school ETGs—greater variation in value added between schools increases the potential between-school ETG and greater within-school variation in teacher value added increases the potential within-school ETG. Therefore, the greater the total variation is, the greater the potential district ETG.

We use the error-adjusted standard deviation of teacher value added to measure the variation in value added within each district. On average across the districts, an increase of one standard deviation in teacher value added is associated with an increase of 0.16 standard deviations of student achievement in ELA and 0.22 in math (Table A.1). See Appendix B for additional details.

3. Aggregation Across Grades

We computed a district-wide ETG by estimating equation (A.8) with teachers in all grades in the district, rather than estimating the equation separately by grade. Because all test scores were converted to z-scores before they were used in the value-added model, the metric for each district-grade ETG is student effect size units relative to the state population of test takers, which is the same metric as the value-added measures for each district-grade. In addition to calculating the ETG for all grades, we also separately calculated the ETG for upper elementary grades and middle school grades, and compared the ETG across districts and years.

4. ETG Based on Student Race and Ethnicity

We also measured Black/White and Hispanic/White gaps in access to effective teaching. To do this, we alternatively replaced FRL status in Equation A.8 with (1) an indicator for being Black, and (2) an indicator being Hispanic. We limited these analyses to districts in which at least 15 percent of the students are White and 15 percent from the relevant minority group. We used value-added results based models that included all eligible students, but we excluded a teacher’s Hispanic students from the calculation of the Black/White gap and non-Hispanic Black students from the calculation of the Hispanic/White gap. For example, when calculating the Black/White gap using Equation A.8, for a teacher with 10 Black students, 8 White students, and 6 Hispanic students, the teacher would have received a weight of 10 for Black students and a weight of 8 for White students. The 6 Hispanic students would not have played a role in the calculation of the Black/White ETG.

This page has been left blank for double-sided copying.

APPENDIX B

DIAGNOSTICS OF VALUE-ADDED MODELS

This page has been left blank for double-sided copying.

APPENDIX B: DIAGNOSTICS OF VALUE-ADDED MODELS

To better understand the properties of the value-added model that underlies the Effective Teaching Gap (ETG) results, we documented the precision and stability of estimates produced by the value-added model, related these estimates to student achievement, and compared value-added estimates derived from different model specifications. We also compared our diagnostic measures to those from the research literature on value added. This appendix presents diagnostic results for the value-added model; Appendix C presents sensitivity analyses for the ETG (showing, for example, the differences in the ETG resulting from different value-added models).

1. Characteristics of the Value-Added Model

To describe the characteristics of estimates derived from the value-added model, we calculated diagnostic measures to address these two questions:

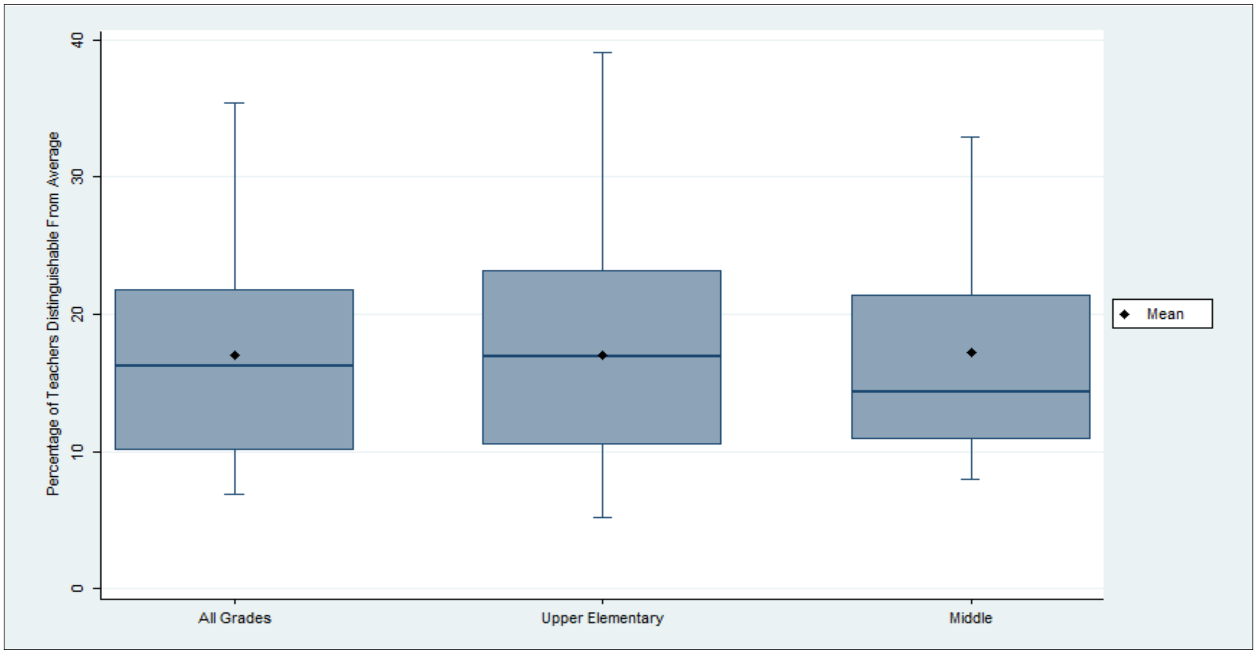
1. Can the value-added model distinguish among more effective, less effective, and average teachers?
2. How stable are teachers' value-added estimates from year to year?

To answer the first question, we conducted two analyses. First, we computed the percentage of teachers whose estimates are statistically distinguishable from the value added of an average teacher in a district, grade, and year (that is, significantly different from zero). Second, we compared the dispersion in the value-added point estimates of all teachers in a district, grade, and year with a measure of how precisely we can estimate value added for an individual teacher. The greater the dispersion in value-added estimates relative to the uncertainty or imprecision of individual value-added estimates (as measured by the estimated standard error), the better able the model was to distinguish among more effective, less effective, and average teachers. Like most of the results in this chapter, the results for each subject are based on 430 individual district-grade-year value-added regressions (29 districts by 5 grades by 3 years, minus some missing data for particular district-grade-year combinations).[46]

We summarize the percentage of teachers with value-added estimates statistically distinguishable from zero across districts in Figure B.1 for English/language arts (ELA) and Figure B.2 for math. Each box-plot presents the distribution of results from the study districts, including the minimum, 25th percentile, median, mean, 75th percentile, and maximum across the 29 study districts. We pooled estimates across grades and years to calculate the average and create a single measure for each district. Because value-added estimates are measured with error, some are statistically indistinguishable from zero. Value-added estimates can be significantly different from zero in either direction, negatively or positively. Negative value-added estimates indicate below-average teaching, while positive estimates indicate above-average teaching.

[46] The math results are based on 427 regressions because we are unable to run value-added regressions for grade 8 in three districts in year 1 for math.

Figure B.1. Percentage of Teachers with English/Language Arts Value-Added Estimates Statistically Distinguishable from Average Across Districts, Years 1 to 3



Source: District administrative data.

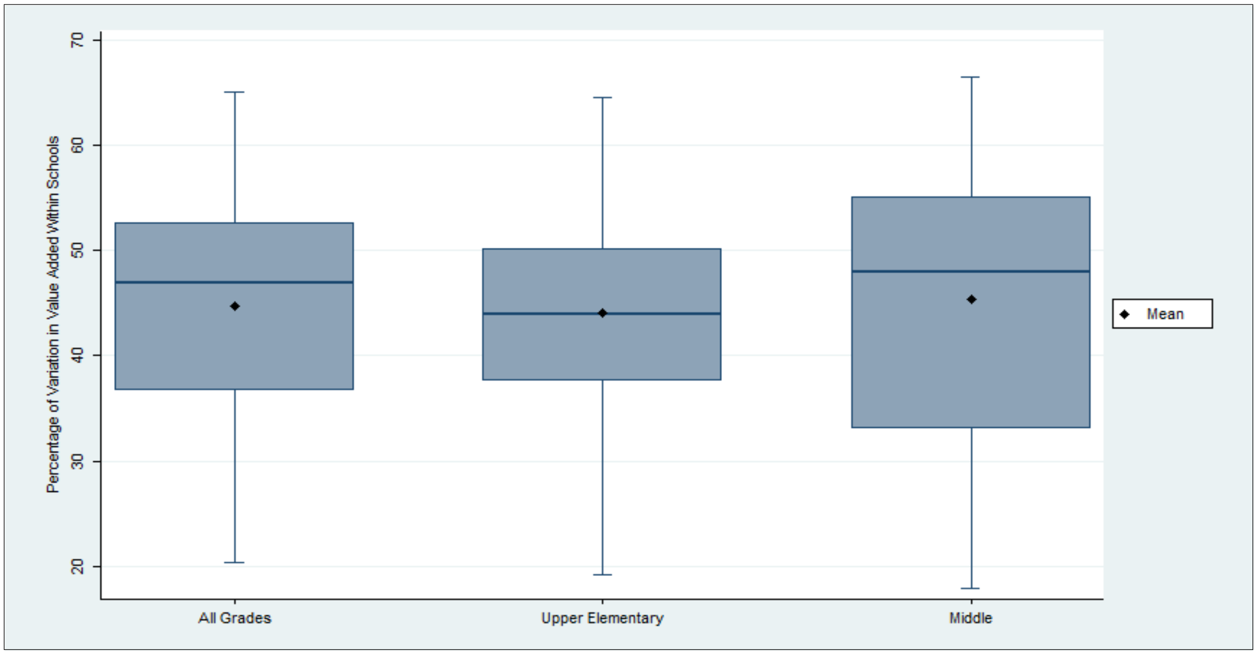
Note: Results are for 29 districts, grades 4 through 8, and years 1 to 3. District-level percentages are weighted across grades and years by the number of teachers. Overall percentages are weighted equally across districts. Distinction of a teacher from average is calculated after removing sampling variance from the value-added estimates through an empirical Bayes shrinkage procedure.

The level indicated by the line in the middle of each box shows the result for the district at the 50th percentile of the distribution of districts. The level indicated by the top of each box shows the result for the district at the 75th percentile. The level indicated by the bottom of each box shows the result for the district at the 25th percentile. The diamond indicates the result for the mean of all districts. The line segments outside the box show the results for the maximum and minimum districts.

We find that, in the average study district, the value-added model distinguished 17 percent of ELA teachers from the average teacher. For 22 districts (that is, 75 percent of the districts), at least 10 percent of teachers were distinguishable from average. The cross-district average is 17 percent for both upper elementary and middle school grades, although the distribution of percent distinguishable is wider across districts for upper elementary grades. This average is higher than the 3 percent of elementary ELA teachers distinguishable from average in a study of a school district in Tennessee that used one year of value-added data (Ballou 2005) and lower than a study in Los Angeles that distinguished 47 percent of teachers from average using six years of data (Briggs and Domingue 2011).

As shown in Figure B.2, on average, 45 percent of math teachers were distinguishable from the average teacher, and at least 37 percent were distinguishable from average for over 22 districts. This compares to 17 percent of elementary math teachers in a Tennessee district for a study using one year of value-added data Ballou (2005), and 57 percent in the study of Los Angeles teachers that used six years of data (Briggs and Domingue 2011). In the average (and median) district, more teachers were distinguishable from average in middle school grades than in upper elementary grades. The higher distribution of percent distinguishable for middle school grades implies that math value-added estimates were more precise for these grades than for upper elementary grades.

Figure B.2. Percentage of Math Teachers with Value-Added Estimates Statistically Distinguishable from Average Across Districts, Years 1 to 3



Source: District administrative data.

Note: Results are for 29 districts, grades 4 through 8, and years 1 to 3. District-level percentages are weighted across grades and years by the number of teachers. Overall percentages are weighted equally across districts. Distinction of a teacher from average is calculated after removing sampling variance from the value-added estimates through an empirical Bayes shrinkage procedure.

The level indicated by the line in the middle of each box shows the result for the district at the 50th percentile of the distribution of districts. The level indicated by the top of each box shows the result for the district at the 75th percentile. The level indicated by the bottom of each box shows the result for the district at the 25th percentile. The diamond indicates the result for the mean of all districts. The line segments outside the box show the results for the maximum and minimum districts.

The second analysis measured the standard deviation of teacher value-added estimates, the average standard error of these estimates, and the ratio of the standard deviation to the average standard error. The ratio of the standard deviation to the average standard error compares the degree with which we can distinguish a group of value-added estimates from each other to the level of precision with which individual value added is estimated. The larger this ratio was, the greater the variability of teachers' value added relative to their sampling variance. Table B.1 presents the standard deviation of teacher value-added estimates in terms of standard deviations of student test scores, the average standard error of value-added estimates, and the ratio of the two for different grade levels across districts.

Across districts, the average standard deviation of value-added estimates over grades 4 through 8 was 0.16 for ELA and 0.22 for math (Table B.1), suggesting that the value-added model we use performed well, at least compared to value-added models used by researchers in other settings. This is within the range of the standard deviation of value-added estimates found in some California, Florida, and North Carolina districts, which range from 0.05 to 0.19 in ELA and 0.12 to 0.27 for math for upper elementary school grades (Koedel and Betts 2007; Sass et al. 2010). For both subjects, the standard deviations are largest for grade 4 and decrease through grade 8. The average standard errors show a similar pattern, decreasing across grades. This results in a relatively stable ratio across grades, averaging 1.74 for ELA and 2.35 for math. So, despite having more students per teacher on which to base value-added estimates in middle school grades, these estimates do not distinguish among middle school teachers more precisely than among elementary school teachers.

Table B.1. Average Variation in Teacher Value-Added Estimates, Years 1 to 3

	English/Language Arts			Math		
Grade	Standard Deviation (SD)	Average Standard Error (SE)	Ratio of SD to Average SE	Standard Deviation	Average Standard Error	Ratio of SD to Average SE
4	0.18	0.10	1.80	0.25	0.11	2.34
5	0.16	0.09	1.69	0.23	0.10	2.33
6	0.15	0.08	1.83	0.21	0.08	2.59
7	0.13	0.08	1.75	0.18	0.07	2.41
8	0.12	0.07	1.65	0.18	0.07	2.32
Grades 4 to 5	0.17	0.10	1.75	0.24	0.10	2.34
Grades 6 to 8	0.14	0.08	1.75	0.19	0.08	2.45
All Grades	0.16	0.09	1.74	0.22	0.09	2.35

Source: District administrative data.

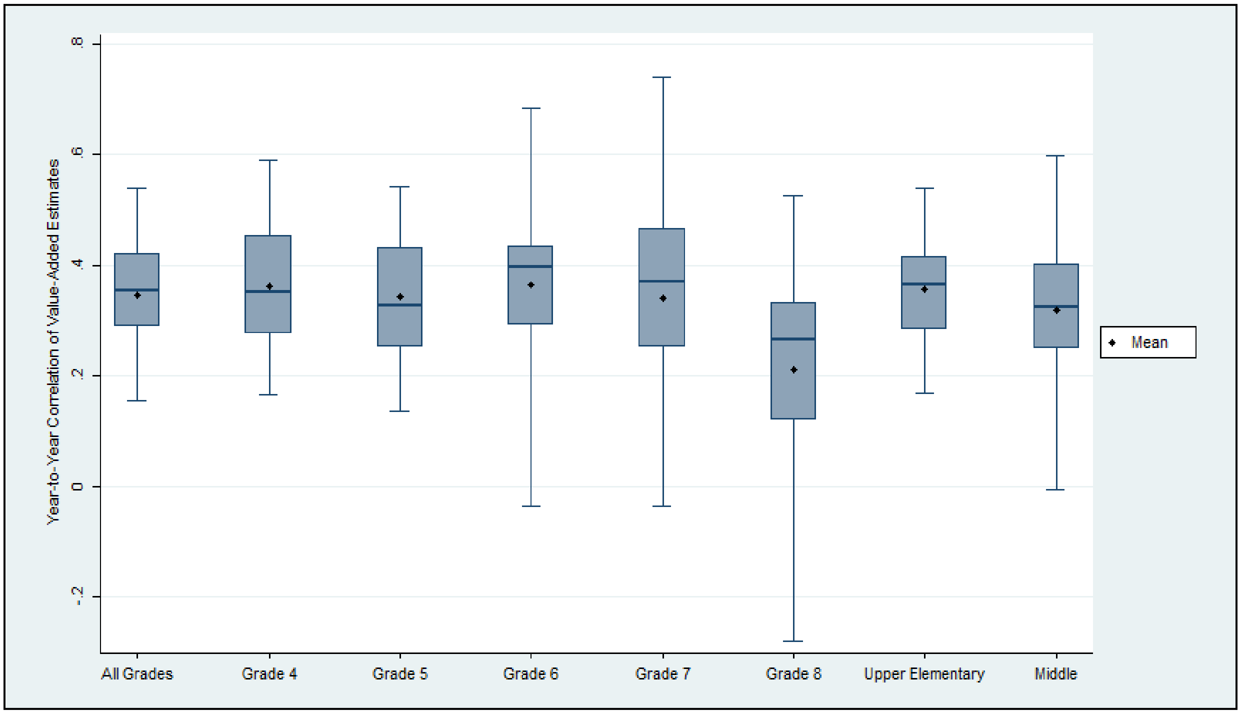
Note: Results are for 29 districts and years 1 to 3. District-level results are weighted across grades and years by the number of teachers. Overall results are weighted equally across districts. The standard deviation is calculated after removing sampling variance from the value-added estimates through an empirical Bayes shrinkage procedure.

For the second question—on year-to-year stability—we calculated the average year-to-year correlation of teacher value-added estimates across districts. Figures B.3 and B.4 present the distribution of these correlations. For the upper elementary grades, we restricted these calculations to 9 districts that have teacher–student–course links. When estimating these correlations within each district, we also note that between 22 and 84 percent of ELA teachers and 27 and 84 percent of math teachers have value-added estimates in the same grade in adjacent years, which is required to be included in the correlation statistic. This implies variation in mobility into and out of the same tested grade from year to year across districts, which could affect the estimated correlations.

The overall year-to-year correlation for ELA value-added estimates is 0.35. As shown in Figure B.3, 21 of the study districts (that is, 75 percent of those included in these correlations) had overall year-to-year correlations above 0.29, although a few had correlations close to or less than zero. Of the 252 year-to-year comparisons we made, 12 of them had negative correlations. The distribution of these correlations does change by grade, with a greater dispersion of results across districts for middle school teachers.

The overall year-to-year correlation for math value-added estimates is 0.49. Twenty-one of the districts had correlations above 0.45, and the minimum average correlation across districts is 0.30. The distribution also varies by grade, with all districts having correlations above 0.46 for grades 5 to 7. Of the 252 year-to-year comparisons we make, 5 of them had negative correlations (with 1 of these in grade 4, 1 in grade 7, and 3 in grade 8). When McCaffrey et al. (2009) studied year-to-year correlations for math teachers in five districts in Florida, they found overall correlations ranging from 0.16 to 0.46 for elementary school teachers and 0.28 to 0.67 for middle school teachers, whereas our study districts ranged from 0.25 to 0.65 for elementary school math teachers and 0.22 to 0.71 for middle school math teachers.

Figure B.3. Year-to-Year Correlation of English/Language Arts Value-Added Estimates Across Districts, Years 1 to 2 and 2 to 3

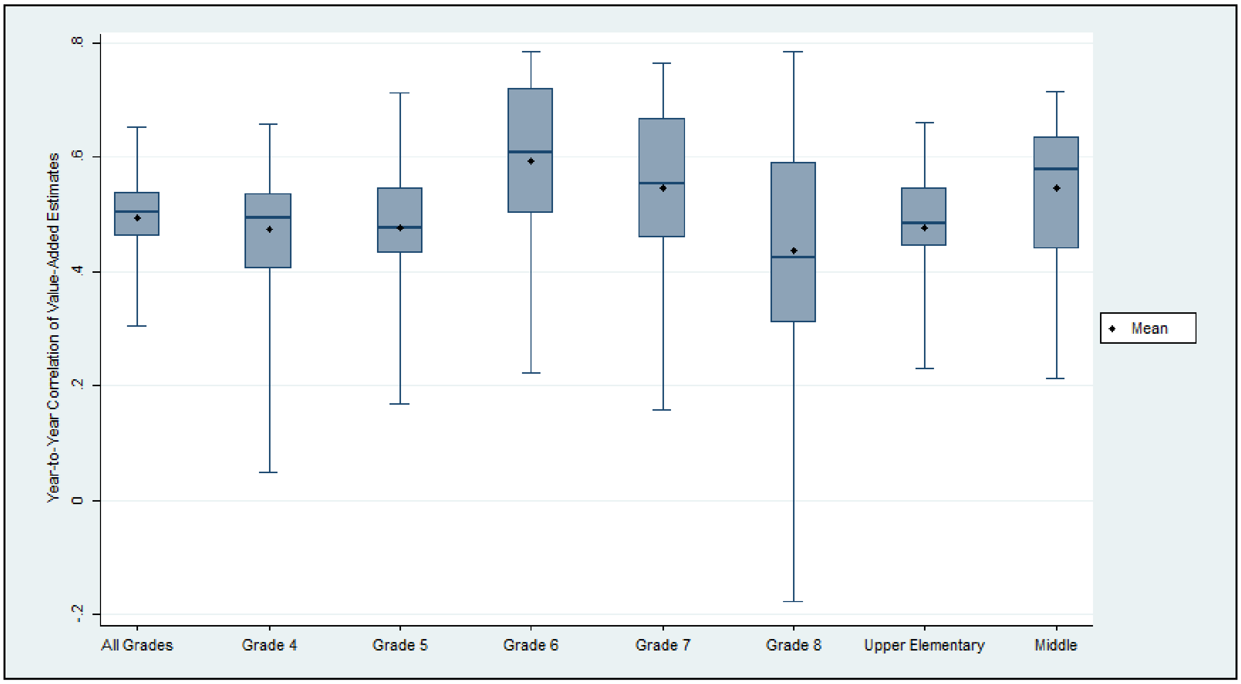


Source: District administrative data.

Note: Results are for 28 districts and 23,156 English/language arts teachers. District-level differences are weighted across grades and years by the number of teachers. Overall results are weighted equally across districts. Year-to-year correlation is calculated after removing sampling variance from the value-added estimates through an empirical Bayes shrinkage procedure.

The level indicated by the line in the middle of each box shows the result for the district at the 50th percentile of the distribution of districts. The level indicated by the top of each box shows the result for the district at the 75th percentile. The level indicated by the bottom of each box shows the result for the district at the 25th percentile. The diamond indicates the result for the mean of all districts. The line segments outside the box show the results for the maximum and minimum districts.

Figure B.4. Year-to-Year Correlation of Math Value-Added Estimates Across Districts, Years 1 to 2 and 2 to 3



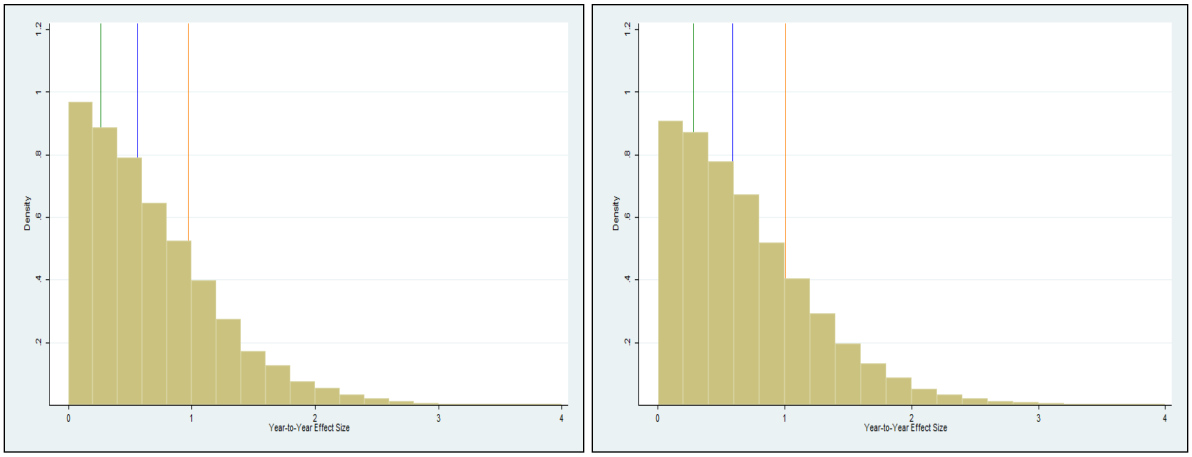
Source: District administrative data.

Note: Results are for 28 districts and 20,061 math teachers. District-level differences are weighted across grades and years by the number of teachers. Overall results are weighted equally across districts. Year-to-year correlation is calculated after removing sampling variance from the value-added estimates through an empirical Bayes shrinkage procedure.

The level indicated by the line in the middle of each box shows the result for the district at the 50th percentile of the distribution of districts. The level indicated by the top of each box shows the result for the district at the 75th percentile. The level indicated by the bottom of each box shows the result for the district at the 25th percentile. The diamond indicates the result for the mean of all districts. The line segments outside the box show the results for the maximum and minimum districts.

When pooling all teachers across districts, in both ELA and math, over 75 percent of teachers moved less than one standard deviation in the distribution of teacher value-added estimates from one year to the next. Figure B.5 presents the year-to-year changes in a teacher's value-added estimates relative to the standard deviation of value-added estimates in that teacher's grade and district with vertical lines representing the 25th, 50th, and 75th percentiles.

Figure B.5. Distribution of Year-to-Year Changes in Math (left) and English/Language Arts (right) Teacher Value-Added Estimates from Years 1 to 3



Source: District administrative data.

Notes: Results are for 28 districts, 23,156 English/language arts teachers, and 20,061 math teachers. Value-added estimates are standardized by dividing each teacher's value-added estimate by the standard distribution of all value-added estimates in that teacher's grade and district. Year-to-year changes are calculated after removing sampling variance from the value-added estimates through an empirical Bayes shrinkage procedure. Vertical lines represent 25th, 50th, and 75th percentile of teachers' year-to-year changes.

2. Relationship Between Value-Added Estimates and Student Achievement

To explore the relationship between the value-added estimates and student achievement, we addressed the following questions:

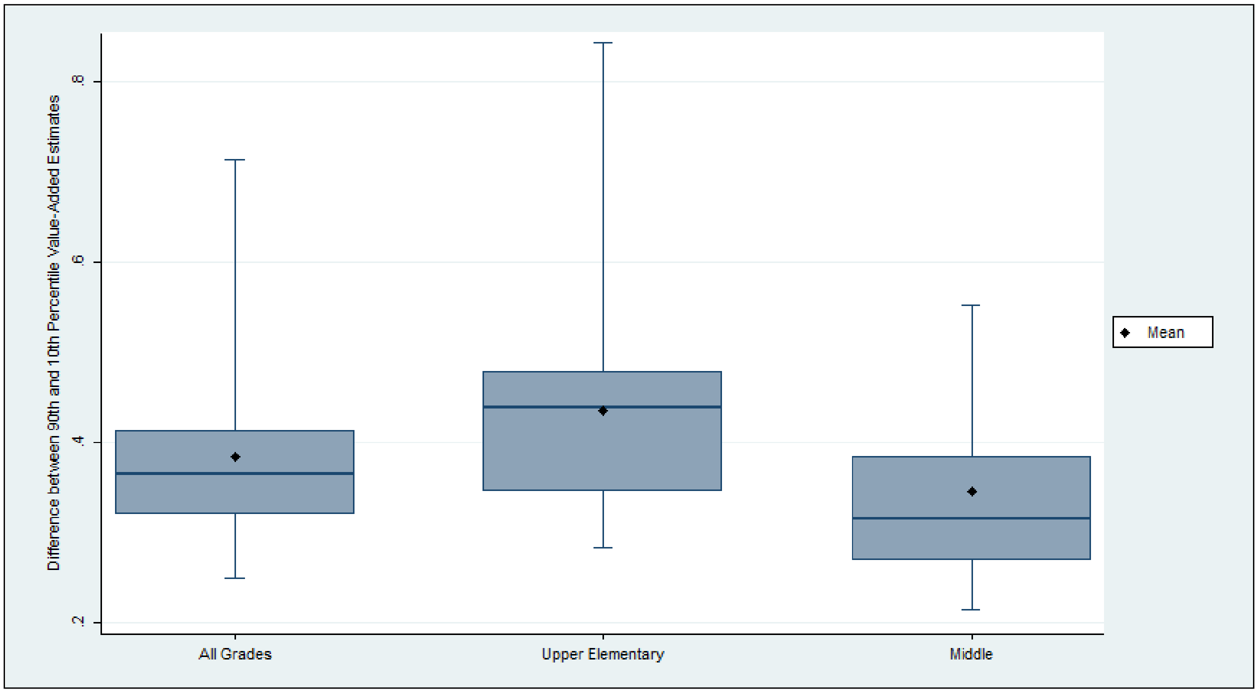
1. What is the difference between the value-added estimate of a teacher at the 90th percentile and a teacher at the 10th percentile?
2. How does the variation in effective teaching compare to the gap in student achievement between FRL and non-FRL students?
3. What proportion of the variation in teacher value-added estimates is within schools and how much is between schools?

Understanding the variation in value-added estimates, and where that variation is concentrated, helps us gauge the importance of effective teaching to student outcomes.

The 90th–10th percentile difference in value-added estimates is 0.38 for ELA and 0.54 for math. In other words, a relatively effective teacher (one at the 90th percentile) had a value-added estimate 0.38 standard deviations of student test scores greater than a relatively ineffective teacher in ELA, and 0.54 standard deviations greater in math. Figures B.6 and B.7 present the distribution of the 90th–10th percentile differences for our sample of districts. By comparison, Sass et al. (2010) reported 90th–10th percentile differences within both high and low poverty schools of 0.25 for ELA and differences ranging from 0.33 to 0.50 for math for districts in Florida and North Carolina. As shown in Figure B.6, the average 90th–10th percentile ELA differences were larger for our study districts, but our measure is based on percentiles of teacher value added across teachers in all schools in a district, not separately by high and low poverty.

As shown in Figure B.7, the distribution of math differences overlaps with those reported by Sass et al. (2010).

Figure B.6. Measured Effectiveness Difference Between 90th and 10th Percentile English/Language Arts Value-Added Estimates Across Districts, Years 1 to 3

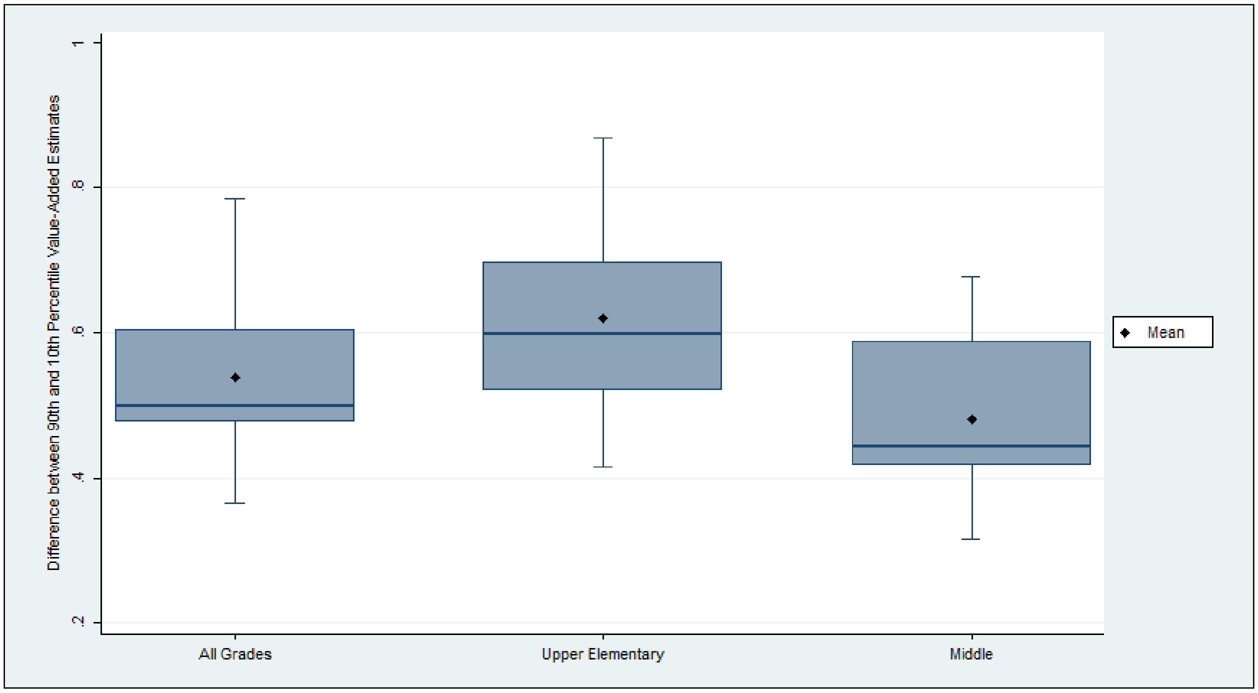


Source: District administrative data.

Note: Results are for 29 districts, grades 4 through 8, and years 1 to 3. District-level differences are weighted across grades and years by the number of teachers. Overall results are weighted equally across districts. The difference between the 90th and 10th percentiles is based on a normal distribution and the standard deviation of value-added estimates after removing sampling variance through an empirical Bayes shrinkage procedure.

The level indicated by the line in the middle of each box shows the result for the district at the 50th percentile of the distribution of districts. The level indicated by the top of each box shows the result for the district at the 75th percentile. The level indicated by the bottom of each box shows the result for the district at the 25th percentile. The diamond indicates the result for the mean of all districts. The line segments outside the box show the results for the maximum and minimum districts.

Figure B.7. Measured Effectiveness Difference Between 90th and 10th Percentile Math Value-Added Estimates Across Districts, Years 1 to 3



Source: District administrative data.

Note: Results are for 29 districts, grades 4 through 8, and years 1 to 3. District-level differences are weighted across grades and years by the number of teachers. Overall results are weighted equally across districts. The difference between the 90th and 10th percentiles is based on a normal distribution and the standard deviation of value-added estimates after removing sampling variance through an empirical Bayes shrinkage procedure.

To gauge the extent to which a reassignment of teachers with high value added to disadvantaged students could potentially close the student achievement gap, we next examined how the 90/10 difference in teacher value added compared to the gap in student achievement between FRL and non-FRL students. For the average study district, the difference between the value added of the 90th percentile teacher and the 10th percentile teacher was equivalent to 62 percent of the ELA achievement gap and 94 percent of the math achievement gap. Table B.2 presents this comparison as the ratio of the 90th–10th percentile difference to the average student achievement gap. For both subjects, these ratios decreased as grade levels increased, consistent with the decrease in the standard deviation of teacher value-added estimates.

Table B.2. Ratio of 90th Percentile/10th Percentile Difference in Teacher Value Added to Student Achievement Gap, 29 Districts, Years 1 to 3

Grade	English/Language Arts			Math		
	Average 90th-10th Percentile VA Score Difference	Average Student Achievement Gap (S.A.G.)	Ratio of VA Difference to Average S.A.G.	Average 90th-10th Percentile VA Score Difference	Average Student Achievement Gap	Ratio of VA Difference to Average S.A.G.
4	0.46	0.73	0.71	0.65	0.66	1.10
5	0.41	0.73	0.65	0.59	0.65	1.03
6	0.39	0.72	0.61	0.54	0.67	0.89
7	0.34	0.70	0.59	0.46	0.65	0.81
8	0.30	0.65	0.52	0.45	0.61	0.86
Grades 4 to 5	0.44	0.73	0.68	0.62	0.66	1.06
Grades 6 to 8	0.35	0.69	0.57	0.48	0.64	0.85
All Grades	0.38	0.70	0.62	0.54	0.65	0.94

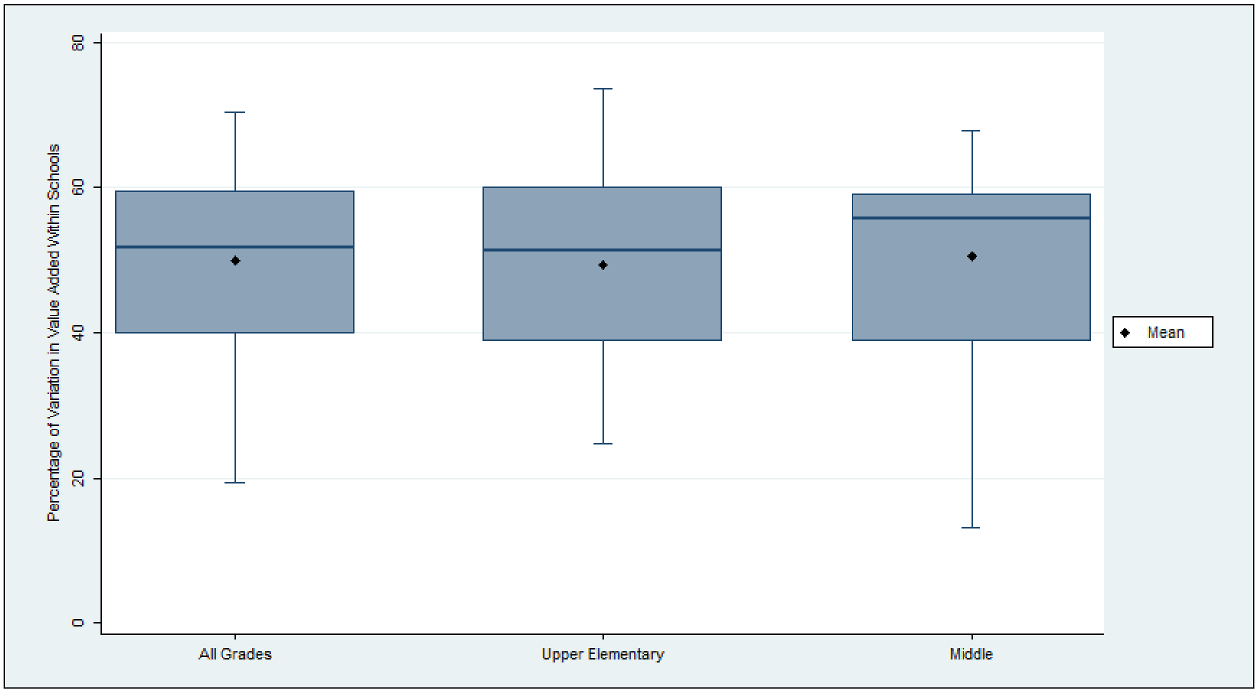
Source: District administrative data.

Note: Results are for 29 districts and years 1 to 3. District-level results are weighted across grades and years by the number of teachers. Overall results are weighted equally across districts. The difference between 90^{th} and 10^{th} percentile is based on a normal distribution and the standard deviation of value-added estimates after removing sampling variance through an empirical Bayes shrinkage procedure. Grade-level ratios are computed within each district-grade-year combination, averaged with equal weight across years within each district, and then averaged with equal weight across districts

Finally, on average, the percentage of variation in teacher value-added estimates from within schools was 50 percent for ELA and 45 percent for math. The remainder was between-school variation. This suggests that the potential of within-school assignments of students to teachers to affect the ETG was approximately equal to that of between-school differences in student sorting. Figure B.8 presents the distribution of these percentages across districts for ELA, and Figure B.9 presents the distribution for math. The average percentage of within-school variation was 1.3 percentage points higher for middle school grades for ELA and 1.2 percentage points higher for math.

We calculated the percentage of the variation from within schools by estimating an ANOVA model for the value-added estimates over schools. The within-school variation equals one minus the percentage of the variation explained by the school indicators. Previous studies have estimated the within-school variation in value-added estimates by including school fixed effects directly in their value-added models, such as Aaronson et al. (2007) and Koedel and Betts (2007). We obtained estimates of the within-school variation from these studies by calculating the ratio of the variance in value-added estimates from models that include school fixed effects to those that exclude them. Using the values reported in these papers, we calculated within-school variation of 46 percent for 9th grade math students in Chicago (Aaronson et al. 2007), and 86 percent for math and 80 percent for ELA for elementary school students in San Diego (Koedel and Betts 2007).

Figure B.8. Percentage of Variation in Teacher Effectiveness Within Schools, English/Language Arts, 29 Districts, Years 1 to 3

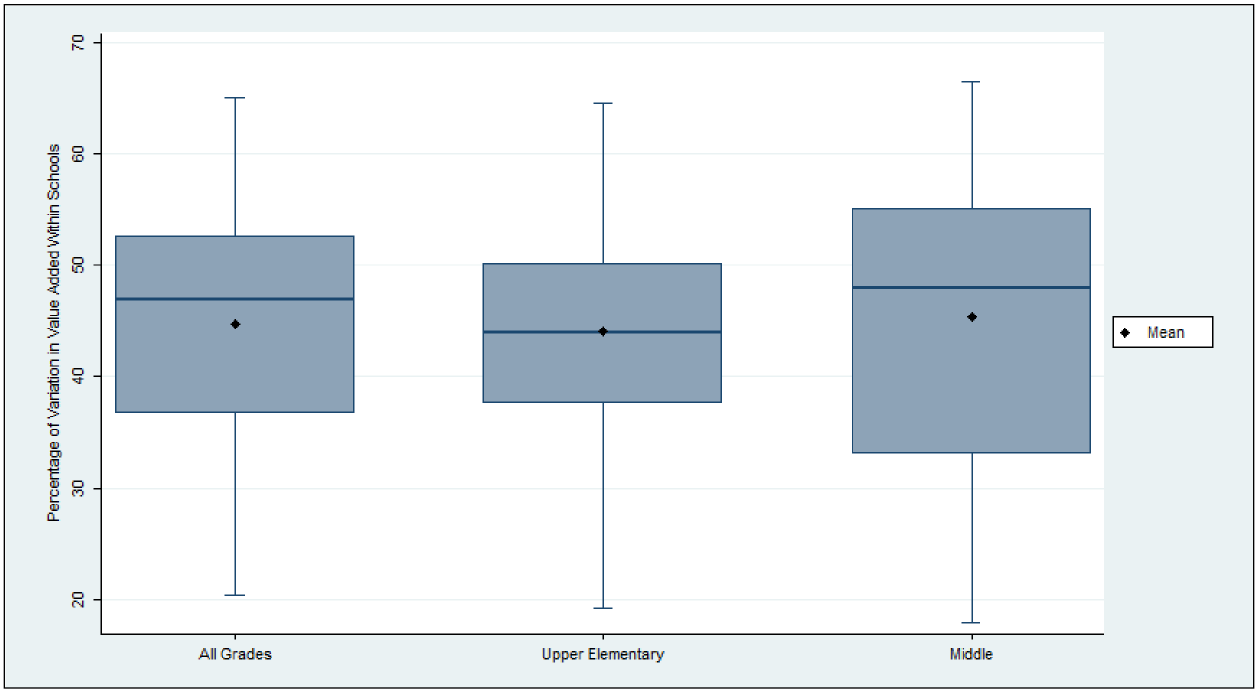


Source: District administrative data.

Note: Results are for 29 districts. District-level percentages are weighted across grades and years by the number of teachers. Overall percentages are weighted equally across districts. Percentage of variation is calculated after removing sampling variance from the value-added estimates through an empirical Bayes shrinkage procedure.

The level indicated by the line in the middle of each box shows the result for the district at the 50th percentile of the distribution of districts. The level indicated by the top of each box shows the result for the district at the 75th percentile. The level indicated by the bottom of each box shows the result for the district at the 25th percentile. The diamond indicates the result for the mean of all districts. The line segments outside the box show the results for the maximum and minimum districts.

Figure B.9. Percentage of Variation in Teacher Effectiveness Within Schools, Math, 29 Districts, Years 1 to 3



Source: District administrative data.

Note: Results are for 29 districts. District-level percentages are weighted across grades and years by the number of teachers. Overall percentages are weighted equally across districts. Percentage of variation is calculated after removing sampling variance from the value-added estimates through an empirical Bayes shrinkage procedure.

The level indicated by the line in the middle of each box shows the result for the district at the 50th percentile of the distribution of districts. The level indicated by the top of each box shows the result for the district at the 75th percentile. The level indicated by the bottom of each box shows the result for the district at the 25th percentile. The diamond indicates the result for the mean of all districts. The line segments outside the box show the results for the maximum and minimum districts.

3. Comparison of Value-Added Estimates Using Different Models

We compared the estimates from the value-added model to those from two alternative specifications by calculating the correlation between the resulting value-added estimates. To make these comparisons, we limited the value-added estimates to a common pool of teachers for whom it is possible to generate estimates using both sets of models. The two alternative models we estimated were: (1) a model that used pre-test scores from two prior years rather than from a single year (as in our main model), and (2) a model that accounted for three classroom-level peer characteristics: the classroom average pre-test, the percentage of FRL students in the classroom, and the standard deviation of pre-test scores in the class. Both of these alternative models (as well as our main model) provided measurement error correction in the pre-test using an errors-in-variables correction. The alternative models could further reduce potential bias in value-added estimates if students were matched to teachers on the basis of past pre-test scores or classroom characteristics.

The correlation between value-added estimates from the main model and those from the model using multiple years of pre-test scores was 0.98 for ELA and 0.99 for math (Table B.3). This aligns with what others have found when making the same comparison, such as Chetty et al.

(2011), who report a correlation of 0.98 for math. For the main model and a model including peer effects, the correlation in value-added estimates is 0.96 for ELA and 0.97 for math (Table B.4).

Table B.3. Correlation of Teacher Value-Added Estimates Between Main Model and Model Using Two Years of Pre-Test Scores, 29 Districts, Years 2 and 3

	Correlation of Value-Added Estimates	
Grade	English/Language Arts	Math
5	0.98	0.99
6	0.99	0.99
7	0.98	0.99
8	0.98	0.99
Grades 6 to 8	0.99	0.99
All grades	0.98	0.99

Source: District administrative data.

Note: Results are from Pearson correlations between the main model and a model with two years of pre-tests for 29 districts for years 2 and 3. District-level results are weighted across grades and years by the number of teachers. Overall results are weighted equally across districts. Results are calculated after removing sampling variance from the value-added estimates through an empirical Bayes shrinkage procedure.

Table B.4. Correlation of Teacher Value-Added Estimates Between Main Model and Peer Effects Model, 29 Districts, Years 1 to 3

	Correlation of Value-Added Estimates	
Grade	English/Language Arts	Math
4	0.99	0.99
5	0.98	0.99
6	0.96	0.96
7	0.94	0.97
8	0.95	0.97
Grades 4 to 5	0.98	0.99
Grades 6 to 8	0.95	0.96
All grades	0.96	0.97

Source: District administrative data.

Note: Results are from Pearson correlations between the main model and peer effects model for 9 districts for grades 4 through 8, and 14 districts for grades 6 to 8, for years 1 to 3. District-level results are weighted across grades and years by the number of teachers. Overall results are weighted equally across districts. Results are calculated after removing sampling variance from the value-added estimates through an empirical Bayes shrinkage procedure.

This page has been left blank for double-sided copying.

APPENDIX C

ADDITIONAL TABLES AND SENSITIVITY ANALYSES FOR CHAPTER IV

This page has been left blank for double-sided copying.

APPENDIX C: ADDITIONAL TABLES AND SENSITIVITY ANALYSES FOR CHAPTER IV

Chapter IV presented results for the Effective Teaching Gap (ETG) in a sample of 29 regionally diverse districts. This appendix supplements that analysis with tables that provide more detail, including results on the sensitivity of the findings to alternative models and estimation approaches.

A. Student Achievement Gap

In Chapter IV, we described how eliminating the ETG in a given year would reduce the student achievement gap between FRL and non-FRL students. We calculated the average student achievement gap in terms of standard deviations of student test scores, and then translated the difference between FRL and non-FRL students into percentile terms using a cumulative normal distribution function. Detailed information on average the student achievement gap in study districts is provided in Table C.1. On average, FRL students in grades 4 through 8 were 28 percentile points lower achieving than non-FRL students in English/language arts (ELA), and 26 percentile points lower achieving in math, with the achievement gap ranging from 10 to 42 percentile points in ELA across districts, and from 10 to 36 percentile points in math.

We also present the student achievement gap by grade level, including grade 3, which shows that the average student achievement gap differed by three percentile points across grades for ELA and two percentile points across grades for math. We used an ANOVA model to test the statistical significance of these differences across grades and found that they were not statistically significant for either subject (p-value=0.586 in ELA and p-value=0.821 in math). This indicates that in the study districts the student achievement gap was constant across grades 3 to 8, even while the average ETG is positive.

Finally, we show how the distribution of student achievement compares for FRL and non-FRL students. As suggested by average differences in student achievement, there were more FRL students in the lower range of the distribution of student achievement and more non-FRL students in the higher range. However, there was substantial overlap in the achievement distributions of FRL and non-FRL students. In other words, there were some students of both types throughout all parts of the distribution of student achievement. Figure C.1 shows the distribution of student achievement for ELA in the upper elementary grades and Figure C.2 shows the distribution for the middle school grades. Figures C.3 and C.4 show these distributions for math. Student test scores from all districts and all years are combined for each figure.

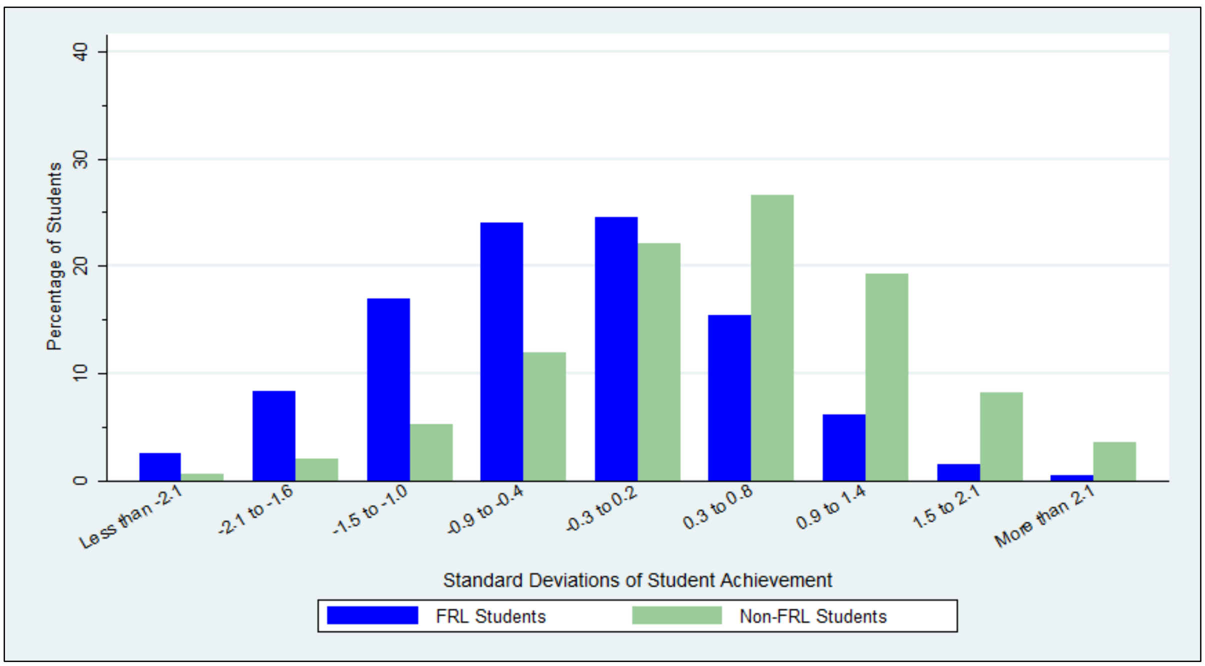
Table C.1. Student Achievement Gap for FRL and Non-FRL Students, Percentiles of Student Achievement, 29 Districts, Years 1 to 3

	English/Language Arts				Math			
Grade	Mean	Standard Deviation	Minimum	Maximum	Mean	Standard Deviation	Minimum	Maximum
3	28	7	11	39	25	7	10	38
4	29	7	13	39	26	7	12	39
5	28	7	10	40	26	7	10	36
6	28	7	11	42	26	7	10	36
7	27	9	8	50	25	7	9	37
8	26	7	10	40	24	7	8	36
Grades 4 to 8	28	7	10	42	26	7	10	36
Grades 3 to 8	28	7	10	41	26	7	10	36

Source: District administrative data.

Note: Results are for years 1 to 3. Grade-level student achievement gaps are calculated within each district-grade-year combination, weighted across grades and years by the number of students, and then averaged with equal weight across districts.

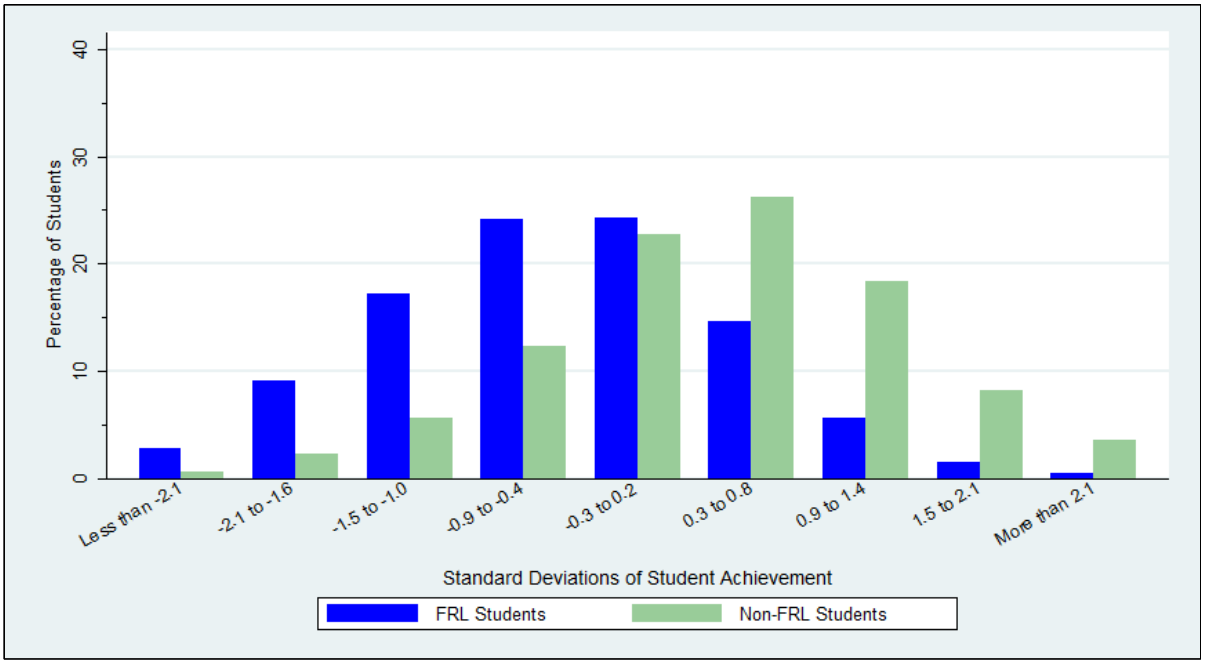
Figure C.1. Distribution of Student Achievement in English/Language Arts for FRL and Non-FRL Students, Upper Elementary Grades, 29 Districts, Years 1 to 3



Source: District administrative data.

Note: Results are for 29 districts, grades 4 and 5, and years 1 to 3. Test scores are standardized relative to state means and standard deviations for each district. Each student with an English/language arts post-test score contributes one observation per year.

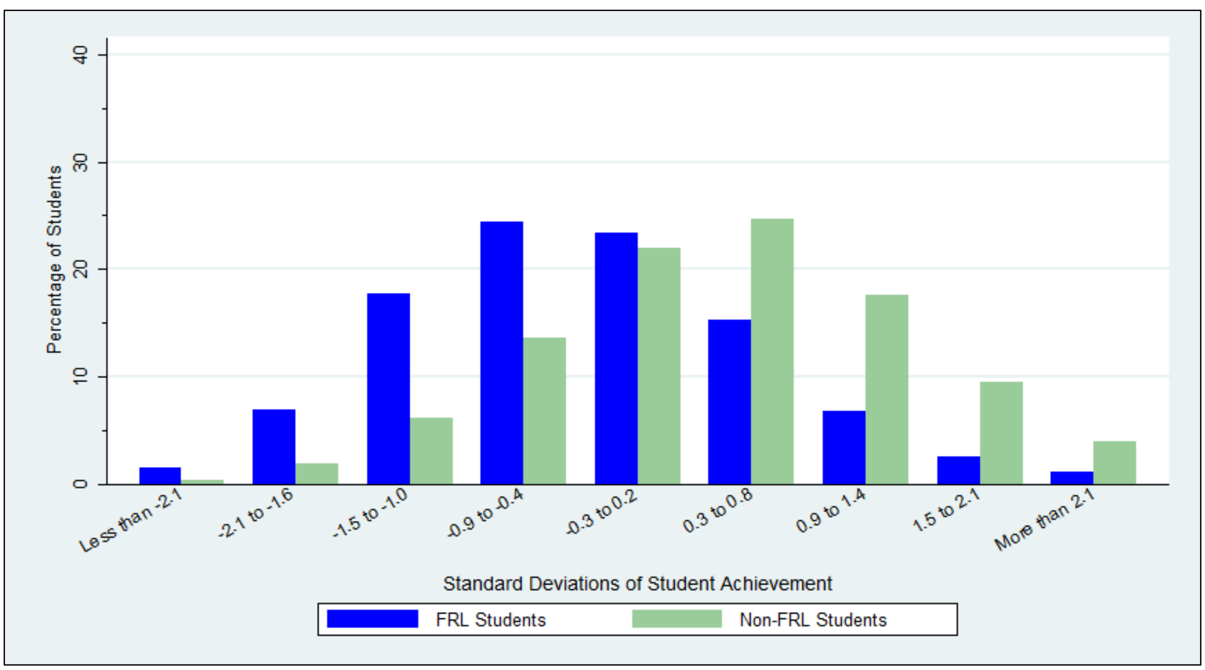
Figure C.2. Distribution of Student Achievement in English/Language Arts for FRL and Non-FRL Students, Middle School Grades, 29 Districts, Years 1 to 3



Source: District administrative data.

Note: Results are for 29 districts, grades 6 through 8, and years 1 to 3. Test scores are standardized relative to state means and standard deviations for each district. Each student with an English/language arts post-test score contributes one observation per year.

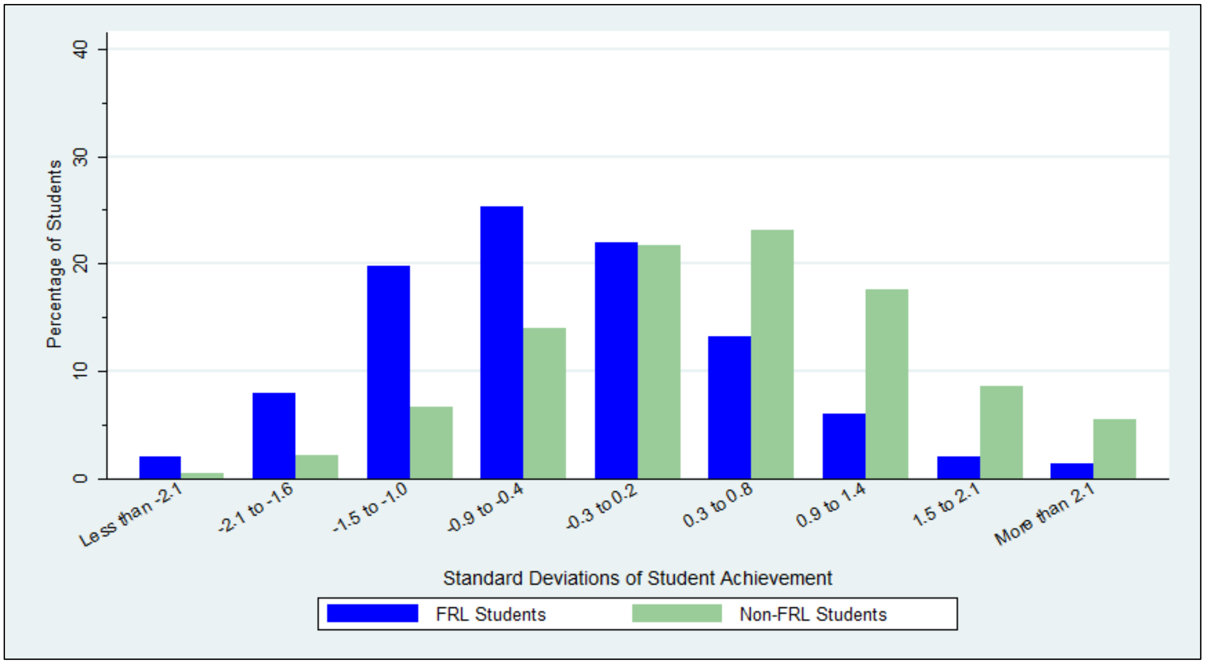
Figure C.3. Distribution of Student Achievement in Math for FRL and Non-FRL Students, Upper Elementary Grades, 29 Districts, Years 1 to 3



Source: District administrative data.

Note: Results are for 29 districts, grades 4 and 5, and years 1 to 3. Test scores are standardized relative to state means and standard deviations for each district. Each student with a math post-test score contributes one observation per year.

Figure C.4. Distribution of Student Achievement in Math for FRL and Non-FRL Students, Middle School Grades, 29 Districts, Years 1 to 3



Source: District administrative data.

Note: Results are for 29 districts, grades 6 through 8, and years 1 to 3. Test scores are standardized relative to state means and standard deviations for each district. Each student with a math post-test score contributes one observation per year.

B. Supplementary Information for ETG Results

We document in Chapter IV that there were no statistically significant differences by year in the average ETG for ELA or math across the first three years of the study, suggesting the absence of an overall upward or downward trend in the ETG. Table C.2 provides additional information on the stability of the estimates across years, including the number of districts for which we have ETG results each year and the number that have positive and significant results, statistically insignificant results, and negative and significant results. For the 25 districts for which we have ETG results in all three years, between 6 and 8 districts changed categories for ELA between any two pairs of years. For math, between 7 and 9 districts changed categories.

As mentioned in Chapter IV, the ETG equals 9 percent of the ratio between the value added of teachers at the 90th and 10th percentiles for ELA and 4 percent of this ratio for math. Table C.3 provides detailed information on the ratio by grade level, with results pooled over the three years of the study. For each grade, for the upper elementary and middle school grade spans, and for all grades, the table shows the average ETG, the average difference in the value added of teachers at the 90th and 10th percentiles of the distribution of value added, and the average ratio of the two (the ETG Ratio).

Table C.2. Statistical Significance of Effective Teaching Gap in Individual Districts, 29 Districts, Years 1 to 3

		Number of Districts			
Subject	Comparison	Total	Positive and Significant ETG	Statistically Insignificant ETG	Negative and Significant ETG
English/Language Arts	Year 3	29	22	7	0
	Year 2	28	22	6	0
	Year 1	25	20	5	0
	Years 1–3	**29**	**27**	**2**	**0**
Math	Year 3	29	12	16	1
	Year 2	28	15	13	0
	Year 1	25	14	10	1
	Years 1–3	**29**	**19**	**10**	**0**

Source: District administrative data.

Note: Results are for 29 districts, grades 4 through 8, and years 1 to 3.

The difference between teachers at the 90th and 10th percentiles was calculated by multiplying a standard deviation of teacher value added (that has been purged of sampling error in the estimates) by 2.56, the number of standard deviations between the 10th and 90th percentiles of a normal cumulative distribution function. This will closely correspond to the actual 90/10 gaps if teacher value-added estimates are approximately normally distributed. Because 90/10 differences in teacher value added are calculated this way, the ETG Ratios are proportional to ETGs calculated as a proportion of the standard deviation of teacher value added; that is, in terms of an effect size of teacher value added. As shown in Table C.3, on average, across all districts and grades, the 90/10 difference equals 0.39 standard deviations of student test scores in ELA and 0.54 standard deviations of student test scores in math.

Table C.3. Ratio of Effective Teaching Gap to 90th Percentile/10th Percentile Difference in Teacher Value Added, 29 Districts, Years 1 to 3

	English/Language Arts			Math		
Grade	Average Effective Teaching Gap	Average 90th–10th Percentile VA Score Difference	Ratio of Average ETG to 90/10 Difference	Average Effective Teaching Gap	Average 90th–10th Percentile VA Score Difference	Ratio of Average ETG to 90/10 Difference
4	0.042*	0.462	0.09	0.019*	0.649	0.03
5	0.038*	0.410	0.10	0.023*	0.590	0.04
6	0.040*	0.393	0.10	0.032*	0.541	0.06
7	0.028*	0.342	0.08	0.019*	0.460	0.04
8	0.022*	0.304	0.07	0.031*	0.450	0.06
Grades 4 to 5	0.040*	0.436	0.09	0.021*	0.620	0.03
Grades 6 to 8	0.030*	0.346	0.08	0.026*	0.484	0.05
All grades	0.034*	0.385	0.09	0.024*	0.540	0.04

Source: District administrative data.

Note: Results are for 29 districts and years 1 to 3. Grade-level ratios are computed within each district-grade-year combination, averaged with equal weight across years within each district, and then averaged with equal weight across districts.

*Indicates statistical significance at the 0.05 level, two-tailed test.

We used the ETG Ratio to compare access to effective teaching across grade spans. We used the ETG Ratio rather than the ETG because, as shown in Appendix B, the variance of teacher value-added estimates becomes tighter at higher grade levels. This leads to smaller estimated ETGs in middle school grades relative to elementary school grades. However, the variation in true teacher effectiveness was not necessarily lower in the middle school grades because the observed decrease in the standard deviation of teacher value added across grades may have been an artifact of the tests rather than a true compression of teaching ability in middle school. So, to capture whether gaps in effective teaching differ by grade span, we divided the ETG for each district-grade by the 90/10 difference in teacher value added for that district-grade to construct an ETG Ratio, and compared ETG Ratios in upper elementary school grades (4 and 5) versus middle school grades (6 through 8).

In ELA, the ETG Ratio was similar at the upper elementary and middle school levels. The ETG ratio was 0.091 for upper elementary grades and 0.082 for middle school grades (Table C.4). The ETG Ratios for the two grade spans were not significantly different from each other overall or in 21 of the 29 districts. The results were mixed for the other districts: 3 had significantly larger ETG Ratios in the middle school grades, and 5 had significantly larger ETG Ratios in the upper elementary grades.

Table C.4. Ratio of Effective Teaching Gap to 90/10 Difference (ETG Ratio) by Grade Span, 29 Districts, Years 1 to 3

Subject	Upper Elementary	Middle	Difference
English/Language Arts			
Average	0.091*	0.082*	0.009
Standard Deviation	0.050	0.055	n.a.
Minimum	0.018	0.011	n.a.
Maximum	0.210	0.210	n.a.
Math			
Average	0.035*	0.052*	-0.017*
Standard Deviation	0.043	0.057	n.a.
Minimum	-0.057	-0.035	n.a.
Maximum	0.137	0.184	n.a.
Sample Size (Districts)	**29**	**29**	**n.a.**

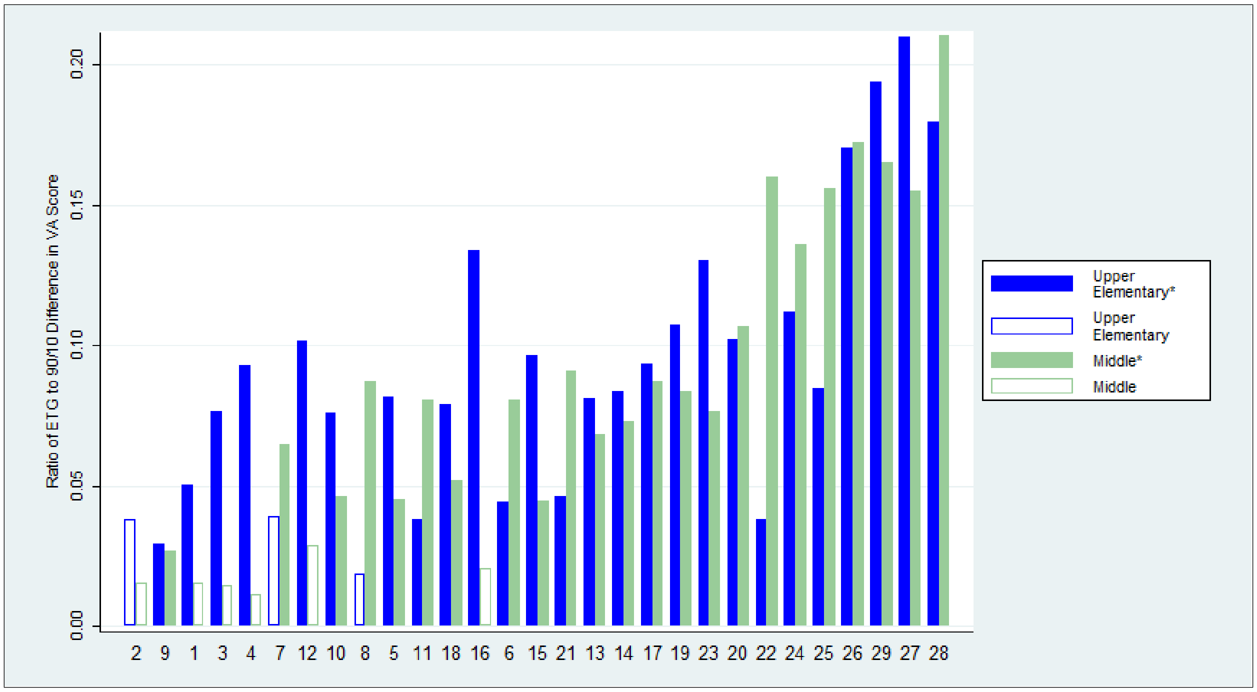
Source: District administrative data.

Note: Results are for 29 districts, grades 4 through 8, and years 1 to 3. District-level results are weighted across grades and years by the number of students. Overall results are weighted equally across districts. The difference in the ETG by grade span is the average of the differences for individual districts.

*Indicates statistical significance at the 0.05 level, two-tailed test. We test whether the average ETGs are statistically significant from zero using variation across districts, and test whether differences in the ETG across grade spans are statistically significant using variation within districts.

ETG Ratios in ELA were consistently positive and significant for both grade spans. The ETG Ratios were positive and significant in 26 of 29 districts in upper elementary grades and in 23 of 29 districts in middle school grades. In other districts, the ETG Ratios were not statistically significant. Figure C.5 shows results by upper elementary and middle school grade spans by district. Districts are ordered from lowest to highest by the ETG Ratio in ELA for all grades. ETG Ratios for upper elementary grades are shown in blue (dark shading); those for middle school grades are shown in green (light shading). Statistically significant results are indicated by solid bars; results that are not significantly different from zero are shown with hollow bars.

Figure C.5. English/Language Arts: Ratio of Effective Teaching Gap to 90/10 Difference by District and Grade Span, 29 Districts, Years 1 to 3



Source: District administrative data.

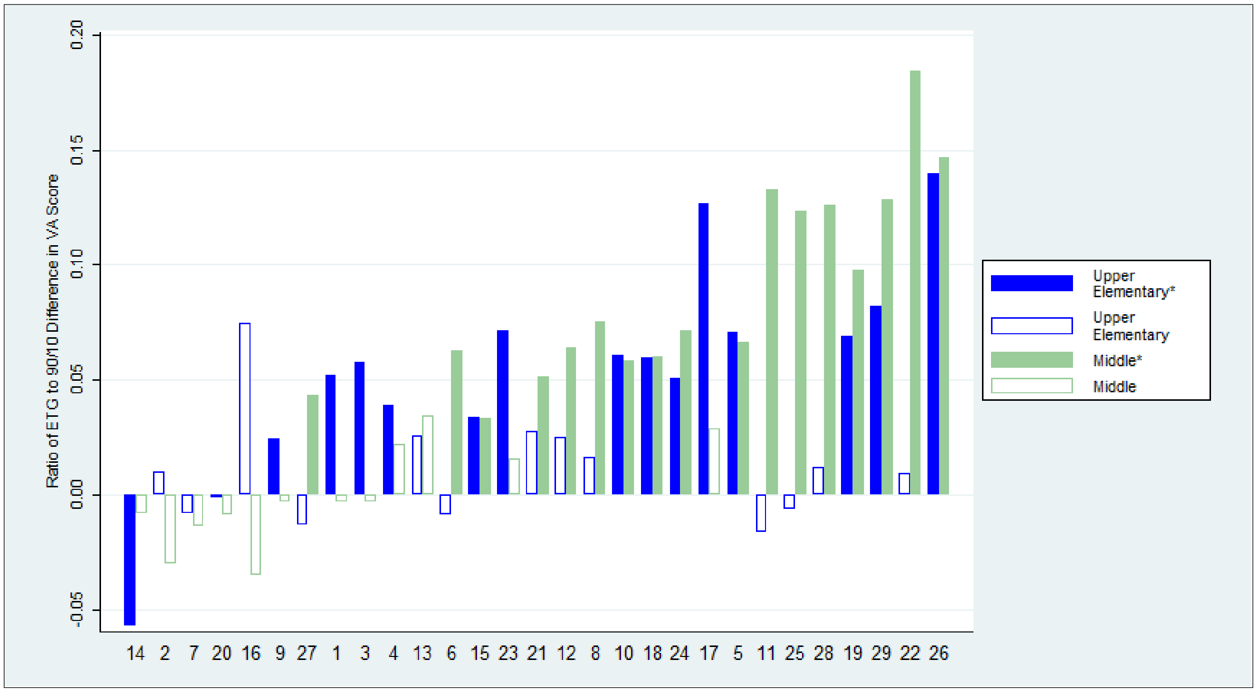
Note: Results are for 29 districts and years 1 to 3. Upper elementary is grades 4 and 5, middle school is grades 6 to 8. District-level results are weighted across grades and years by the number of students. Districts are ordered by the size of the ETG Ratio for all grades. District codes 1 to 29 are assigned in order of the size of the district ETG in English/language arts and are consistent across figures. The ETG Ratio is the ratio of the ETG to the difference in value added for teachers at the 90th and 10th percentiles. The solid bars show ETG Ratios that are significantly different from zero at the 0.05 level; the hollow bars are ETG Ratios that are not significantly different from zero.

*Indicates statistical significance at the 0.05 level, two-tailed test.

In math, ETG Ratios were significantly larger at the middle school level. The ETG Ratio was 0.052 in middle school grades, compared to 0.035 for upper elementary school grades (Table C.4). We found larger ETG Ratios at the middle school level for 6 districts, and larger ETG ratios at the upper elementary level for 5 districts, the other districts have differences by grade span that were not significant (18 districts). The difference in the ETG Ratio for the upper elementary and middle school grades is shown in Figure C.6 by the difference in height between the blue (dark shading) and green (light shading) bars.

The ETG Ratios for math tended to be positive and significant but were not universally so. In upper elementary grades, ETG Ratios in math were positive and significant in 14 districts, statistically insignificant in 14 districts, and negative and significant in 1 district. In middle school grades, ETG Ratios in math were positive and significant in 17 districts, statistically insignificant in 12 districts, and never negative and significant. District-by-district results for ETG Ratios are shown in Figure C.6.

Figure C.6. Math: Ratio of Effective Teaching Gap to 90/10 Difference by District and Grade Span, 29 Districts, Years 1 to 3



Source: District administrative data.

Note: Results are for 29 districts and years 1 to 3. Upper elementary is grades 4 and 5, middle school is grades 6 to 8. The ETG Ratio is the ratio of the ETG to the difference in value added for teachers at the 90th and 10th percentiles. District-level results are weighted across grades and years by the number of students. Districts are ordered by the size of the ETG Ratio for all grades. District codes 1 to 29 are assigned in order of the size of the district ETG in English/language arts and are consistent across figures. The solid bars show ETG Ratios that are significantly different from zero at the 0.05 level; the hollow bars are ETG Ratios that are not significantly different from zero.

*Indicates statistical significance at the 0.05 level, two-tailed test.

In ELA, districts with relatively large gaps in effective teaching in one grade span tended to have relatively large gaps in the other grade span, with a correlation coefficient between the grade spans of 0.53. In math, the correlation between grade spans is lower, with a correlation coefficient of 0.11. In math, this suggests that policies affecting gaps in effective teaching may be more differentiated across grade spans in math than in ELA. For example, a shift by some districts to tracking by ability level in middle school grades in math but not in ELA could produce both an overall jump in the ETG Ratio for math and a lower overall correlation between grade spans for math compared to ELA.

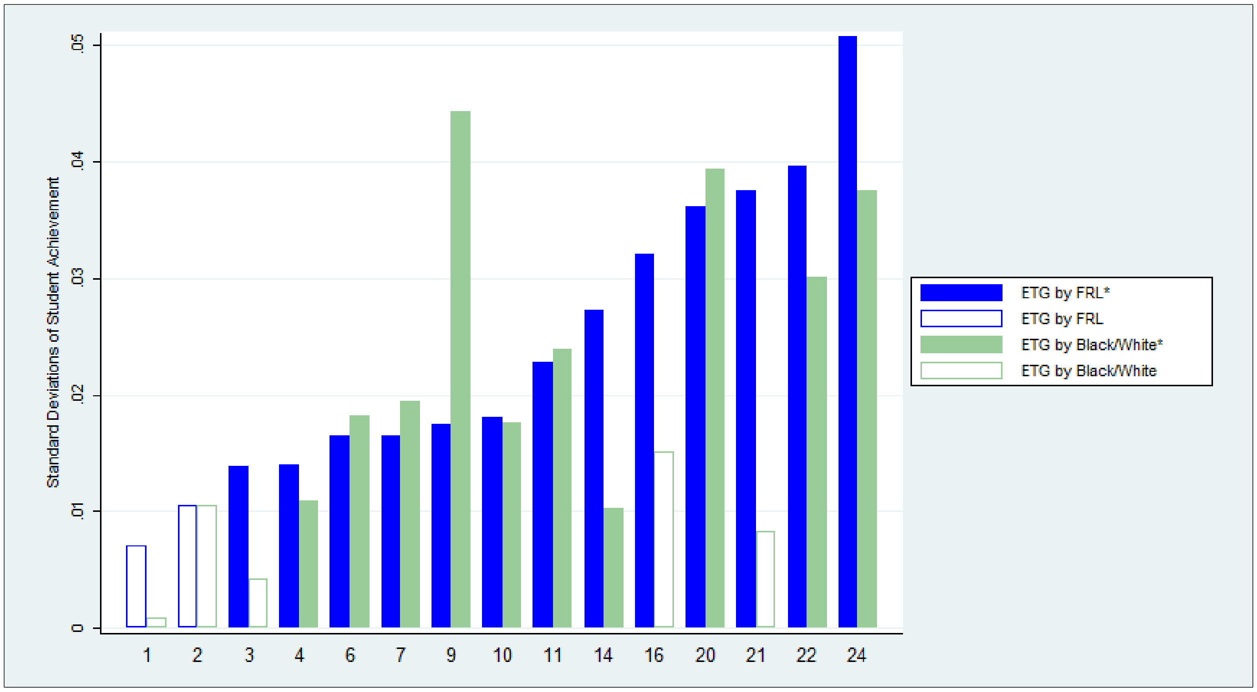
C. ETGs by Race and Ethnicity

We examined ETGs based on race and ethnicity in addition to presenting the gaps by FRL status. We restricted the sample of districts to those that have at least 15 percent White students and 15 percent students of one minority group. In the 15 study districts with at least 15 percent Black and 15 percent White students we found a Black/White ETG of 0.019 in ELA and 0.021 in math, compared to FRL ETGs of 0.024 in ELA and 0.015 in math in those districts. In the 18 study districts with at least 15 percent Hispanic and 15 percent White students, we found a Hispanic/White ETG of 0.033 in ELA and 0.029 in math, compared to FRL ETGs of 0.033 in ELA and 0.024 in math.

Individual district results showed the degree of similarity between ETGs based on FRL status and ETGs based on race. In ELA, for the 15 districts in this analysis, ETGs based on FRL were positive and significant for 13 districts and statistically insignificant for 2 districts. Black/White ETGs were positive and significant in 10 districts and statistically insignificant in 5. None of the districts had negative and significant ETGs. For math, ETGs based on FRL were positive and significant in 7 districts and statistically insignificant in 8. For Black/White ETGs, the 7 districts with positive and significant ETGs based on FRL also had positive and significant ETGs based on race. One other district also had a positive and significant ETG and one had a negative and significant ETG. District-level results are shown in Figures C.7 and C.8 in side-by-side comparisons of Black/White ETGs with ETGs based on FRL status. The height of each bar shows the overall ETG. Blue bars (dark shading) indicate the FRL gap, and corresponding green (light shading) bars show the Black-White gap. A solid bar means the ETG was significantly different than zero, and a hollow bar means it was not. One district (number 9) was an outlier in both subjects: in ELA, the Black/White ETG was 0.03 standard deviations of student test scores larger than the FRL-based ETG; in math, the Black/White ETG was 0.04 standard deviations larger than the FRL-based ETG.

Comparable information for Hispanic/White ETGs, presented in Figures C.9 and C.10, shows the overlap between FRL and race and ethnicity results. In ELA, shown in Figure C.9, all 18 districts had positive and significant gaps based on FRL, and 15 districts had positive and significant gaps based on Hispanic ethnicity. The other districts had statistically insignificant ETGs. For math, shown in Figure C.10, 12 districts had positive and significant gaps for both FRL and Hispanic ethnicity. Of the other 6 districts, none had a statistically significant FRL-based ETG, but one had a positive and significant gap for Hispanic ethnicity. District 9 was also an outlier in these results: in both subjects, the Hispanic/White ETG was 0.06 standard deviations of student test scores larger than the FRL-based ETG. Excluding District 9 from both analyses, the correlation in FRL and Black/White gaps increases from 0.52 to 0.70 in ELA and from 0.81 to 0.88 in math. Doing the same for Hispanic/White gaps, the correlation increases from 0.52 to 0.85 in ELA and from 0.75 to 0.89 in math.

Figure C.7. Comparison of Effective Teaching Gaps by FRL and Race, English/Language Arts, 15 Districts, Years 1 to 3

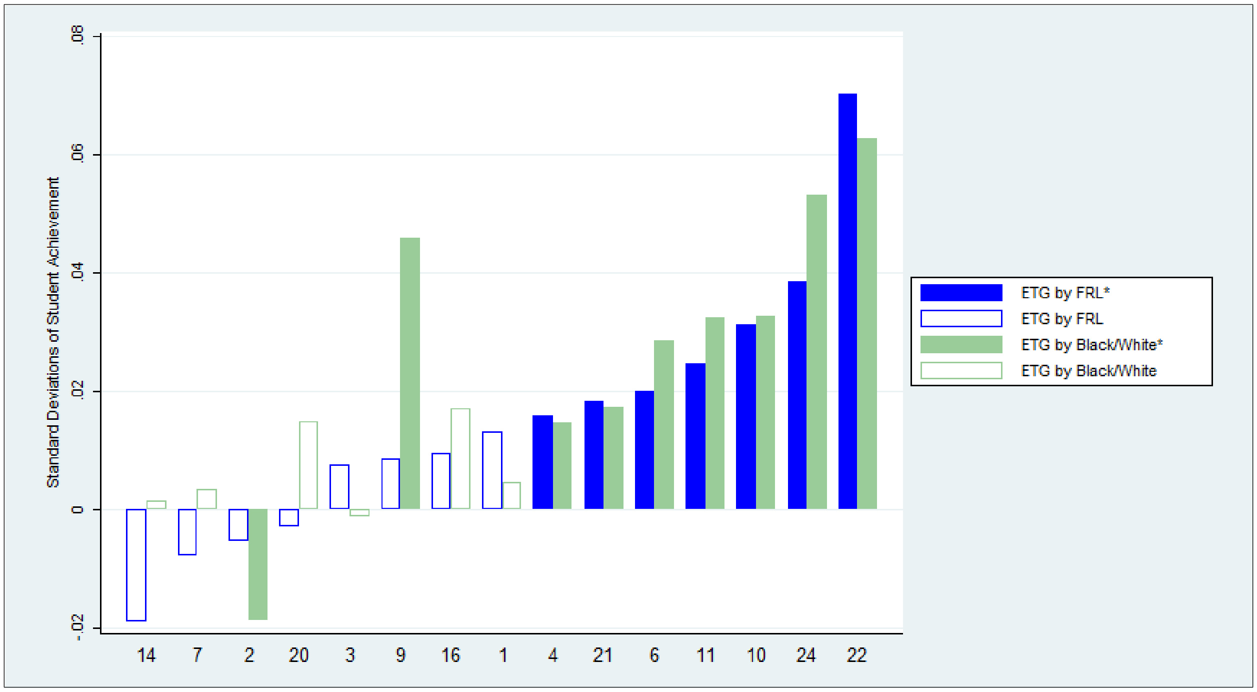


Source: District administrative data.

Note: Results are for 15 districts, grades 4 through 8, years 1 to 3. District-level results are weighted across grades and years by the number of students. Districts are ordered by the size of the FRL ETG. District codes 1 to 29 are assigned in order of the size of the district FRL ETG in English/language arts, and are consistent across figures. The solid bars show FRL or Black/White ETGs that are significantly different from zero at the 0.05 level; the hollow bars indicate ETGs that are not significantly different from zero. The ETG is expressed in terms of standard deviations of student test scores.

*Indicates statistical significance at the 0.05 level, two-tailed test.

Figure C.8. Comparison of Effective Teaching Gaps by FRL and Race, Math, 15 Districts, Years 1 to 3

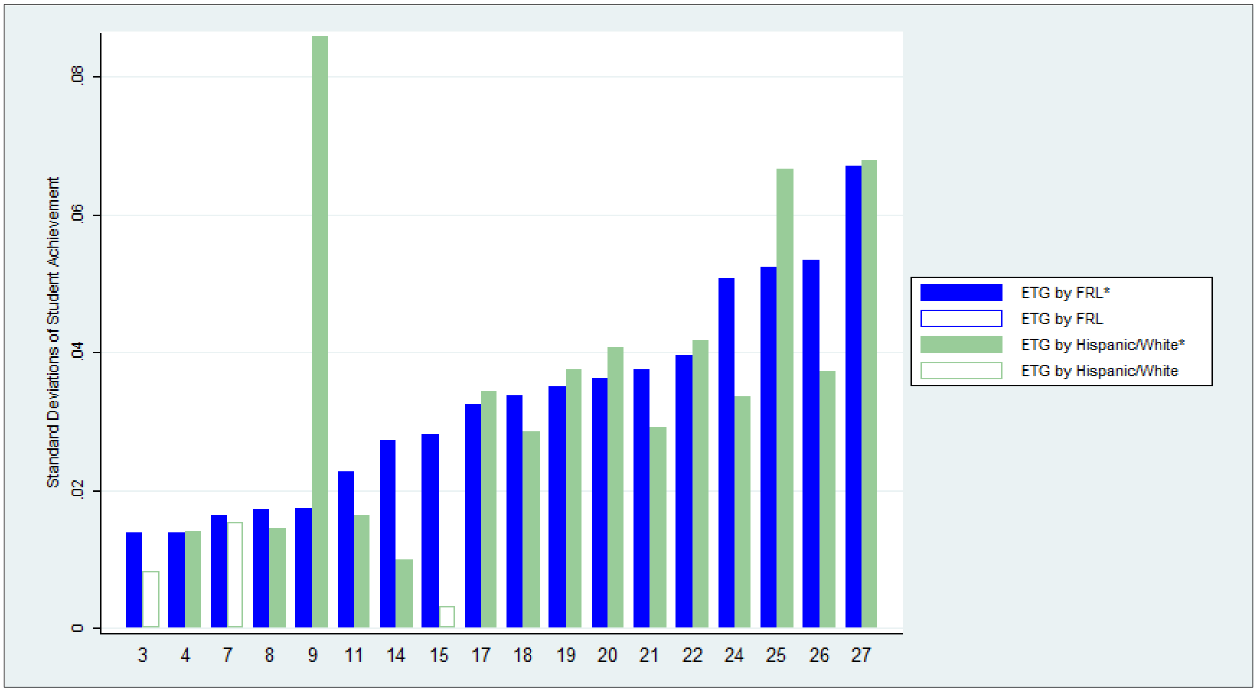


Source: District administrative data.

Note: Results are for 15 districts for grades 4 through 8. District-level results are weighted across grades and years by the number of students. Districts are ordered by the size of the FRL ETG. District codes 1 to 29 are assigned in order of the size of the district FRL ETG in English/language arts, and are consistent across figures. The solid bars show FRL or Black/White ETGs that are significantly different from zero at the 0.05 level; the hollow bars indicate ETGs that are not significantly different from zero. The ETG is expressed in terms of standard deviations of student test scores.

*Indicates statistical significance at the 0.05 level, two-tailed test.

Figure C.9. Comparison of Effective Teaching Gaps by FRL and Ethnicity, English/Language Arts, 18 Districts, Years 1 to 3

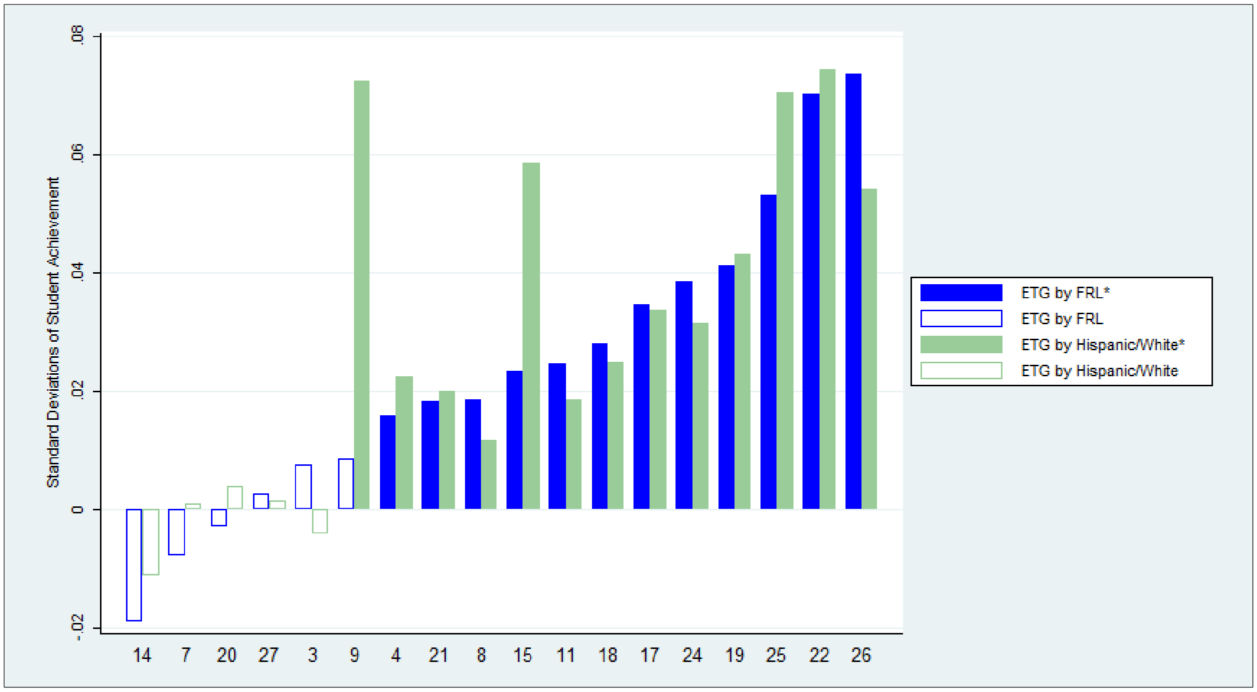


Source: District administrative data.

Note: Results are for 18 districts, grades 4 through 8, and years 1 to 3. District-level results are weighted across grades and years by the number of students. Districts are ordered by the size of the FRL ETG. District codes 1 to 29 are assigned in order of the size of the district FRL ETG in English/language arts, and are consistent across figures. The solid bars show FRL or Hispanic/White ETGs that are significantly different from zero at the 0.05 level; the hollow bars indicate ETGs that are not significantly different from zero. The ETG is expressed in terms of standard deviations of student test scores.

*Indicates statistical significance at the 0.05 level, two-tailed test.

Figure C.10. Comparison of Effective Teaching Gaps by FRL and Ethnicity, Math, 18 Districts, Years 1 to 3



Source: District administrative data.

Note: Results are for 18 districts for grades 4 through 8. District-level results are weighted across grades and years by the number of students. Districts are ordered by the size of the FRL ETG. District codes 1 to 29 are assigned in order of the size of the district FRL ETG in English/language arts, and are consistent across figures. The solid bars show FRL or Hispanic/White ETGs that are significantly different from zero at the 0.05 level; the hollow bars indicate ETGs that are not significantly different from zero. The ETG is expressed in terms of standard deviations of student test scores.

*Indicates statistical significance at the 0.05 level, two-tailed test.

D. Sensitivity Analyses

We conducted five sensitivity analyses of the main results presented in Chapter IV: (1) a comparison of the ETG results generated from the main model to two alternative calculations of the ETG that consider the possible impact of measurement error in FRL status; (2) a comparison of ETG results generated from the main model to those calculated from value-added estimates generated from two additional value-added models; (3) an examination of how the ETG in the urban core of countywide districts compares to the ETG across the district; (4) a test of whether the ETG results are sensitive to excluding the two districts in which over 20 of students are enrolled in schools eligible for Provision 2 or Provision 3; and (5) a test of whether the ETG results for middle school math are sensitive to excluding three districts in which students take end-of-course tests instead of end-of-grade tests.

Comparison to calculations that account for measurement error in FRL status. As discussed in Chapter 2, FRL status is measured with error, which could potentially lead us to understate the ETG. Here, we examined two alternative calculations of the ETG that consider the possible impact of measurement error in FRL status on the magnitude of the ETG. First, we considered using a statistical technique that adjusts for the imperfect measurement of an indicator used as a covariate in a regression model—in this case, when the families of some students identified as FRL actually earn more than 185 percent of the poverty line and when families of some students identified as non-FRL actually earn less than that amount. Second, we recalculated the ETG removing students who are eligible for a reduced-price lunch in the 19 study districts for which we have information on reduced-price lunch eligibility as well as free-lunch eligibility.

To assess the consequences of measurement error in FRL status on the ETG, we used a statistical technique developed by labor economists to measure the effect of union membership on wages when data on union status are sometimes mismeasured (Aigner 1973; Freeman 1984) and applied to other settings with similar data issues (Savoca 2000). The method relies on knowing the percentage of the sample that is mismeasured in both directions, and then multiplying the coefficient on the mismeasured variable by the following adjustment factor:

$$\text{(C.1)} \qquad [\tilde{P}\left(1-\tilde{P}\right)]/[P\left(1-P\right)\left(1-r_0-r_1\right)]$$

where P equals the proportion of students who truly meet the FRL criteria, $\tilde{P}$ equals the proportion of students measured as meeting FRL criteria, r_0 is the "false positive rate," or the proportion of non-FRL students incorrectly classified as FRL, and r_1 is the "false negative rate," or proportion of FRL students incorrectly classified as non-FRL.[47] As mentioned in Chapter 2, Ponza et al. (2007) found that 9.1 percent of FRL students who are classified as FRL are not truly eligible for FRL based on family income. This differs from the false negative rate, which asks what percentage of students who are truly FRL are incorrectly classified as non-FRL. However, we can use this estimate along with an estimate of the percentage of students who are classified as non-FRL but are truly FRL-eligible to back out the false positive and false negative rates required to compute (C.1). Given these two estimates, the adjustment factor does not depend on the proportion of students in both groups.

There is no estimate from the research literature on the percentage of students who are classified as non-FRL but are truly FRL-eligible, so we apply a range of plausible assumptions: 5 percent, 10 percent, and 20 percent. Doing so yields the result that the ETG may be underestimated by between 16 to 41 percent, as shown in Table C.5. Revised estimates of the ETG range from 0.040 to 0.048 for ELA and 0.028 to 0.034 in math.

[47] We follow the notation used in Savoca (2000), where a derivation of the adjustment factor can be found.

Table C.5. Comparing Effective Teaching Gaps With and Without Correction for Measurement Error in FRL Status, 29 Districts, Years 2 and 3

	Adjustment Factor	ELA ETG	Math ETG
Without Correcting for Measurement Error in FRL	1.00	0.034	0.024
Assuming 5 percent of non-FRL students are truly FRL-eligible	1.16	0.040	0.028
Assuming 10 percent of non-FRL students are truly FRL-eligible	1.24	0.042	0.030
Assuming 20 percent of non-FRL students are truly FRL-eligible	1.41	0.048	0.034

Source: District administrative data.

Note: For each of the three rows correcting for measurement error, we assume that 9.1 percent of FRL students who are classified as FRL are not truly eligible for FRL based on family income.

There are three reasons to believe that these estimates are upper bounds of the effect of measurement error in FRL status on the ETG. First, the misclassification rates reported in Ponza et al. (2007) were based on a broader sample of districts than the study districts, which, as noted in Chapter 3, have higher FRL rates than districts nationally or even than the 100 largest districts. The higher the proportion of students who are identified as FRL, the lower the error rates in certification are likely to be (Ponza et al. 2007). Second, the measurement error adjustment has been applied only to the ETG calculation but not to the FRL indicator in the value-added models on which the ETG is constructed. Were we to have taken this step, the value-added estimates of teachers of FRL students would likely have increased while the value-added estimates of teachers of non-FRL students would likely have decreased, leading to a smaller ETG prior to adjusting it for measurement error. Third, unlike the context of union membership in which the measurement error correction methodology was devised, a student who is misclassified is likely to be close to the margin of being FRL or non-FRL. In other words, although a binary variable, FRL status is derived from an underlying continuous distribution of income, unlike union membership, which is a discrete yes/no indicator.

Due to these concerns about this measurement error correction technique, we followed an alternate strategy for testing the sensitivity of the results to measurement error: we redefined "disadvantaged" to mean free-lunch eligible, excluding students eligible for a reduced-price lunch. In a sample of 19 study districts for which this distinction was available, we compared free-lunch eligible students to students ineligible for a free or reduced-price lunch. This excludes an average of 8.0 percent of students who are eligible for a reduced-price lunch in these districts. The rationale for excluding students eligible for a reduced-price lunch is that this group of students is more error-prone—that is, they are more likely than students eligible for a free lunch to have family income above the 185 percent of the poverty line threshold (Ponza et al. 2007).

In these 19 districts, the ETG when defining disadvantaged students as those eligible for free or reduced-price lunch is 0.032 for ELA and 0.024 for math. When we exclude reduced-price lunch students from the classification of disadvantaged, the ETG is 0.034 for ELA and 0.025 for math.

Comparison to alternative value-added models. We tested the sensitivity of the results to two alternative versions of our value-added model. In the first alternative model, we accounted for an extra year of student pre-test scores. In the second alternative model, in addition to individual student characteristics, we accounted for three measures of the characteristics of a student's classroom peers: the classroom average pre-test score, the percentage of FRL students in the classroom, and the standard deviation of pre-test scores in the class.

As explained in Chapter II, if the pre-tests are imperfect measures of student achievement, value-added estimates can be biased in ways that could lead to ETGs that are too large. We can potentially minimize this bias, however, by including additional years of student pre-tests. So we compared ETGs based on the main model to ETGs based on a model that accounts for an additional year of pre-test scores as a way of further accounting for measurement error in pre-tests, beyond the errors-in-variables adjustment we performed for prior-year pre-test scores in the main model. Because two years of pre-test data were not available for year 1 or for grade 4 in any year, we limited the sample for this analysis for both models to grades 5 to 8 in years 2 and 3. All districts were included in the analysis.

Differences in the ETGs based on the two value-added models were statistically significant in ELA but the differences were 0.007 in ELA and 0.004 in math. The model with two years of pre-tests produced ETGs of 0.025 in ELA and 0.020 in math, compared with 0.031 and 0.024 for the main model with a single year of pre-tests (Table C.6). The last row of Table C.6 indicates that district-by-district results were correlated at 0.98 for ELA and 0.99 for math. Figures C.11 and C.12 compare district-level results for the two models for ELA (C.11) and math (C.12). As in Figures IV.1 and IV.2 in Chapter IV, we present point estimates and 95 percent confidence intervals for the main model (restricted to the common grades and years). We also show the point estimates of the alternative value-added model. As can be seen in these figures, the point estimates for ELA were within the 95 percent confidence interval of the original estimates for 25 of the 29 districts in ELA and for all 29 districts in math.

Table C.6. Comparing Effective Teaching Gaps Based on Main Model and Model Using Two Years of Pre-Test Scores, 29 Districts, Years 2 and 3

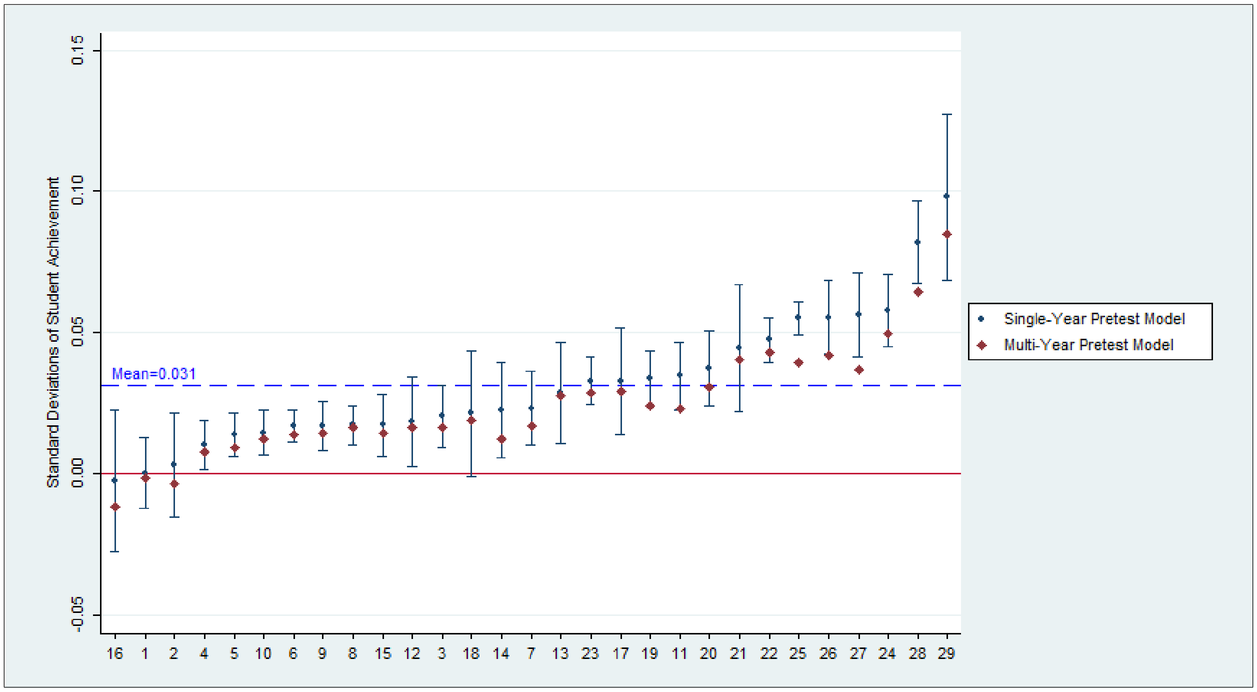
	English/Language Arts			Math		
	Main Model	Model Using Two Years of Pre-Test Scores	Difference Between Two Models	Main Model	Model Using Two Years of Pre-Test Scores	Difference Between Two Models
Cross-district Average	0.031*	0.025*	0.007*	0.024*	0.020*	0.004
Standard Deviation	0.023	0.020		0.026	0.025	
Correlation	0.98			0.99		

Source: District administrative data.

Note: Results are from the main model and a model using two years of pre-test scores for 29 districts, grades 5 to 8, and years 2 and 3. District-level results are weighted across grades and years by the number of students. Overall results are weighted equally across districts. Differences between the two models are based on the average of the differences for individual districts.

*Indicates statistical significance at the 0.05 level, two-tailed test. We test whether the average ETGs for each model are statistically significant from zero using variation across districts, and test whether differences between the two models are statistically significant using variation within districts.

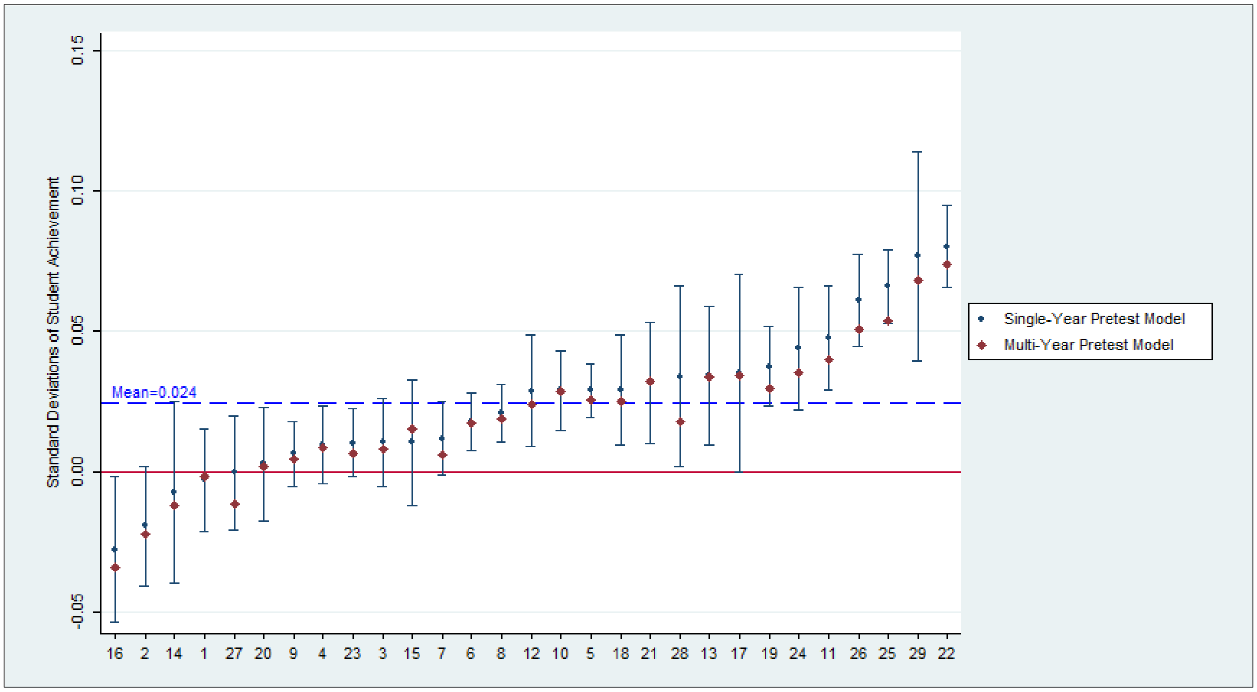
Figure C.11. Effective Teaching Gaps Based on Main Model and Model Using Two Years of Pre-Test Scores, English/Language Arts, 29 Districts, Years 2 and 3



Source: District administrative data.

Note: Results are for 29 districts, grades 5 to 8. District-level results are weighted across grades and years by the number of students. Circles represent the district-level ETGs from the main model and the vertical lines show the 95 percent confidence intervals around each point. The red diamonds represent the district-level ETGs from the multi-year pre-test model. Districts are ordered by the size of the ETGs for the main model. District codes 1 to 29 are assigned in order of the size of the district ETG in English/language arts for the main model, and are consistent across figures. The ETG is expressed in terms of standard deviations of student test scores.

Figure C.12. Effective Teaching Gaps Based on Main Model and Model Using Two Years of Pre-Test Scores, Math, 29 Districts, Years 2 and 3



Source: District administrative data.

Note: Results are for 29 districts, grades 5 to 8. District-level results are weighted across grades and years by the number of students. Circles represent the district-level ETGs from the main model and the vertical lines show the 95 percent confidence intervals around each point. The red diamonds represent the district-level ETGs from the multi-year pre-test model. Districts are ordered by the size of the ETGs for the main model. District codes 1 to 29 are assigned in order of the size of the district ETG in English/language arts for the main model, and are consistent across figures. The ETG is expressed in terms of standard deviations of student test scores.

We also examined the sensitivity of the ETG estimates from the main model to those produced by an alternative value-added model that includes peer effects. Because the estimation of a peer effects model depends on the ability to link teachers to multiple classrooms, we restricted this analysis to 23 districts, including grades 4 through 8 for 9 districts and grades 6 to 8 for the other 14 districts.

The inclusion of peer effects in the value-added model led to differences in the calculation of the average ETG. As shown in Table C.7, average ETGs based on the peer effects model were statistically different from those based on the main model by 0.022 in both subjects. The peer effects model produced ETGs of 0.006 in ELA and 0.002 in math, compared with 0.029 and 0.024 for the main model. The correlation across districts of ETGs based on the two value-added models, as shown in the bottom row of Table C.7, was 0.34 for ELA and 0.57 for math. However, when excluding one outlier district (District 27), the correlations increased to 0.52 in ELA and 0.62 in math. Figures C.13 and C.14 present point estimates and 95 percent confidence intervals of ETG estimates for each district based on the main model (restricted to the common grades and years) as well as the alternative peer effects model. As can be seen in these figures, the point estimates were within the 95 percent confidence interval of the original estimates for 10 of the 23 districts in ELA and for 14 of the 23 districts in math.

The results of the peer effects model may be influenced by the particular peer effects specification. As described by Hoxby and Weingarth (2006), there are a variety of ways that classroom characteristics may be specified in a peer effects model. For example, there may be a nonlinear relationship between classmates' prior achievement and a given student's achievement, as opposed to the linear specification estimated here. To better understand the influence of peers on disadvantaged students' access to effective teaching, we plan to estimate alternative specifications of the peer effects model in the study's final report. The data we analyze for the final report will include five years of value-added estimates, which will provide additional variation in classroom characteristics for a given teacher.

Table C.7. Comparing Effective Teaching Gaps Based on the Main and Peer Effects Models, 23 Districts, Years 1 to 3

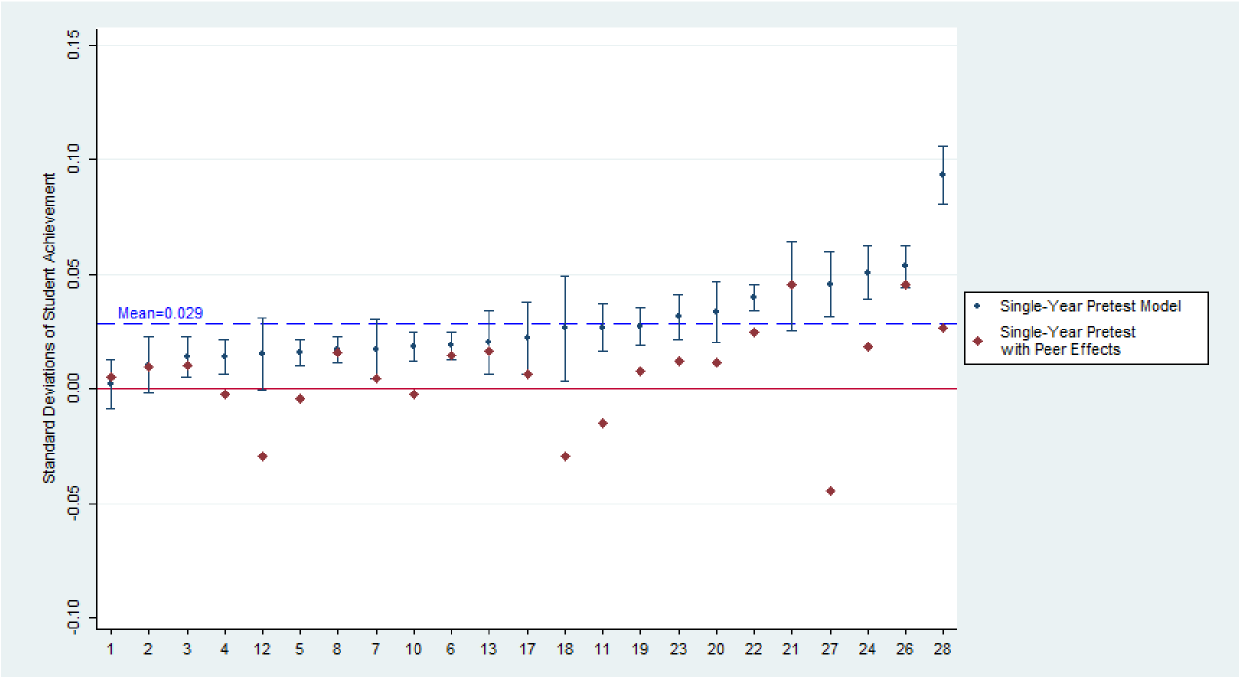
	English/Language Arts			Math		
	Main Model	Peer- Effects Model	Difference Between Two Models	Main Model	Peer- Effects Model	Difference Between Two Models
Cross-district Average	0.029*	0.006	0.022*	0.024*	0.002	0.022*
Standard Deviation	0.019	0.022		0.023	0.034	
Correlation	0.34			0.57		

Source: District administrative data.

Note: Results are from the main model and peer effects model for 23 districts. District-level results are weighted across grades and years by the number of students. Overall results are weighted equally across districts. Teachers of grades 6 through 8 in 23 districts and for grades 4 and 5 in 9 districts are included in the calculation of value added.

* The difference between models is statistically significant at the 0.05 level, two-tailed test.

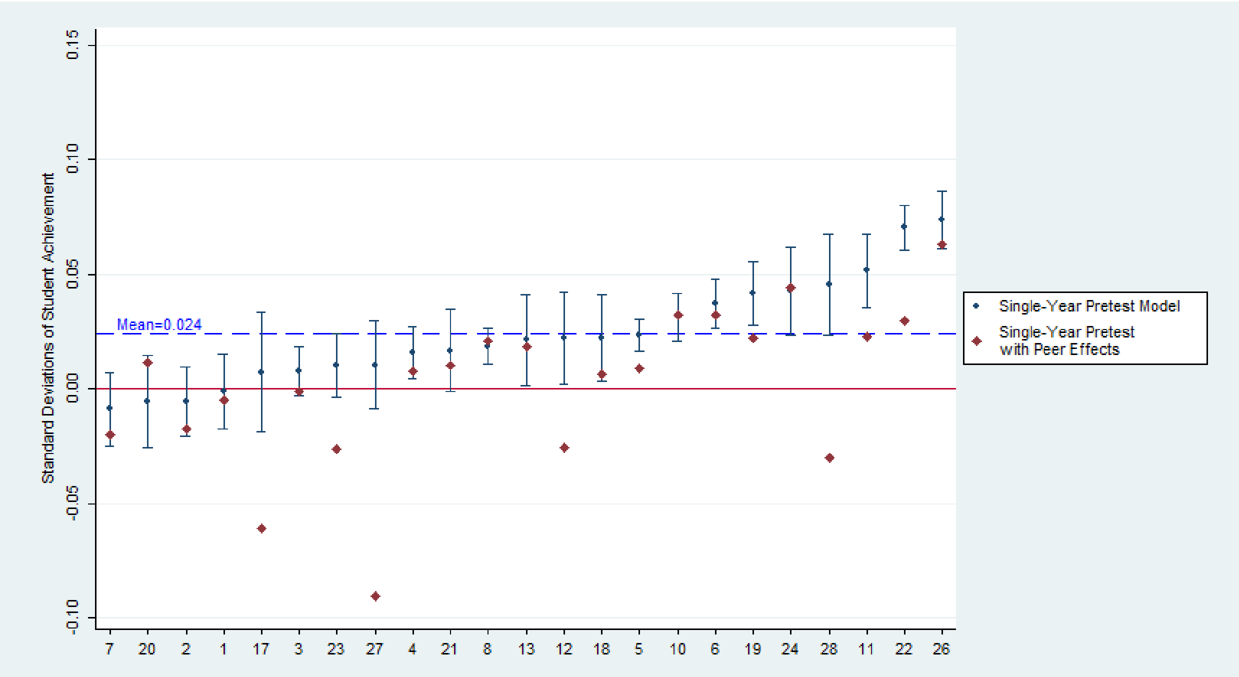
Figure C.13. Effective Teaching Gaps Based on Main and Peer Effects Models, English/Language Arts, 23 Districts, Years 1 to 3



Source: District administrative data.

Note: Results are from the main model and peer effects model for grades 6 through 8 for 23 districts and grades 4 and 5 for 9 districts. District-level results are weighted across grades and years by the number of students. Circles represent the district-level ETGs from the main model and the vertical lines show the 95 percent confidence intervals around each point. The red diamonds represent the district-level ETGs from the multi-year pre-test model. Districts are ordered by the size of the ETGs for the main model. District codes 1 to 29 are assigned in order of the size of the district ETG in English/language arts for the single year pre-test model, and are consistent across figures. The ETG is expressed in terms of standard deviations of student test scores.

Figure C.14. Effective Teaching Gaps Based on Main and Peer Effects Models, Math, 23 Districts, Years 1 to 3



Source: District administrative data.

Note: Results are from the main model and peer effects model for grades 6 through 8 for 23 districts and grades 4 and 5 for 9 districts. District-level results are weighted across grades and years by the number of students. Circles represent the district-level ETGs from the main model and the vertical lines show the 95 percent confidence intervals around each point. The red diamonds represent the district-level ETGs from the multi-year pre-test model. Districts are ordered by the size of the ETGs for the main model. District codes 1 to 29 are assigned in order of the size of the district ETG in English/language arts for the single year pre-test model, and are consistent across figures. The ETG is expressed in terms of standard deviations of student test scores.

Comparison of the ETG for urban schools to the ETG for the whole district. We also examined how the ETG in the urban core of countywide districts compared to the ETG across the district. This was a means to gauge the external validity of our results, which mostly included districts defined by city boundaries; competing school districts (for which data were not available) are located in suburban areas. If highly effective teachers can easily move between an urban core district and the surrounding suburban districts, then the relevant area for measuring ETGs from the perspective of families may include the urban core and surrounding districts.[48] In this broader area, the ETGs may have been larger (that is, more disadvantageous to FRL students) than might have been suggested by those based on the urban area alone. We identified 5 county-wide districts in our sample that contained between 36 percent and 66 percent of students attending schools in urban areas, according to the 2008–09 Common Core of Data (CCD). We compared the ETGs for these districts to those for the subset of schools located in urban areas (in the same districts) to test the sensitivity of our results.

[48] This contrasts with the perspective taken in this study, which is the perspective of school districts.

On average, the ETG for schools in the urban core did not differ from the ETG for all schools by more than 0.005 standard deviations of student test scores in either subject across these 5 county-wide districts. The overall ETG in ELA was 0.017 on average for all schools in these districts and 0.018 for schools in the urban core, and the math ETG was 0.021 for all schools and 0.024 for those in the urban core. The between- and within-school ETGs did not differ by more than 0.003 standard deviations in either subject when including only students in the urban core. This suggests that the ETGs in our urban-only districts were not omitting an important dimension of inequity in the distribution of teachers. However, because urban and suburban areas within a countywide district were subject to the same salary schedule and other teacher policies, the results may not fully generalize to settings in which urban districts are surrounded by competing suburban districts with different policy environments.

Comparison when excluding districts with more than 20 percent of students in Provision 2 or Provision 3 schools. The administrative data provided by two districts did not include accurate FRL data for more than 20 percent of students enrolled in schools participating in Provision 2 or Provision 3. As described in Appendix A, these schools offered meals free to all students, without determining the students' actual eligibility for these benefits based on their household circumstances. All students in these schools may appear as FRL, even though not all students met the eligibility criteria. As explained in Appendix A, we imputed FRL status for students at these schools. The overall average ETGs in ELA and math changed by 0.001 or less when excluding these two districts. We also examined the relative size of the ETGs for these two districts and found that one had an ETG in the top 25 percent of districts for ELA and math, while the other district had an ETG near the median for ELA and in the bottom 25 percent of districts for math.

Comparison when excluding districts with multiple math tests. Three of the districts had middle school students take end-of-course tests rather than end-of-grade tests for math. As a result, middle school students in the same grade could have taken different math tests. We measured value added separately for the different math tests and then equated the value-added estimates of teachers across the two tests. To test the sensitivity of our results to this approach, we recalculated the middle school math ETGs when excluding these districts. The middle school math ETG across districts differed by less than 0.01 when excluding the three districts.

DEPARTMENT OF EDUCATION
UNITED STATES OF AMERICA