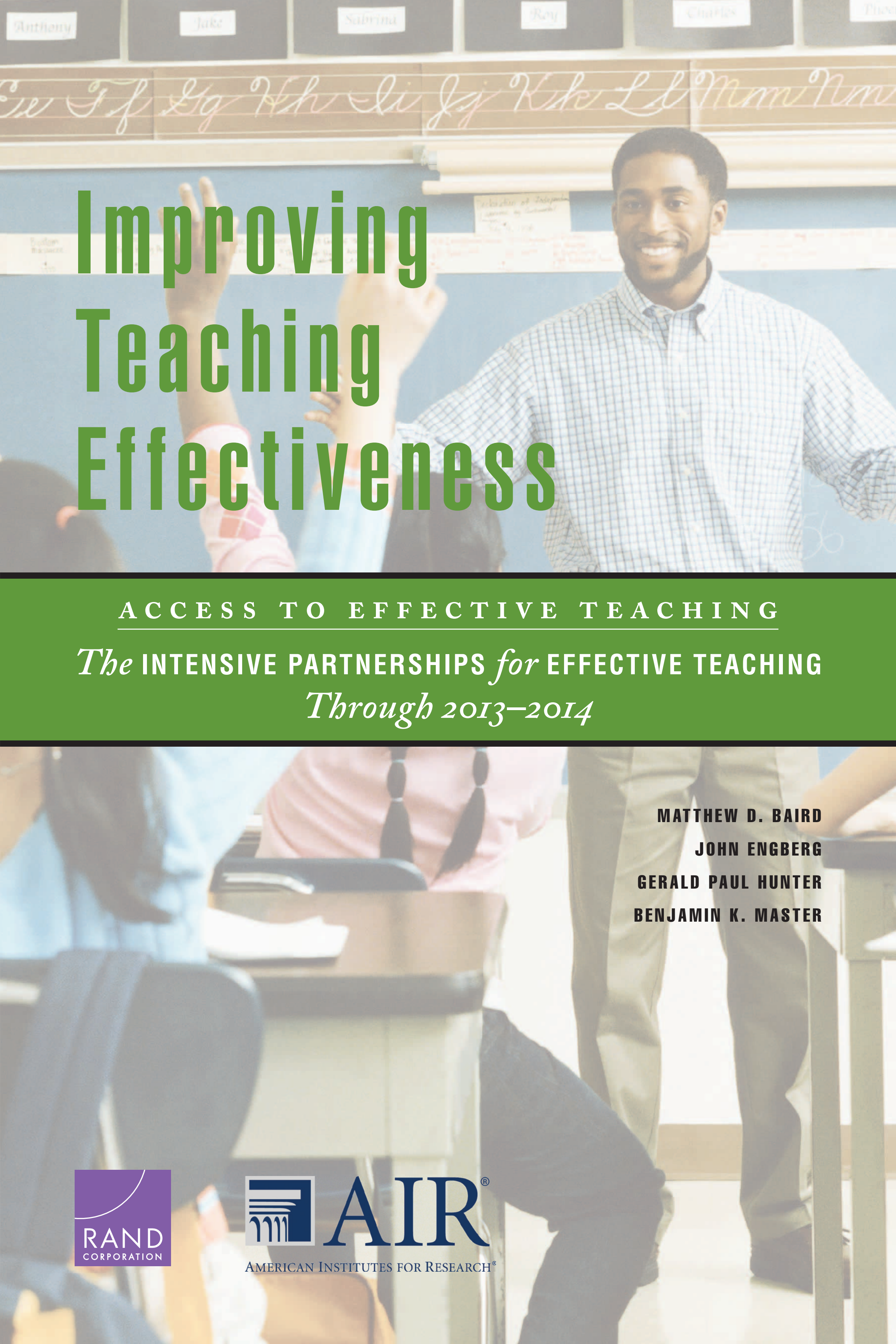


Improving Teaching Effectiveness

ACCESS TO EFFECTIVE TEACHING

The INTENSIVE PARTNERSHIPS *for* EFFECTIVE TEACHING
Through 2013–2014

MATTHEW D. BAIRD
JOHN ENGBERG
GERALD PAUL HUNTER
BENJAMIN K. MASTER

RAND CORPORATION

AIR®
AMERICAN INSTITUTES FOR RESEARCH®

For more information on this publication, visit www.rand.org/t/RR1295z4

Library of Congress Cataloging-in-Publication Data is available for this publication.
ISBN: 978-0-8330-9561-9

Published by the RAND Corporation, Santa Monica, Calif.

Cover: Teacher Standing in Front of a Class of Raised Hands, Digital Vision.

Preface

The Bill & Melinda Gates Foundation launched the Intensive Partnerships for Effective Teaching in school year 2009–2010. After careful screening, the foundation identified seven Intensive Partnership sites—three school districts and a cluster of four charter management organizations (CMOs)—to implement strategic human-capital reforms over a six-year period.[1] The foundation also selected the RAND Corporation and its partner, the American Institutes for Research (AIR), to evaluate the Intensive Partnerships efforts. The RAND/AIR team is conducting three interrelated studies examining the reforms' implementation, the reforms' effect on student outcomes, and the extent to which the reforms are replicated in other districts. The evaluation began in July 2010 and collected its first wave of data during the 2010–2011 school year; it will continue through the 2015–2016 school year and produce a final report in 2017. During this period, the RAND/AIR team is producing a series of internal progress reports for the foundation and the Intensive Partnership sites.

The present report is the first public report on the relationship between teachers' value-added estimates in mathematics and reading and the demographic characteristics of the students they serve.[2] We refer to this relationship as the sorting of teachers among students—i.e., the assignment of teachers to schools and to classrooms composed

1 We use the word *site* to describe the three school districts and the four CMOs that received funding from the foundation to implement the Intensive Partnerships initiative.

2 *Teacher value added* refers to statistical estimates of teachers' contributions to growth in student test scores.

of various demographics of students. We analyze data through school year 2013–2014 and focus on sorting that affects the access that low-income minority students have to teachers of differing levels of value added. The report presents sorting patterns for three participating districts in the three school years prior to implementation of the Intensive Partnerships intervention and the four years following the start of the intervention. In the appendix, we also include the patterns for one of the CMOs but caution the reader that the findings for this site are very imprecise because the site is so small.

In addition to examining the sorting of teachers' value-added estimates in mathematics and reading using a common model for all sites, the report examines the sorting of teachers by the sites' own achievement growth measures and by the sites' composite effectiveness measures. The composite measures are based not only on students' test performance but also on classroom observations and other measures.

We intend the report not only to provide feedback to the foundation and the sites but also to be of use to other educators and policymakers. This report also contributes to the growing research literature on the sorting of effective teachers by student characteristics, exemplified by the recent National Center for Education Evaluation and Regional Assistance report (Isenberg et al., 2013). More information about the Intensive Partnerships initiative is available in accompanying reports (Stecher, Garet, Hamilton, et al., in production; Gutierrez, Weinberger, and Engberg, in production), which provide detail on the reforms' implementation and the reforms' effect on students' level of achievement, respectively.

Contents

Figures

Tables

Summary

As part of its effective-teaching initiative, Intensive Partnerships for Effective Teaching, the Bill & Melinda Gates Foundation has partnered with three urban school districts across the United States and a group of four charter management organizations to undertake a set of strategic human-capital reforms. The reforms are intended to improve teachers' overall effectiveness and to ensure that students from historically disadvantaged backgrounds—specifically, low-income minority (LIM) students—have access to highly effective teachers. Lack of access to effective teaching has been identified as a possible contributor to the well-documented achievement gap between LIM students and their more-advantaged peers.

This report attends to the distribution of effective teachers within and across schools in the sites, collectively known as the Intensive Partnership sites. We examine the trends in the distribution of effective teachers between LIM students and other students. We also examine whether any of a variety of mechanisms can explain changes in LIM students' access to effective teaching. These mechanisms include increasing the percentage of LIM students whom effective teachers teach, increasing the effectiveness of teachers with large percentages of LIM students, and replacing less effective teachers of LIM students with more-effective teachers.

The first step in our analysis is to estimate each teacher's contributions to his or her students' achievement—that is, that teacher's value added. We use a common value-added model with teacher-linked data

on the mathematics and reading performance of students in grades 3 through 8 in the 2006–2007 through 2013–2014 school years.

We then examine the sorting of teachers by their value added between LIM students and other students for each site. First, we estimate the annual within-school association between a teacher's value-added estimate and the proportion of that teacher's students who are LIM students. We also examine sorting of teacher effectiveness across schools within each Intensive Partnership site, estimating the annual association between average teacher value added in each school and the proportion of students at the school who are LIM students. Finally, we estimate the sorting of effective teachers within schools by comparing a teacher's value added and the proportion of a teacher's students who are LIM students, holding the school constant. Our focus is on how these overall, within-school, and between-school associations have changed over time in each site—particularly on whether LIM students' access to effective teachers in their schools and school systems has improved in the four academic years since the Intensive Partnerships initiative commenced. We also repeat the analysis separately for elementary and middle school grades.

We next examine the sorting of teacher effectiveness measures that the sites provided. Each site uses a different achievement growth measure and calculates a different composite measure of effectiveness. The composites include achievement growth as one component worth 30 to 40 percent of the overall effectiveness score, with the rest made up of classroom-observation measures and other inputs.

We conclude our report with an accounting of the mechanisms that sites might use to change LIM students' access to effective teaching. We first focus on teachers at the top 20 percent and bottom 20 percent of the value-added performance distribution in each year and examine whether they teach more or fewer LIM students or quit teaching the subject in the following year. Although this analysis captures the changes in access based on assignments as related to the past year's performance, it does not reflect performance of new teachers brought in to teach tested subjects or changes in returning teachers' performance from one year to the next. We conduct a second analysis, which decomposes the change in access over time into portions attributable to teacher replacement, teacher improvement, or reassignment of LIM students to better-performing teachers.

Findings

We find that preintervention sorting patterns generally favored LIM students in most sites, subjects, and years and that those patterns have persisted in some cases during the intervention years. In other words, teachers with more LIM students have higher value added, on average, than teachers with fewer LIM students do. This was largely true before the intervention and has remained fairly consistent since the intervention began.

Despite this generalization, the study sites varied notably in both their longitudinal sorting trends and their recent amounts of sorting:[1]

- In Hillsborough County Public Schools in Florida, the largest site in the study, sorting in mathematics and reading has been fairly stable over time and close to neutral with respect to LIM students. The most recent year suggests slightly regressive sorting: In 2014, a teacher with 10 percentage points more LIM students than other teachers have was estimated to produce 1.1 percent of a standard deviation less achievement than those other teachers in mathematics and 0.3 percent of a standard deviation less in reading—small but statistically significant differences.[2]
- In Memphis City Schools in Tennessee (which merged with Shelby County Schools shortly before the last year of the analysis period for this report), LIM students' access to effective teachers was generally trending downward until the most recent year, in which there was a jump toward more-favorable sorting. In 2014, a

[1] Our findings focus on the three districts. The four charter management organizations have fewer schools and students, a situation that leads to less precise findings. Indeed, of these, only Aspire Public Schools has an adequate sample size to yield any findings at all. Because of the tentative nature of the Aspire findings, we relegate their presentation and discussion to the appendix.

[2] Our references to statistical significance use a *p*-value of 0.05.

teacher with 10 percentage points more LIM students than other teachers have was estimated to produce 4.2 percent of a standard deviation more achievement than those other teachers in mathematics and 0.9 percent of a standard deviation more in reading, both statistically significant.

- In Pittsburgh Public Schools in Pennsylvania, sorting has not been consistently positive or negative, although the estimates of the degree of sorting fluctuate substantially from year to year. In 2014, a teacher with 10 percentage points more LIM students than other teachers have was estimated to produce 0.8 percent of a standard deviation more achievement than those other teachers in mathematics and 0.9 percent of a standard deviation more in reading, with neither statistically significant.

Additional analysis of sorting of LIM students with respect to the site-generated composite effectiveness measure, which includes classroom-observation scores, shows significantly more-negative sorting. In other words, observation scores are consistently more negative for teachers with more LIM students, which is consistent with other evidence that observation scores do not account for classroom context (Whitehurst, Chingos, and Lindquist, 2014).

During both the preintervention and postintervention periods, sorting between schools has generally been more favorable to LIM students than sorting within schools has been. The sites are more successful at placing the most-effective teachers in schools with a high percentage of LIM students than they are in placing the most-effective teachers within each school in high-LIM classrooms.[3]

We examine whether the more-negative within-school sorting appears to be due to greater within-school sorting in middle schools. The division of classes into advanced and regular tracks in many middle schools provides a clear opportunity for such sorting. However, we do not find any consistent pattern of greater negative within-school

[3] For this report, we define *success* as a higher coefficient of the sorting parameters; with regard to sorting, we define *success* as teachers with higher value added having more LIM students.

sorting for middle school grades than for elementary school grades. As we look across subjects and districts, we find that, in some years, within-school sorting is more beneficial for LIM students in middle school grades and, in some years, it is greater in elementary school grades. Therefore, the tendency to have more-effective teachers with fewer LIM students within schools cannot be solely attributed to middle school academic tracks. Although there is less variation in LIM percentage among teachers within schools than between schools, and traditions, such as rewarding effective teachers by having them teach more-advanced classes, which might work against providing LIM students access to the most-effective teachers within schools, we recommend that sites determine whether there are feasible opportunities to improve within-school sorting.

Our analysis of the sites' own achievement growth and composite measures of teacher effectiveness showed sorting patterns that tend to be less favorable to LIM students than those using our achievement growth measure. The difference varies by site but is most pronounced for the composite measures. The changes over time and the between-school/within-school split for the sites' measures, however, are similar to the result we find using our value added–based measures. The most-likely explanations for the discrepancies in the amount of sorting are that our value-added models adjust for both student-level and classroom-level background characteristics in an attempt to isolate teacher effects, that our models are based on single-year estimates of teachers' value added in order to capture true year-to-year changes in effectiveness, and that the classroom-observation scores included in the composite measures do not adequately account for differences in student background. Variation among the sites' measures underscores the importance of our using a single value-added model to make apples-to-apples comparisons when evaluating this multisite intervention.

In our accounting of what sites do to change the sorting from one year to the next, we find little evidence of increased systematic use of particular mechanisms to improve access. Neither the first analysis based on prior-year effectiveness nor the second analysis that accounts for changes of effectiveness from year to year and changes in personnel indicates a consistent strategy across the sites. In some sites, more-effective teachers are replacing less effective teachers of LIM students; in other sites, less effective teachers of LIM students are increasing their effectiveness.

In sum, we find that LIM students enjoyed slightly better-than-average access to high-performing teachers before the Intensive Partnerships intervention commenced. This favorable pattern has largely persisted and increased slightly overall, although all sites are not taking the same steps toward improving access. Finally, the fact that LIM students appear to benefit more from between-school sorting of teacher effectiveness than from within-school sorting suggests that the Intensive Partnership sites should pay particular attention to within-school dynamics that might restrict LIM students' access to the top teachers in their schools.

Acknowledgments

We are grateful to the large number of district administrative staff and community stakeholders who reviewed our tables and figures and helped us correctly analyze the Intensive Partnerships initiative's effect on access to effective teaching in each site. These people include Anna Brown and Ted Dwyer in Hillsborough County Public Schools; Bradley Leon and Jessica Lotz in Shelby County Schools; and Tara Tucci and Ashley Varrato in Pittsburgh Public Schools.

We appreciate the data that the sites provided that enabled us to perform these analyses. In particular, Jeffery Shive, Leah Halteman, and Tangie Ray of Shelby County Schools; Lorraine Marnet of Pittsburgh Public Schools; Marianne Aman and David Russell of Hillsborough County Public Schools; and David Roth and Veronica Chew of Aspire Public Schools helped assemble and interpret these data.

We also appreciate the Bill & Melinda Gates Foundation staff's willingness to engage with us; we are especially grateful to Sara Allan, Nate Brown, Sarah Buhayar, Steve Cantrell, Josh Edelman, Eli Pristoop, Irvin Scott, Greg Sommers, and Ky Vu for their advice and responsiveness throughout the project. We thank Charles A. Goldman and Catherine H. Augustine of RAND and Cory Koedel of the University of Missouri for reviewing the document and providing constructive feedback. Finally, we acknowledge other members of the RAND team: Stephanie Lonsinger, Jan M. Hanley, Stephanie Williamson, Jennifer L. Steele (now of the American University), Brian M. Stecher, and Elizabeth D. Steiner. We also acknowledge members of the American Institutes for Research team, especially Michael S. Garet.

Abbreviations

AIR	American Institutes for Research
CMO	charter management organization
HCPS	Hillsborough County Public Schools
LIM	low-income minority
MCS	Memphis City Schools
NAEP	National Assessment of Educational Progress
NCEE	National Center for Education Evaluation and Regional Assistance
PD	professional development
PPS	Pittsburgh Public Schools
SCS	Shelby County Schools
SGP	student growth percentile
TCRP	the College-Ready Promise
VAM	value-added measure
WLS	weighted least squares

CHAPTER ONE

Introduction

One key objective of the Intensive Partnerships for Effective Teaching is to promote increased disadvantaged students' access to the most-effective teachers in their respective sites. With funding from the Bill & Melinda Gates Foundation, the Intensive Partnerships for Effective Teaching intervention commenced in the 2010–2011 school year. It is being carried out in three large, urban school districts—Hillsborough County Public Schools (HCPS) in Florida, Shelby County Schools (SCS) in Tennessee (previously, Memphis City Schools [MCS]), and Pittsburgh Public Schools (PPS) in Pennsylvania—as well as in four California charter management organizations (CMOs) that make up a coalition called the College-Ready Promise (TCRP).[1] The CMOs that make up TCRP are Aspire Public Schools, which operates in California and Tennessee; Alliance College-Ready Public Schools in the Los Angeles area; Green Dot Public Schools, which are mostly in and around Los Angeles; and Partnerships to Uplift Communities Schools, which also operates mostly in the Los Angeles area. The Intensive Partnerships initiative includes an integrated set of programs, or *levers*, designed to improve teacher human capital in each site and to improve disadvantaged students' access to the best teachers working within a site. These levers include efforts to improve how teachers are recruited, hired, and professionally developed; how they are evaluated, rewarded,

[1] On July 1, 2013, MCS merged with SCS. We refer to the Memphis-based site primarily as MCS rather than as SCS because the data used in this report either predate the merger of the two districts or, for 2013–2014, use only the schools from MCS because we need a prior year with which to estimate the value-added measure (VAM).

and dismissed; and how they are assigned to—or incentivized to work with—high-need students.

In this report, we describe trends over time in the association between a teacher's value-added estimate and the proportion of students in that teacher's classes and school who are both low-income and minority (LIM). In this context, we define *low-income* as eligible for free or reduced-price meals, and *minority* refers to students classified in districts' administrative data sets as black, Hispanic, or Native American, or combinations of any of these with other ethnicities or races. Insofar as data are available, we examine this association during the three school years before the Intensive Partnerships initiative commenced (2007–2008, 2008–2009, and 2009–2010), as well as during the first four years of Intensive Partnerships implementation (2010–2011, 2011–2012, 2012–2013, and 2013–2014).[2]

In a scenario in which teachers are randomly assigned, we would expect to see no relationship between students' LIM status and the quality of their teachers. However, if there were positive sorting, in which more-effective teachers were more likely to be assigned to higher-LIM concentration classrooms and less effective teachers were more likely to be assigned to lower-LIM concentration classrooms, we would expect to see a positive association between teacher value added and student LIM status. Key questions include whether sorting patterns change over time and whether within-school sorting patterns appear to reinforce, be similar to, or offset between-school sorting.

We also investigate sorting of teacher effectiveness using the various effectiveness measures that each site developed as part of the Intensive Partnerships initiative. Finally, we examine the extent to which sites are changing access over time, using such mechanisms as changing classroom composition differently based on teacher effectiveness or improving teachers' effectiveness differentially based on their classroom composition.

---

[2] In an earlier working paper, we described how teacher effectiveness estimates are sorted in the three years before and three years after the Intensive Partnerships intervention commenced (Steele et al., 2014).

This report focuses on the LIM and non-LIM students' relative access to effective teachers, which has been the focus of recent research and policies (Isenberg et al., 2013). Of course, changing the relative access to effective teachers is only one of the ways in which sites can improve the quality of the education that LIM students receive. Even if the distribution of teaching effectiveness by LIM status does not change, LIM students will have access to more-effective teaching if the quality of all teachers improves. Furthermore, achievement can improve for some or all students for reasons other than an increase in teaching effectiveness. The RAND/American Institutes for Research evaluation has produced an accompanying report that examines the reforms' impact on achievement for all students and select subgroups (Gutierrez, Weinberger, and Engberg, in production). The remainder of this report proceeds as follows: Chapter Two briefly describes the context of Intensive Partnerships implementation to date. Chapter Three describes our empirical approach to estimating value added and examining its sorting by students' LIM status. Chapter Four presents longitudinal trends in our estimates of teachers' value added for the three districts, as well as in the districts' estimates of teachers' value added and their overall teacher effectiveness composites.[3] In Chapter Five, we present our analyses of ways in which sites can change LIM students' access to effective teaching. Chapter Six concludes. An appendix provides further tables and figures, details regarding the methodology for decomposing the sorting mechanisms, and the results for the Aspire CMO.

---

[3] Our findings focus on the three districts. The four CMOs have fewer schools and students, which leads to less precise findings. Indeed, of these, only Aspire has an adequate sample size to yield any findings at all. Because of the tentative nature of the Aspire findings, we relegate their presentation and discussion to an appendix.

CHAPTER TWO

Implementation Background

Since the Intensive Partnerships initiative began in the fall of 2010, each Intensive Partnership site has taken a distinctive approach to its slate of human-capital reforms. Stecher, Garet, Hamilton, et al., in production, provides detailed summaries of the Intensive Partnership sites' implementation of various aspects of the human-capital reforms. Because this report focuses on changes over time in LIM students' access to effective teachers, we focus in particular on two aspects of the implementation context that are relevant to the ensuing discussion. The first is the extent to which each site has emphasized levers aimed at shifting LIM students' access to effective teachers—what we call *distribution levers*. The second is the weighted combination of measures that each site is using in its composite measures of teacher effectiveness.

We briefly highlight recent data about distribution levers in each site because these are the principal mechanisms through which the sites intend to change LIM students' access to higher–value added teachers. Table 2.1 summarizes key factors that might especially influence changes in LIM students' access to quality teaching in each site. We note that, as of 2011–2012, HCPS was increasingly using teacher-evaluation data in placing teachers within schools and in making decisions about which teachers would be allowed to change schools. This could affect the extent to which high-performing teachers are assigned to LIM students within and between schools. In MCS, we also notice levers that emphasize distributing higher-quality teachers to high-need students—namely, by restricting who can transfer to high-need schools and basing transfer hiring decisions on effectiveness data rather than

Table 2.1
Context Shaping Disadvantaged Students' Access to Effective Teachers

Site	Teacher Placement Policy
Aspire	Principals make hiring decisions. There are no centralized placement procedures.
HCPS	Teacher-evaluation data are used to assess within-school placements at high-need schools and by principals in hiring transfer teachers.
PPS	Career-ladder opportunities are available to entice high-performing teachers to historically low-achieving schools. Hiring of new teachers into the district has been very limited because of enrollment declines and budget restrictions.
SCS and MCS	Transferring teachers must have high scores to teach in the highest-need schools. Seniority is not used in placement decisions.

SOURCE: Stecher and Garet, 2014.

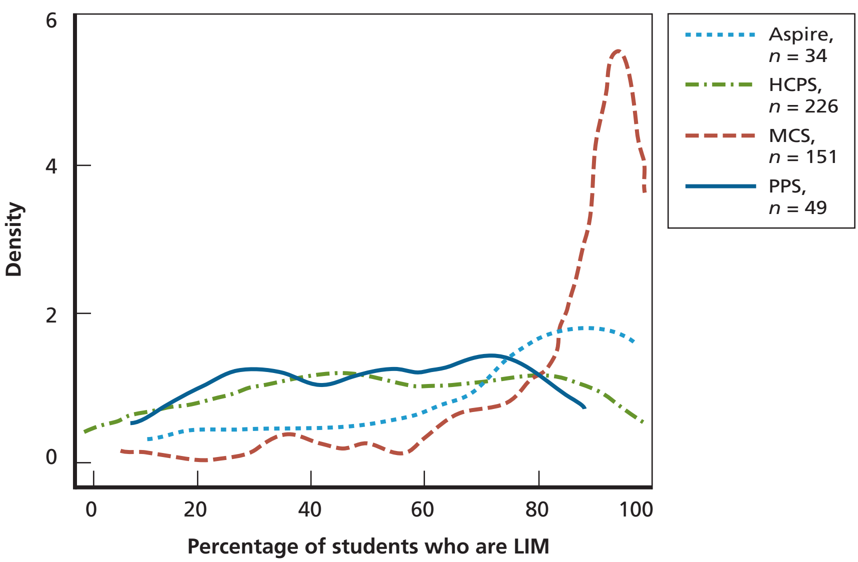
on seniority. We see less emphasis on sorting teachers progressively in PPS's stated policies, but that is not to say that other efforts to improve teacher evaluation and capacity might not also improve the relative quality of teachers who have many LIM students. And in Aspire, as is true for the other TCRP CMOs, principals have hiring authority, so centralized teacher placement policies do not play a role in where teachers are asked to work.

It is important to acknowledge that these distribution levers are not the only Intensive Partnerships–related mechanisms that might affect LIM students' instructional experiences. For instance, if teachers or schools that serve many LIM students get priority access to quality professional development (PD), or if high-LIM schools are staffed by newer principals who are more aggressive in recruiting and evaluating teachers, those factors can also shift the distribution of effective teaching within a site. Furthermore, many reform policies intended to affect teachers of all students, such as bonuses to retain the most-effective teachers or training for all principals to hire more-effective teachers, can differentially affect LIM students. Still, the purpose of Table 2.1 is to highlight policies that can especially influence how teachers are assigned to students and schools.

It is also important to bear in mind that the foundation chose the Intensive Partnership sites as grantees in part because they had already shown innovation in their teacher human-capital reforms. In other words, they might have been attending to concerns about LIM students' access to effective teaching even before the Intensive Partnerships initiative took effect. For example, HCPS staff have noted in interviews that high-need Renaissance School Services schools had priority hiring rights in place before Intensive Partnerships. Moreover, given that they are CMOs, TCRP sites had school-based hiring policies in place before Intensive Partnerships began.

A few other factors are worth bearing in mind when viewing the findings below. First, as shown in Figure 2.1, the distribution of schools' percentage of LIM students is not the same for all sites. In

Figure 2.1
Distribution of School-Level Low-Income Minority Proportions, by Site, in School Year 2013–2014



NOTE: The vertical axis can be interpreted by taking an interval on the horizontal axis and multiplying its width by the average height of a curve over that interval. For example, the Aspire curve has an average height of about 2 over the interval between 80 and 100 percent on the horizontal axis, indicating that about 40 percent (= 100 × 2 × 20%) of Aspire schools' student populations are between 80 and 100 percent LIM students.

RAND *RR1295/4-2.1*

MCS, the majority of schools are 80 percent LIM or more; in the other sites, the proportion of LIM students in each school is more evenly distributed, with a few low-LIM schools, a few high-LIM schools, and numerous mid-LIM schools (20 to 80 percent LIM). The consequence is that teachers' and schools' proportions of LIM students are less variable in a site that, like MCS, has mostly high-LIM schools than in a site that, like PPS or HCPS, has a fairly even distribution of high-, low-, and mid-LIM schools.

Second, the sites differ in the variation in percentage of a teacher's students who are LIM students and the extent to which this variation reflects differences in the percentage of a school's students who are LIM students or within-school differences among teachers' students. Table 2.2 shows that the standard deviation among schools ranges from 19 to 30, and the standard deviation among teachers within schools ranges from 5 to 10. These values limit sites' ability to reassign teachers, especially within schools, to improve access.

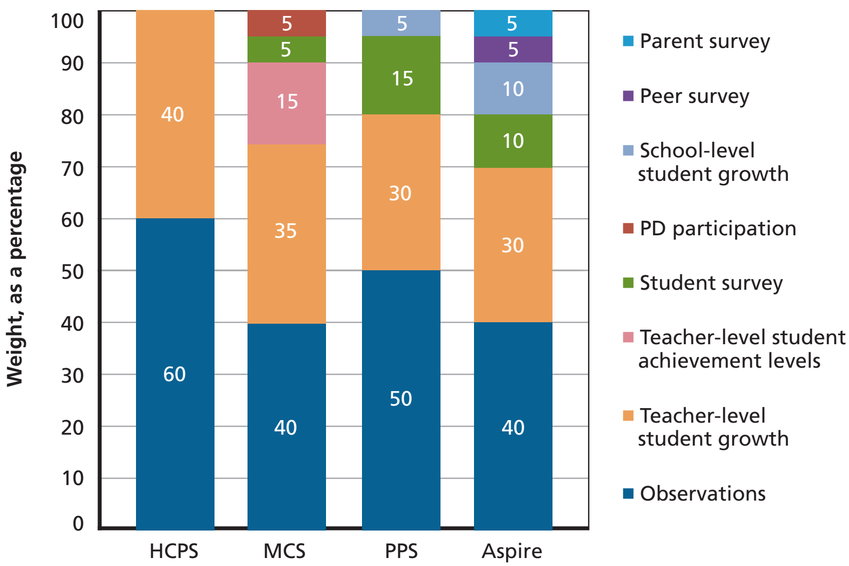
Finally, the concentration of LIM students for MCS changed on July 1, 2013, when the district merged with SCS. As noted above, because we have only one year of data for the schools previously not in MCS, we cannot estimate VAM for these teachers (needing baseline scores for the students from the previous year). For this reason, and to maintain the continuity of the schools represented rather than focus on the large change in the schools, the analysis focuses on MCS in terms of its district boundaries at the start of the study.

Table 2.2
Percentage of Students Who Are Low-Income Minority Students, by Teacher

		Standard Deviation		
Site	Average	Overall	Between Schools	Within a School
Aspire	67.1	30.1	29.61	5.36
HCPS	40.1	25.3	23.35	9.80
MCS	84.7	20.1	18.98	6.73
PPS	50.0	25.6	24.49	7.58

The other implementation feature of particular importance to this report is the way in which each site combines various effectiveness measures to create a composite measure. We report below on the sorting with respect to three measures of teacher effectiveness: measures based on our own value-added estimates for teachers in mathematics and reading over a six-year time span; teacher-level estimates of value added or student achievement growth that the sites reported; and teacher effectiveness *composites* that the sites reported. To interpret differences between sorting parameters for the sites' value-added estimates and their composites, it is useful to know how the composites are constructed. Figure 2.2 shows the weights that the sites reported *for teachers of students in tested subjects and grades*. (Note that, because we report only on composites for teachers of mathematics and reading in grades 4 through 8, we display weights pertaining only to teachers for whom we can estimate student growth.)

Figure 2.2
Component Weights in Effectiveness Composites for Teachers of Tested Grades and Subjects



RAND *RR1295/4-2.2*

We see that the predominant measure in all of the composites is teacher-observation scores, ranging from 40 percent in MCS and Aspire to 60 percent in HCPS. These are followed in importance by teacher-level estimates of student growth (value added or student growth percentiles [SGPs]), which are weighted between 30 and 40 percent. Other data sources play small roles in some sites, including student achievement levels and participation in PD in MCS, student surveys in PPS and Aspire, and parent and peer surveys in Aspire. Value added as the sites estimated (which we discuss further in Chapter Five) is but one component, and not the predominant one, in the composite estimates, so we would anticipate that sorting estimates based on the composites might differ from sorting estimates based on value added alone.

CHAPTER THREE

Empirical Approach

We begin our cross-site analysis of teacher sorting by estimating teacher value added separately for each subject, mathematics and reading, using the same model in all sites. A teacher of both subjects will have two separate value-added estimates in each year in which that teacher appears in the data—one for mathematics and one for reading. A teacher of only one of the two subjects will have one estimate per year. This is distinct from the approach that the sites took, which instead gives each teacher one value-added estimate per year, taking into account the teacher's students' performance in both mathematics and reading.

We estimate the value-added estimates and sorting parameters in separate stages, employing a generalized least-squares hierarchical fixed-effects approach that Borjas and Sueyoshi, 1994, describes and Aaronson, Barrow, and Sander, 2007, applies to teacher value added. In the first-stage model,

$$A_{icjt} = \alpha_0 + \alpha_1 A_{it-1} + X_{it}\alpha_X + Z_{ct}\alpha_Z + \mu_{jt} + \varepsilon_{icjt}. \qquad (3.1)$$

A_{icjt} is student achievement for student i assigned to teacher j in year t and classroom section c. It is first scaled to a state-level z-score using the state/year/grade standard deviations and means and, from there, scaled to the national level using the National Assessment of

Educational Progress (NAEP).[1] Achievement is a function of lagged achievement (A_{it-1}), which is an estimate of the combination of innate ability and prior learning; observed student-level covariates (X_{it}), including gender, race and ethnicity, socioeconomic status, being over age for one's grade, gifted status, and status as an English language learner; classroom-level covariates (Z_{ct}), which include lagged student-level test scores and the other covariates aggregated to the classroom level, as well as class size. μ_{jt} is the teacher value added in year t, and ε_{icjt} is the random noise (unexplained variation in student test scores). Student-level and classroom-level covariates (i.e., measures except for lagged test scores) are centered at their site-specific (i.e., district- or CMO-specific) means.

The inclusion of classroom-level covariates allows us to separate teachers' contributions to student learning and the aggregate effects of the classroom composition. We identify the effects of the classroom-level covariates within teacher, taking advantage of the fact that many teachers across grades and sites teach more than one class section in a given content area each year. Tables A.1 and A.2 in the appendix provide details.

Equation 3.1 could alternatively be estimated in two stages, one that regresses student test scores on student covariates and classroom dummy variables and a second that regresses the estimated classroom fixed effects on classroom-level covariates and teacher dummy variables. However, in sensitivity analyses, we found very little difference in value-added model estimates or in associations between value-added estimates and students' LIM status when we collapsed the first two stages, as shown in Equation 3.1. This suggests that the classroom-level covariates capture the important sources of variation for teachers' classroom-level deviations from their overall value added.

---

1 We want value-added estimates to be in units that allow us to compare across sites and over time, which scaling to the external NAEP allows us to do. A sample of students in grades 4 and 8 takes the exam every two years in each state. We use the means and standard deviations for each state and nationally to rescale scores to the national norm. We use linear egression to interpolate means and standard deviations for grades in between grades 4 and 8 and for each untested year.

Our models account for the fact that test scores (and thus lagged test scores) are measured with error. Like Briggs and Domingue, 2011, in accounting for this measurement error, we use two-stage least squares and instrument lagged test scores using the lagged test scores from the other subject (e.g., lagged mathematics score is instrumented by lagged reading score).[2]

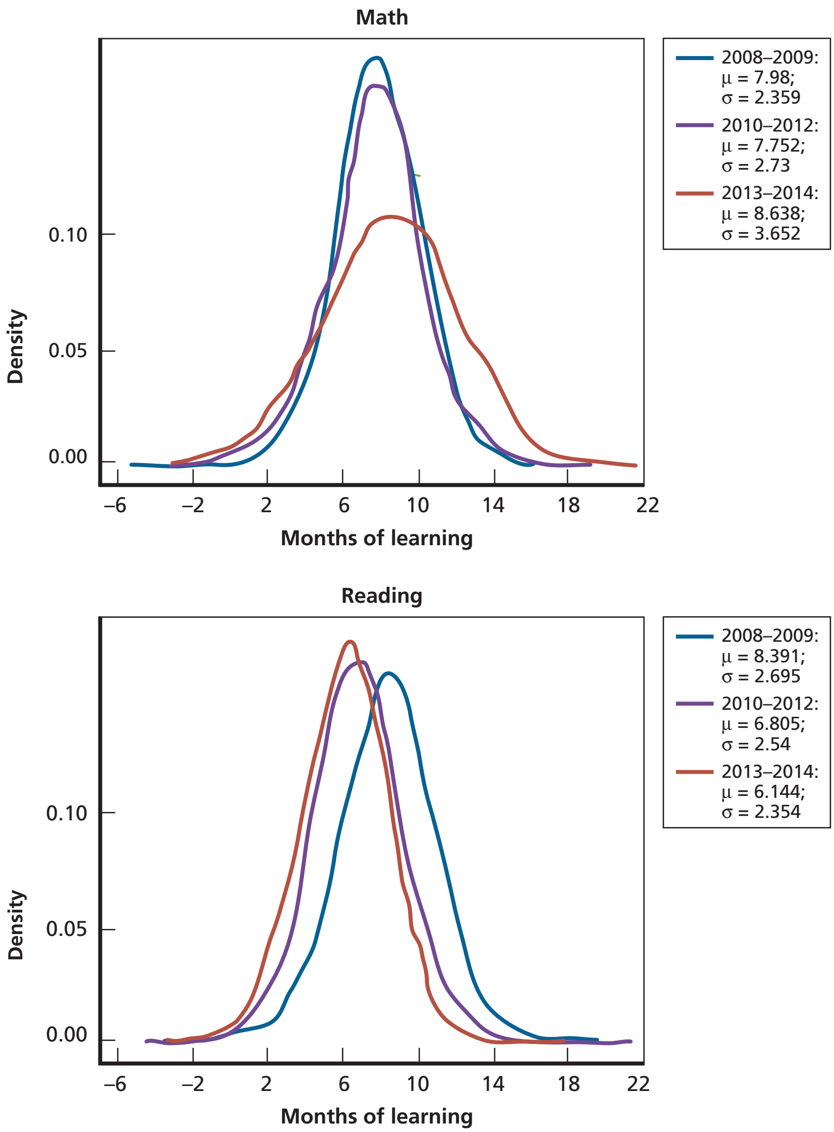
To estimate Equation 3.1, we use weighted least squares (WLS), with weights given by the proportion of the year that students were taught by a given teacher in the tested subject. In other words, following the Hock and Isenberg, 2012, full-roster method, a student's test score might appear as multiple observations in the data, with one record for each course in which the student was taught the tested subject. Weights reflect the proportion of the school year that the student spent in a particular course and are constrained not to exceed 1. This constraint means that we anticipate 0 marginal return to supplemental doses of mathematics or reading instruction beyond the first course. Weights are calculated as p/k, where p is the proportion of the school year the student spent in a given school (using modal enrollment days at that school as a denominator) and k is the number of unique mathematics or reading class sections in that school to which the student is linked in a given year.[3]

---

[2] We experimented with various instruments, such as double lags in the same subject and in the other subject, and found little difference in the value-added estimates or in the teacher sorting coefficients. Likewise, we tested the inclusion of lagged other test score as a control variable instead of as an instrument and found similar results. We settled on the specification used here to be consistent with the literature that accounts for measurement error and to retain as many observations as possible (hence, not using double lags). We note that Lockwood and McCaffrey, 2014, investigates a variety of methods for correcting for measurement error and uses simulation methods to show that a well-identified instrumental variable method performs just as well as a more burdensome method based on conditional standard errors of measurement. It does not, however, investigate whether using an additional score as an instrument, like we do, is preferable or whether using it as an additional covariate is.

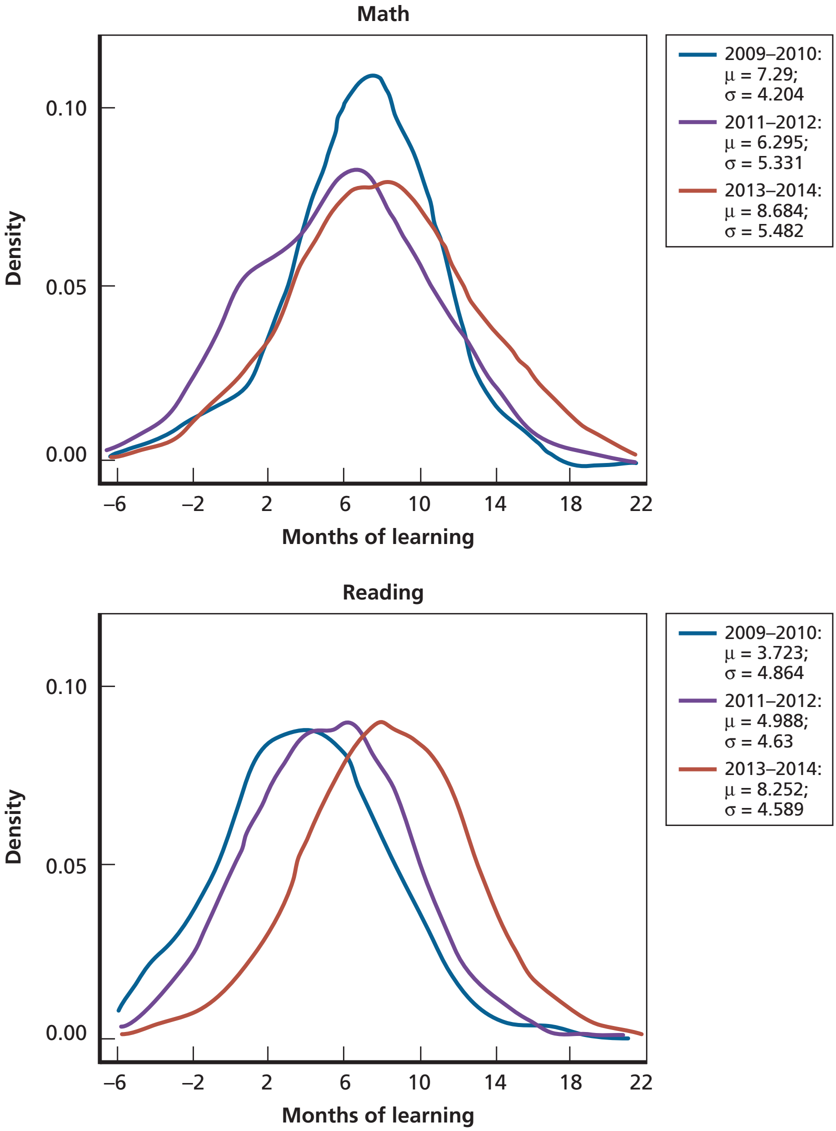
[3] In sensitivity tests, we gave each record a weight of p rather than p/k, thereby allowing the sum of a student's weights to exceed 1. Our results were not sensitive to the use of this alternative weighting approach.

Figure 3.1
Hillsborough County Public Schools Distribution of Value-Added Measure, by Period: Math and Reading Learning Density



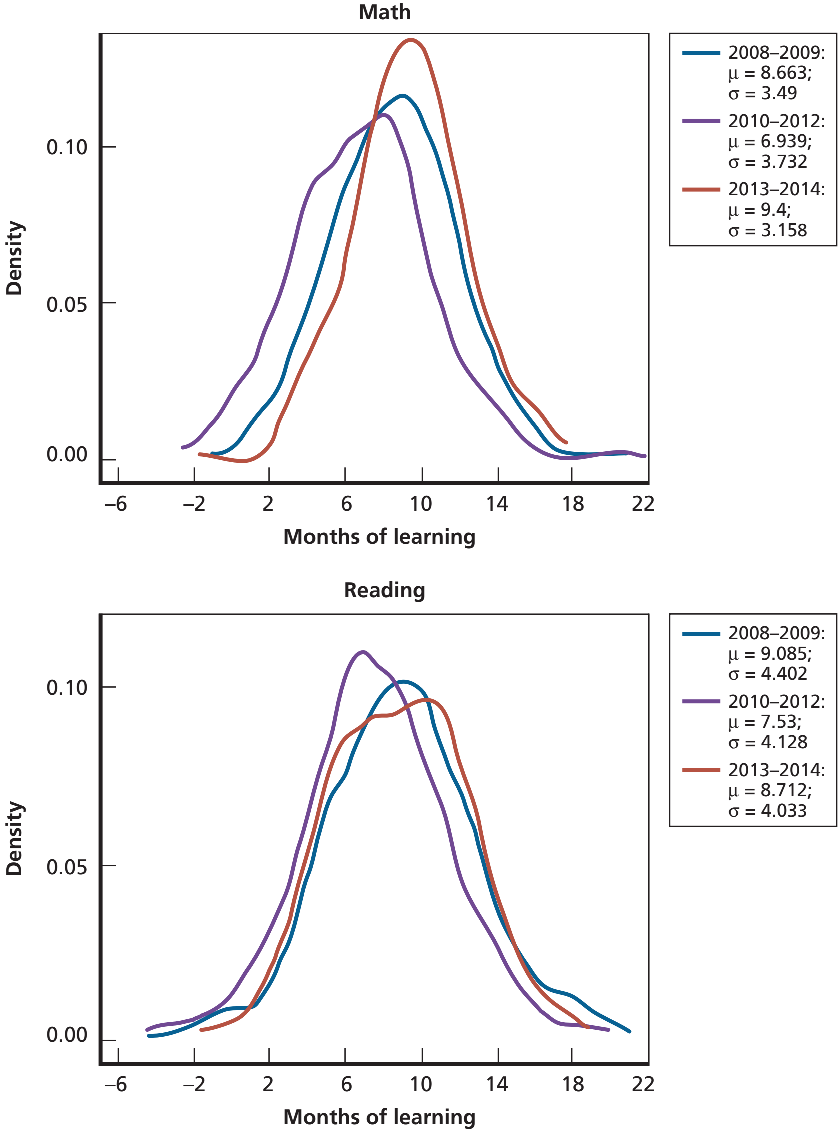
RAND *RR1295/4-3.1*

Figure 3.2
Memphis City Schools Distribution of Value-Added Measure, by Period: Math and Reading Learning Density



RAND *RR1295/4-3.2*

Figure 3.3
Pittsburgh Public Schools Distribution of Value-Added Measure, by Period: Math and Reading Learning Density



RAND *RR1295/4-3.3*

Figures 3.1 through 3.3 show the distributions of estimated value added for each subject in each of the three districts. Each figure shows the distribution during three time periods: prereform, early reform, and recent. In addition to graphing the distributions, we provide the means and standard deviations for each district in each period. For ease of interpretation, we have translated value added from the units that we use elsewhere in this report (i.e., student-level nationwide standard deviation of achievement, as implied by our transformation of achievement to a *z*-score on the NAEP scale) to months of learning (Bloom et al., 2008).[4]

As these figures suggest, there are substantial differences among sites and over time both in the average value added and in the shape.[5] The distributions for HCPS are much less spread out than those for MCS.[6] In some sites and subjects, we see large changes in the average value added over time; in other cases, the average value added is fairly stable.

Our aim in this report is not to analyze differences in these distributions between sites and years but to analyze how the teachers within these distributions are sorted between LIM and non-LIM students. As all of these distributions show, there are considerable differences in value added among teachers, suggesting that students in the same site might be taught by teachers of very different performance levels. Therefore, after estimating teacher effects, we estimate three relationships between teachers' proportions of students who are LIM students in year *t* and teacher effectiveness in year *t*. The first is the overall relationship, representing the extent to which each teacher's fitted value added,

---

4 We set national average achievement equal to nine months.

5 Gutierrez, Weinberger, and Engberg, in production, provides estimates of the Intensive Partnerships reforms' effects on student achievement up through 2014. The present measures of value added, although scaled to the NAEP scale so as to be comparable across sites and over time, are not true impact measures. The estimates in Gutierrez, Weinberger, and Engberg, in production, measure effects by comparing average achievement in the sites and predicted achievement based on that of other schools in the same states and on prereform relative site performance.

6 We graph shrunken estimates, so differences between sites in the estimate precision is not driving the difference in distribution spread. See McCaffrey et al., 2004.

$\hat{\mu}_{jt}$, is related to the proportion of that teacher's students who are LIM, regardless of the school in which the teacher works. This relationship is captured with a second-stage regression, in which the parameter of interest, β_1, represents the difference in $\hat{\mu}_{jt}$ associated with a unit difference in the share of all of teacher *j*'s students in year *t* who are LIM students:[7]

$$\hat{\mu}_{jt} = \beta_0 + \beta_1 LIM_{jt} + \upsilon_{jt}. \tag{3.2}$$

To account for the randomness associated with the estimation of $\hat{\mu}_{jt}$, we estimate these two stages using generalized least squares; that is, we weight the second-stage regression by the Cholesky decomposition of the inverse of the variance–covariance matrix associated with the estimation of μ_{jt}. Note that this also shrinks noisy estimates of the value added and so is comparable to empirical Bayes shrinkage, a common postestimation strategy for teacher value-added models (McCaffrey et al., 2004).

We are also interested in decomposing the relationship of LIM and teacher value added into the within-school and between-school components to see whether sorting is particularly strong in one or both areas. To do so, instead of estimating Equation 3.2 as the second stage, we estimate Equation 3.3. θ_{st} is a fixed effect controlling for the school (*s*) in which the teacher works during year *t*. Controlling for the schools changes the interpretation of the coefficient on teacher's average percentage of students who are LIM students. β'_1 can now be interpreted as the sorting between LIM and value added that occurs within schools because we are holding constant the schools to which the teachers are assigned:

$$\hat{\mu}_{jt} = \beta'_0 + \beta'_1 LIM_{jt} + \theta_{st} + \upsilon_{jt}. \tag{3.3}$$

---

[7] Note that, because LIM_{jt} is coded from 0 to 1, a unit difference is actually a 100-percentage-point difference.

We also estimate a third regression (again, using generalized least squares), replacing the LIM share of the *teacher's* students (LIM_{jt}) with the LIM share of the *school's* students:

$$\hat{\mu}_{jt} = \gamma_0 + \gamma_1 LIM_{st} + \eta_{st}. \tag{3.4}$$

γ_1 represents the relationship of teaching effectiveness among schools based on the percentage of their students who are LIM students. It reflects the sorting of teaching effectiveness between schools. Overall sorting (β_1) is a weighted average of within-school sorting (β'_1) and between-school sorting, with the weights reflecting the ratio of the variances of between-teacher percentage of students who are LIM students and between-school percentage of students who are LIM students (Raudenbush and Bryk, 2002, p. 137).

It is important to note that γ_1 also reflects anything about the school that makes all teachers in the school more or less productive, such as leadership effectiveness, special programs, or resources. Although it has been shown that teachers are the most-important school-based factors in students' achievement growth, the presence of these other factors could bias our estimates of between-school sorting.

In value-added estimation, an important consideration is whether to estimate teachers' value added using just their students in the current year or whether to include the performance of their prior-year students as well. Several studies have demonstrated marked improvement in the reliability of value-added estimates when they incorporate the performance of the students the teachers taught not only in the current year but also in one or more previous years (Goldhaber and Hansen, 2010; Schochet and Chiang, 2010). Presumably for this reason, the PPS and MCS Intensive Partnership sites calculate teachers' value-added estimates based on value-added estimates that average teachers' performance across multiple years. Given that the sites' estimates carry high stakes for teachers, this approach seems appropriate for strengthening the reliability of the estimates.

However, the downside of averaging value added across years is that it likely understates *true* year-to-year variation in teacher performance. In the case of the Intensive Partnerships evaluation, in which we are interested in gauging the Intensive Partnerships initiative's effect on not only teachers' assignments to their schools but also changes in individual teachers' effectiveness relative to other teachers in the same Intensive Partnership sites, we estimate value added based on the performance of a teacher's students in the current year. Our own investigations have revealed this to be the correct choice in our setting across various loss functions. Although this might result in some instability because of the sample of students a teacher is assigned in a given year, it also allows our estimates to capture true year-to-year changes in teachers' relative effectiveness.

A related consideration we face is whether to examine sorting of teacher value added by student LIM composition in terms of teachers' estimated effectiveness in the current or the prior year. In this report, we focus on the sorting of LIM students in terms of teachers' current-year effectiveness estimates. This approach allows us to examine the extent to which LIM students have access to high-quality teaching in each year of the study compared with their non-LIM peers in the same sites. Changes in sorting patterns from year to year can arise for a variety of reasons. These include not only changes in how existing teachers are assigned to classrooms or schools by administrators (or how they are encouraged to take different assignments) but also such factors as how new teachers are assigned and how teachers of LIM students are professionally developed or rewarded for improving their instructional practice. In other words, our approach takes into account all of the factors that can shift the relative quality of teaching that LIM students receive from year to year.

An alternative approach would be to estimate the relationship between teachers' prior-year value-added estimates and the LIM statuses of their current students. This approach would capture the extent to which the sites were assigning teachers to classrooms or schools based on what was previously known about their performance. However, because schools typically do not have value-added estimates available for the prior year until shortly before or even after the start of a new school year, we would actually need to use teachers' value added from two years prior to the current year to report on the extent to which sites were deliberately assigning teachers to schools or classrooms based on prior value-added estimates. Moreover, because the Intensive Partnerships intervention largely precipitated the systematic use of teachers' value added in decisionmaking, schools would not have been able to base assignments on prior-year value added until the 2012–2013 school year in HCPS and the 2013–2014 school year in the other sites, so we would not have much data to detect these effects in our current data. For all of these reasons, we focus instead on sorting of current-year value added by teachers' current-year student LIM compositions. From a student's perspective, this is the most important definition because it captures the relative quality of instruction that LIM students are receiving in a given year.

In general, we pool all teachers in grades 4 through 8 when we examine sorting. However, the greater variety of course offerings in middle school than elementary school suggests that there might be more sorting of students within schools during these years. The greater departmentalization suggests that within-school sorting might differ more between subjects in middle school grades than in elementary school grades. Therefore, we also conduct the same sorting analysis after dividing teachers into elementary grades (grades 4 and 5) and middle school grades (6 through 8).

This report documents not only the sorting of our teacher value-added estimates by students' LIM status but also the sorting of teacher effectiveness estimates that the sites provided for up to three post–Intensive Partnerships–inception years: 2011 and 2012 in HCPS and 2012, 2013, and 2014 in MCS and PPS. We report on the distribution of two kinds of effectiveness estimates that the sites provided: value-added or SGP estimates at the teacher level and teacher effectiveness composites. For teachers of students in tested subjects and grades, the composites include the site-generated value-added estimates or SGPs as one component of a weighted composite. In the composite definitions in Figure 2.2 in Chapter Two, we denote this component as "Teacher-level student growth." For ease of reference, we generally call these the *site-generated value-added estimates* in this report, although, when referring specifically to teacher-level student growth estimates from Aspire, we use the term *SGPs* because this is the method that the CMOs use for calculating achievement growth (Betebenner, 2009).[8] To have comparable sorting parameters, we rescale the site estimates to the same mean and standard deviation as our VAM, by site, subject, and year. Also to aid comparison, we analyze sorting patterns for only the subsample of teachers for whom we can calculate VAM.

This allows us to use the same scale to discuss the sorting of each set of estimates by students' LIM statuses. This means that the units in which we discuss the sorting of the site-generated estimates correspond to standard deviations of our evaluation-model teacher effectiveness estimates in a particular site and year. The question of interest is how well the various measures of teacher effectiveness lend themselves to generalization about sorting levels and trends in each site. Because the value-added model we apply across sites (our evaluation model) yields a consistent apples-to-apples comparison between sites, and because we can estimate it for a six-year period in most sites, we focus on that as our main window into the sorting of teacher effectiveness by students' LIM statuses. However, we also examine how well our estimates line up with the sorting patterns we find using the site-generated effectiveness estimates, and we comment on possible reasons for any observed discrepancies.

Before we present our sorting estimates, we would like to emphasize that there are many ways to estimate value-added models that produce similar but not identical results. For example, some studies omit classroom covariates from the equation. Some use the average of residuals for a teacher's students to estimate the teacher's value added rather than including an indicator (or dosage) variable for each teacher. The conversion to NAEP units also has limitations because it is based on interpolation of average scores for untested years and grades and on the unrealistic assumption that student scores on the NAEP and the state assessment are very highly correlated. Throughout this report, we discuss the limitations of both our value-added estimation and of our estimation of the sorting parameters. However, we have chosen our value-added model and sorting model over the many alternatives because we

---

[8] We include more discussion of the implications of the differences between value added and student growth percentiles when we discuss the Aspire findings in the appendix.

judge our models to be most suited to our particular purpose—namely, to trace differences over time and among districts in overall, between-school, and within-school sorting of teacher effectiveness by the percentage of students in their classes who are LIM students. Other modeling choices with different limitations are likely preferred for other purposes, such as evaluating individual teachers or estimating long-term effects of teacher value added on student outcomes. We now turn to presenting our findings.

CHAPTER FOUR

Longitudinal Trends, by Site

In this chapter, we describe trends over time in the association between a teacher's value-added estimate and the proportion of students in that teacher's classroom and school who are LIM students. We present these trends first for each of the three urban school districts in the study: HCPS, MCS, and PPS. Because of their more tentative nature, we present trends for Aspire Public Schools, which is the largest TCRP CMO and the only one for which we could estimate value-added models, in an appendix.[1] The figures here accompany tables showing the number of students, teachers, and schools included in the analysis in each subject and year. In the section for each of the districts, we also comment on any differences in sorting patterns between elementary and middle school grades, referring to figures in the appendix. The smallness of the sample prevents us from examining sorting by grade level for Aspire.

In addition to showing the association of our estimates of teacher value added with proportion LIM, we present comparable information regarding the association of the sites' own teacher value-added estimates and composite measures of teacher effectiveness. The sites have been constructing these composites as part of the Intensive Partnerships initiative since school year 2010–2011 in HCPS and since 2011–

1 For one of the other TCRP sites, Green Dot Public Schools, we cannot estimate value-added models because all of the schools are high schools, and we can estimate value added only for teachers of grades 4 through 8. For the other two TCRP sites, Partnerships to Uplift Communities Schools and Alliance College-Ready Public Schools, the number of teachers and schools in the sites is too small to provide reliable estimates of the distribution of teacher value added within or between schools.

2012 in the other sites. At present, we can track the distribution of the site-generated teacher composites and value-added estimates for only three years per site because only these have been made available to us. Still, for ease of comparison, we present these sorting estimates on the same scales as those used to present longitudinal sorting of our own value-added estimates. It is important to note that the sites do not estimate teachers' value added separately for mathematics and reading. For example, a teacher who has students who take both the math and reading standardized exam would still only have one site VAM and composite score. Each teacher's value-added estimate reflects the relative performance of that teacher's students for all the subjects in which that teacher teaches them. To compare with our VAM estimates that are by subject, we include a teacher in the math calculations, for example, if the teacher has any students who take the math exam. For comparability with our estimates, our figures present site-generated effectiveness estimates only for grades 4 through 8 because these are grades in which it is also possible to estimate value added in mathematics and reading for all sites.[2] Given that sorting patterns are not always identical in the two sets of graphs, we consider possible reasons for discrepancies in the cross-site discussion later in this section.

The figures that follow show the longitudinal sorting trends in three ways. The light blue dashed line represents the estimated relationship of teacher value-added estimates and students' LIM statuses *within* schools. As such, it corresponds to parameter β'_1 in Equation 3.3 in Chapter Three. The green dash-and-dotted line represents the estimated relationship of teacher value added in each school and the proportion of students in a school who are classified as LIM, or the *between*-school relationship. It corresponds to parameter γ_1 in Equation 3.4 in Chapter Three. Finally, the solid dark-blue line represents overall sorting both between and within schools. It corresponds to parameter β_1 in Equation 3.2 in Chapter Three. In this sense, the

[2] In the tables accompanying the figures showing the site-generated estimates, we provide a row called "Site-based VAM, all teachers." This pertains to all the teachers for whom the sites calculate value added, whether or not we have student data on them. In some cases, it might include students tested in subjects other than mathematics and reading or teachers of upper grades, in which students are not necessarily tested every year.

overall sorting parameter is the most complete measure of teacher sorting for a site in a given year, whereas the within- and between-school parameters can help districts diagnose where and how the sorting is occurring.

In the figures, an estimate of 0 means that there is no relationship in a given subject, site, and year between teachers' value-added estimates and the LIM statuses of their students. In other words, it means that a student's LIM status is, on average, unrelated to the effectiveness of the teacher to whom the student is assigned. To call attention to positive versus negative sorting of teacher value added, the 0 grid line on the y-axis is marked in red. A positive coefficient means that LIM status is associated with *higher* average teacher value-added estimates, whereas a negative coefficient means that LIM students are being taught by teachers with lower value-added estimates than their non-LIM peers are. Coefficients are denominated in nationwide student-level standard deviations of achievement in the testing year as adjusted for the covariates listed above in the description of the value-added model. The proportion LIM ranges from 0 to 1. For example, a coefficient of 0.20 indicates that a teacher with all LIM students (i.e., a proportion LIM equal to 1) produces learning that is equivalent to 0.2 standard deviations in test scores more than a teacher with no LIM students. Throughout the following discussion of the magnitudes of the coefficients, we use the example of teachers who differ in student proportion LIM by 10 percentage points (i.e., 0.10). For two such teachers, a coefficient of 0.20 indicates that the one with more LIM students produces learning that is 0.02 standard deviations, or 2 percent of a standard deviation, more than the other. The figures also indicate whether each sorting coefficient differs from 0 at the 5-percent level of statistical significance. Given that we are reporting on the value-added estimates for all available mathematics and reading teachers in grades 4 through 8 in a given site and year, one could argue that it is the degree of observed sorting in the sample that concerns us, rather than generalizability to a larger population. Still, if a parameter is very imprecisely estimated, it might reasonably be interpreted as noise rather than as a signal of meaningful differences between the teachers of higher-LIM and lower-LIM classes. For this reason, we denote parameter estimates that are statistically significant at the 0.05 level with solid markers and estimates that are statistically indistinguishable from 0 with hollow markers. In all graphs, we denote the inception of the Intensive Partnerships intervention with a vertical line between the 2009–2010 and 2010–2011 school years.[3]

Hillsborough County Public Schools in Florida

Sorting of Value Added in Mathematics and Reading over Time

HCPS is the largest site in the Intensive Partnerships study and thus the site in which the teacher-sorting parameters are most precisely estimated. Table 4.1 presents the three sorting estimates for each subject and year: the between-school coefficient, the within-school coefficient, and the overall coefficient. We show these estimates in the “Beta” columns. In the central columns, headed “Standard Error,” we present the standard errors associated with the sorting coefficients. The right two columns present the numbers of teachers and schools in each analysis.

To facilitate interpretation of the values in Table 4.1, we present longitudinal graphs of the mathematics and reading sorting coefficients in Figure 4.1. The left panel of Figure 4.1 represents the between, within, and overall sorting coefficients for mathematics shown in the top half of Table 4.1, and the right panel represents the sorting coefficients for reading shown in the bottom half of Table 4.1. In the figure, we use a solid marker to denote an estimate that is statistically distinguishable from 0 and a hollow marker to denote a non–statistically significant estimate.

Examining the sorting coefficients represented in Table 4.1 and Figure 4.1, we observe the patterns we describe in the rest of this section.

---

[3] In all cases, we adjust for multiple hypotheses using the Benjamini–Hochberg method (Benjamini and Hochberg, 1995), with each family being for a given estimator across years and subjects within a district. Of course, these methods for inference account only for classical sampling variation and its effect on parameter estimates. As in all estimation, differences between the model and reality will produce additional sources of error.

Table 4.1
Hillsborough County Public Schools Sorting Parameters Using Evaluation Value-Added Measure, by Subject and Year

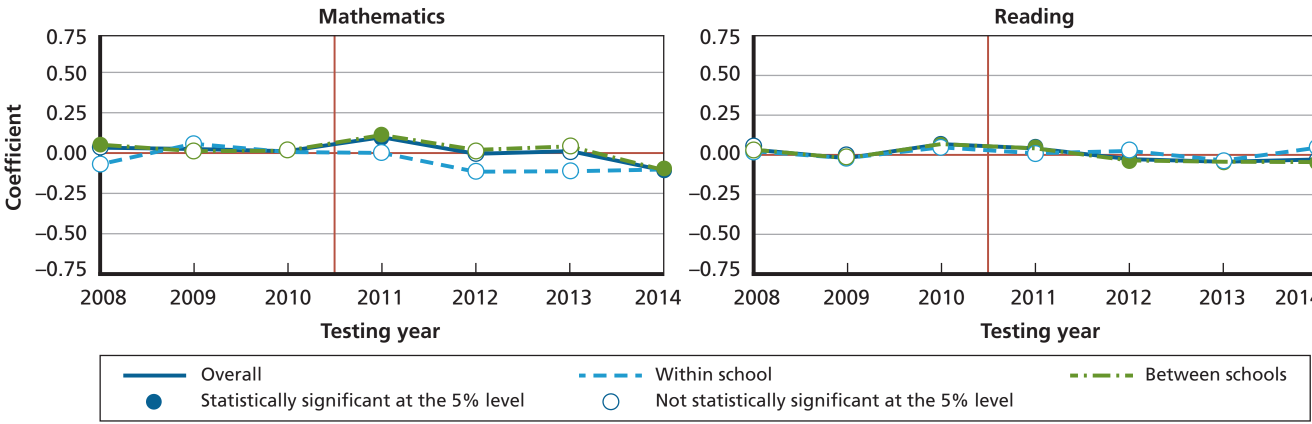
Year	Beta			Standard Error			Teachers	Schools
	Overall	Between	Within	Overall	Between	Within		
Mathematics								
2008	0.032	0.054**	–0.069	0.019	0.020	0.044	1,615	180
2009	0.022	0.011	0.055	0.017	0.017	0.035	1,786	185
2010	0.016	0.015	0.013	0.016	0.017	0.036	1,682	190
2011	0.100***	0.113***	0.003	0.020	0.021	0.050	1,630	188
2012	–0.003	0.017	–0.113	0.023	0.023	0.056	1,555	187
2013	0.016	0.042*	–0.114	0.020	0.021	0.047	1,463	188
2014	–0.111***	–0.104***	–0.102	0.024	0.024	0.054	1,554	189
Reading								
2008	0.032*	0.034*	0.020	0.016	0.016	0.036	1,854	180
2009	–0.024	–0.023	–0.023	0.014	0.015	0.030	2,188	185
2010	0.066***	0.066***	0.046	0.015	0.015	0.034	2,036	189
2011	0.036**	0.040***	0.007	0.014	0.014	0.033	2,077	188
2012	–0.030**	–0.038***	0.024	0.012	0.012	0.029	1,994	188

Table 4.1—Continued

Year	Beta			Standard Error			Teachers	Schools
	Overall	Between	Within	Overall	Between	Within		
2013	–0.046***	–0.046***	–0.037	0.013	0.013	0.031	1,833	188
2014	–0.030*	–0.047***	0.046	0.014	0.015	0.030	1,698	187

NOTE: The "Beta" columns show three sorting estimates for each subject and year: the overall coefficient, the between-school coefficient, and the within-school coefficient. The "Standard Error" columns show the standard errors associated with the sorting coefficients. The right two columns present the numbers of teachers and schools in each analysis. *** = $p < 0.01$. ** = $p < 0.05$. * = $p < 0.1$. We used the Benjamini–Hochberg method to adjust significance levels for multiple hypotheses.

Figure 4.1
Relationship Between Teacher Value-Added Measure and Percentage of Students Who Are Low-Income Minority Students in Hillsborough County Public Schools



NOTE: The red vertical line signifies inception of the Intensive Partnerships initiative. The red horizontal line shows a coefficient of 0, which means that a student's LIM status is unrelated to the effectiveness of the teacher to whom the student is assigned. A positive (i.e., above that line) coefficient means that LIM status is associated with *higher* average teacher value-added estimates; a negative (below the red horizontal) coefficient means that LIM students are being taught by teachers with lower value-added estimates than their non-LIM peers are. The light-blue dashed lines represent the estimated relationship of teacher value-added estimates and students' LIM statuses *within* schools and correspond to parameter β'_1 in Equation 3.3 in Chapter Three. The green dash-and-dotted lines represent the estimated relationship of teacher value added in each school and the proportion of students in a school who are classified as LIM, or the *between*-school relationship, and correspond to parameter γ_1 in Equation 3.4 in Chapter Three. The solid dark-blue lines represent overall sorting both between and within schools and correspond to parameter β'_1 in Equation 3.2 in Chapter Three.

RAND *RR1295/4-4.1*

Mathematics

- Between-school sorting peaked in 2011 and has been declining since and now disadvantages LIM students.
- At the same time, within-school sorting also has been trending in a way that disadvantages LIM students.
- The overall effect, shown with the solid blue line, is a sorting pattern that was mostly positive or neutral for LIM students but that, in the most recent year, displays slightly negative sorting. The estimated coefficient of –0.111 in 2014 indicates that the average value added for a teacher serving students with a 10-percentage-point higher LIM rate than that teacher's subject/grade peers had a value-added estimate that was 1.1 percent of a standard deviation lower than that of those peers, on average.
- Because we weight the between-school trends more heavily than within-school trends, the overall sorting pattern has remained neutral until the most recent year.

Reading

- Sorting between and within schools has hovered close to 0 before and after the intervention began, but with some perturbation in the year immediately before and the two years after intervention inception, and has generally trended slightly negative (or in favor of non-LIM students).
- Overall sorting favored LIM students in 2007–2008, 2009–2010, and 2010–2011 by as much as 0.066 standard deviations in 2009–2010. The 2014 estimated coefficient of –0.030 means that a teacher with 10 percentage points more LIM students than other teachers have had a value-added estimate that was 0.3 percent of a student standard deviation lower than those other teachers had, on average.
- Within- and between-school sorting generally has been much closer for reading than for mathematics.

Figure A.1 in the appendix presents the sorting coefficients separately by grades 4 through 5 and grades 6 through 8.

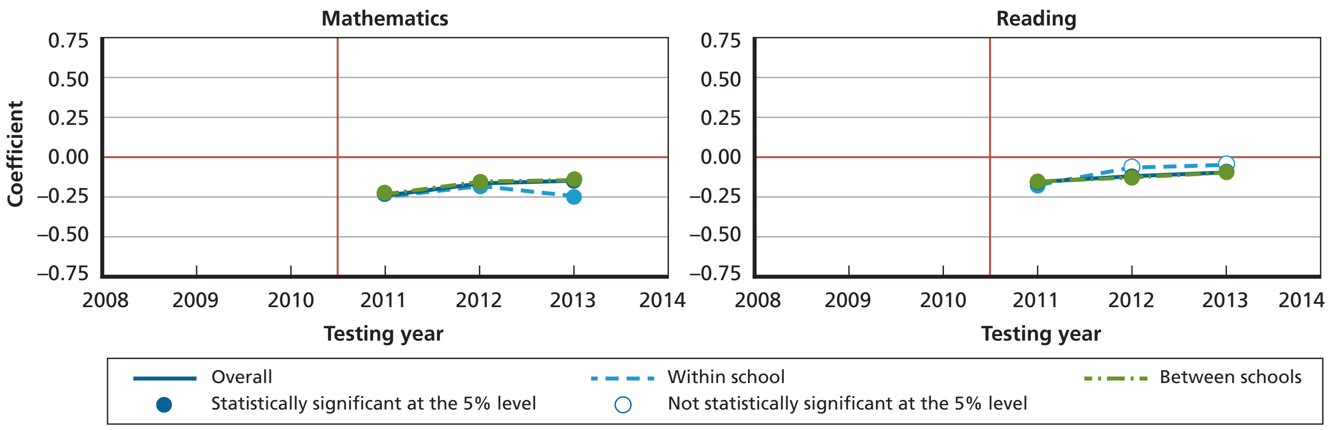
Sorting of Site-Reported Value-Added Measure Estimates and Composites

We turn now to the distribution of the site-generated teacher quality estimates that HCPS reported. For HCPS, we have site-generated value-added estimates and teacher effectiveness composites for the school years 2010–2011 (the first Intensive Partnerships implementation year) through 2012–2013, but estimates are not available for 2013–2014. In Figure 4.2, we examine the sorting of the sites' value-added estimates and, in Figure 4.3, the teacher effectiveness composites. To make the sorting parameters comparable, we scale the effectiveness measures to the same mean and variance as our VAM estimates. In Table 4.2, we present the sorting parameters that use all teachers for whom the site provides effectiveness measures and for a restricted subgroup of teachers for whom we can calculate VAM estimates. We use the restricted subgroup for Figures 4.2 and 4.3.

As was true for the estimates reported above based on our cross-site evaluation model, the sorting estimates we report here reflect the extent to which the site-generated value-added estimates and effectiveness composites differ by the proportion of students in a teacher's classes and school who are LIM students. Our main observations are as follows:

- Between-school sorting is more favorable to LIM students than within-school sorting for VAM and composite scores for mathematics teachers and for composite for reading teachers.
- However, the site-generated estimates for both value added and composites show overall sorting patterns that favor non-LIM students, as illustrated by the negative and statistically significant overall sorting estimates in all years and subjects. This is different from the mostly positive or neutral overall sorting patterns we reported above for mathematics and reading in those years, using the value-added model we are employing across sites. We discuss possible reasons for such differences below.
- There has been little change over time in sorting, with perhaps a slight move toward more-favorable sorting for LIM students than in previous years for site-provided VAM.

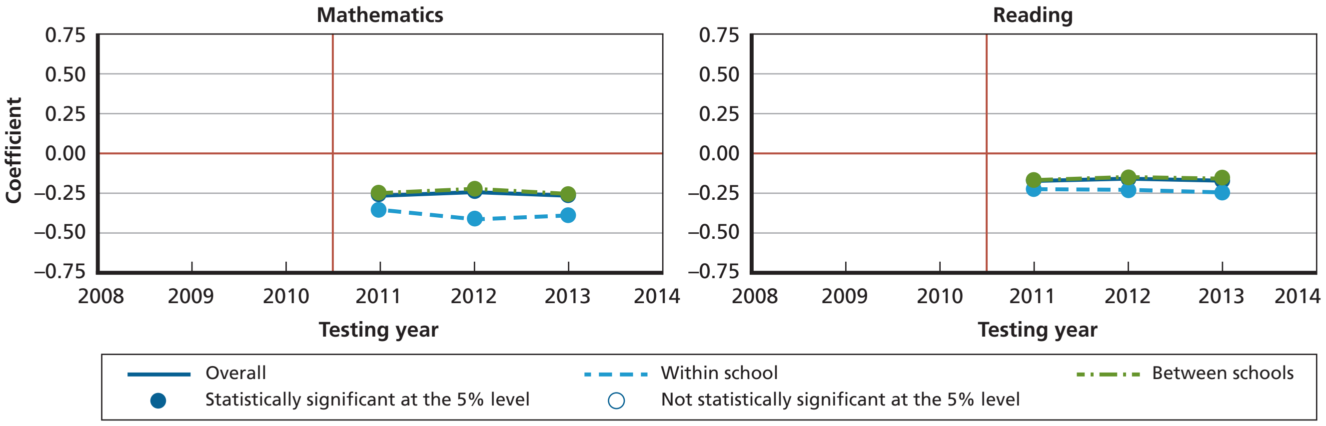
Figure 4.2
Sorting of Site-Provided Value-Added Measure Estimates, by Percentage of Students Who Are Low-Income Minority Students in Hillsborough County Public Schools



NOTE: The red vertical line signifies inception of the Intensive Partnerships initiative. The red horizontal line shows a coefficient of 0, which means that a student's LIM status is unrelated to the effectiveness of the teacher to whom the student is assigned. The negative (below the red horizontal) coefficients mean that LIM students are being taught by teachers with lower value-added estimates than their non-LIM peers are. The light-blue dashed lines represent the estimated relationship of teacher value-added estimates and students' LIM statuses *within* schools and correspond to parameter β_1' in Equation 3.3 in Chapter Three. The green dash-and-dotted lines represent the estimated relationship of teacher value added in each school and the proportion of students in a school who are classified as LIM, or the *between*-school relationship, and correspond to parameter γ_1 in Equation 3.4 in Chapter Three. The solid dark-blue lines represent overall sorting both between and within schools and correspond to parameter β_1' in Equation 3.2 in Chapter Three.

RAND *RR1295/4-4.2*

Figure 4.3
Sorting of Site-Provided Teacher Effectiveness, by Percentage of Students Who Are Low-Income Minority Students in Hillsborough County Public Schools



NOTE: The red vertical line signifies inception of the Intensive Partnerships initiative. The red horizontal line shows a coefficient of 0, which means that a student's LIM status is unrelated to the effectiveness of the teacher to whom the student is assigned. The negative (below the red horizontal) coefficients mean that LIM students are being taught by teachers with lower value-added estimates than their non-LIM peers are. The light-blue dashed lines represent the estimated relationship of teacher value-added estimates and students' LIM statuses *within* schools and correspond to parameter β'_1 in Equation 3.3 in Chapter Three. The green dash-and-dotted lines represent the estimated relationship of teacher value added in each school and the proportion of students in a school who are classified as LIM, or the *between*-school relationship, and correspond to parameter γ_1 in Equation 3.4 in Chapter Three. The solid dark-blue lines represent overall sorting both between and within schools and correspond to parameter β'_1 in Equation 3.2 in Chapter Three.

RAND *RR1295/4-4.3*

Table 4.2
Hillsborough County Public Schools Sorting Parameters, by Model and Year

	Beta			Standard Error				
Year	**Overall**	**Between**	**Within**	**Overall**	**Between**	**Within**	**Teachers**	**Schools**
Math, site-based VAM, restricted subsample								
2011	–0.237***	–0.233***	–0.246***	0.018	0.020	0.050	1,623	187
2012	–0.157***	–0.153***	–0.182***	0.021	0.022	0.059	1,553	187
2013	–0.154***	–0.140***	–0.243***	0.020	0.021	0.055	1,458	187
Reading, site-based VAM, restricted subsample								
2011	–0.158***	–0.155***	–0.175***	0.012	0.013	0.034	2,069	188
2012	–0.119***	–0.126***	–0.067*	0.012	0.012	0.033	1,987	187
2013	–0.096***	–0.103***	–0.050	0.013	0.014	0.037	1,819	187
Math, site-based VAM, all available teachers								
2011	–0.221***	–0.210***	–0.257***	0.013	0.018	0.047	2,883	187
2012	–0.181***	–0.141***	–0.183***	0.015	0.020	0.056	2,785	187
2013	–0.146***	–0.130***	–0.238***	0.014	0.020	0.052	2,643	187

Table 4.2—Continued

Year	Beta			Standard Error			Teachers	Schools
	Overall	Between	Within	Overall	Between	Within		
Reading, site-based VAM, all available teachers								
2011	−0.159***	−0.129***	−0.161***	0.009	0.011	0.030	2,882	188
2012	−0.115***	−0.114***	−0.057*	0.009	0.011	0.031	2,785	187
2013	−0.107***	−0.095***	−0.037	0.011	0.013	0.036	2,643	187
Math, site-based composite, restricted subsample								
2011	−0.262***	−0.245***	−0.353***	0.018	0.020	0.049	1,623	187
2012	−0.244***	−0.219***	−0.410***	0.020	0.021	0.057	1,553	187
2013	−0.269***	−0.249***	−0.386***	0.019	0.021	0.052	1,460	188
Reading, site-based composite, restricted subsample								
2011	−0.173***	−0.166***	−0.223***	0.012	0.013	0.034	2,069	188
2012	−0.158***	−0.148***	−0.226***	0.011	0.012	0.032	1,988	187
2013	−0.170***	−0.157***	−0.246***	0.013	0.014	0.036	1,824	188

Table 4.2—Continued

Year	Beta			Standard Error			Teachers	Schools
	Overall	Between	Within	Overall	Between	Within		
Math, site-based composite, all available teachers								
2011	–0.188***	–0.179***	–0.287***	0.010	0.014	0.037	2,883	187
2012	–0.196***	–0.170***	–0.318***	0.012	0.017	0.045	2,788	187
2013	–0.208***	–0.198***	–0.339***	0.011	0.017	0.043	2,650	188
Reading, site-based composite, all available teachers								
2011	–0.135***	–0.118***	–0.178***	0.007	0.009	0.026	2,881	188
2012	–0.125***	–0.110***	–0.170***	0.007	0.009	0.025	2,789	187
2013	–0.151***	–0.118***	–0.185***	0.008	0.010	0.028	2,649	188

NOTE: The “Beta” columns show three sorting estimates for each subject and year: the overall coefficient, the between-school coefficient, and the within-school coefficient. The “Standard Error” columns show the standard errors associated with the sorting coefficients. The right two columns present the numbers of teachers and schools in each analysis. *** = $p < 0.01$. ** = $p < 0.05$. * = $p < 0.1$. We used the Benjamini–Hochberg method to adjust significance levels for multiple hypotheses. We restricted the sample to teachers for whom we can calculate VAM using the common formula in Equation 3.1 in Chapter Three.

- There appears to be more-systematic negative sorting of effectiveness in terms of the composites than in terms of the site-generated value-added estimates alone. Within-school sorting of the composites is less favorable to LIM students than within-school sorting of the value-added estimates. And in 2013, overall sorting of the composites was less favorable to LIM students than overall sorting of the value-added estimates alone, though both were unfavorable to LIM students. Because the HCPS composites are 40 percent value added and 60 percent teacher observation scores (see Figure 2.2 in Chapter Two), this suggests that teachers with a higher proportion of students who are LIM students receive lower observation ratings, on average, than those with lower proportions of students who are LIM students and that this is especially true for teachers in the same school. It also suggests that observation ratings in schools with higher proportions of students who are LIM students are higher, on average, than observation ratings in schools with lower proportions of students who are LIM students.
- As a result, the sites' composite estimates show greater disadvantage to LIM students than the value-added estimates alone.
- The sorting parameters are not substantially different depending on whether we use all teachers or only the subsample for whom we have estimated VAMs, particularly for the site-based VAM. For the site-based composite scores, using the same subset of teachers as our VAM sample does lead to slightly worse estimated access for LIM students to effective teachers.

Memphis City Schools in Tennessee

Sorting of Value Added in Mathematics and Reading over Time

Table 4.3 presents sorting coefficient estimates, standard errors, and sample sizes for both subjects in all years. Figure 4.4 then presents the sorting between schools, within schools, and overall. As before, a solid marker indicates that a sorting-pattern estimate is statistically sig-

Table 4.3
Memphis City Schools Sorting Parameters Using Evaluation Value-Added Measure, by Subject and Year

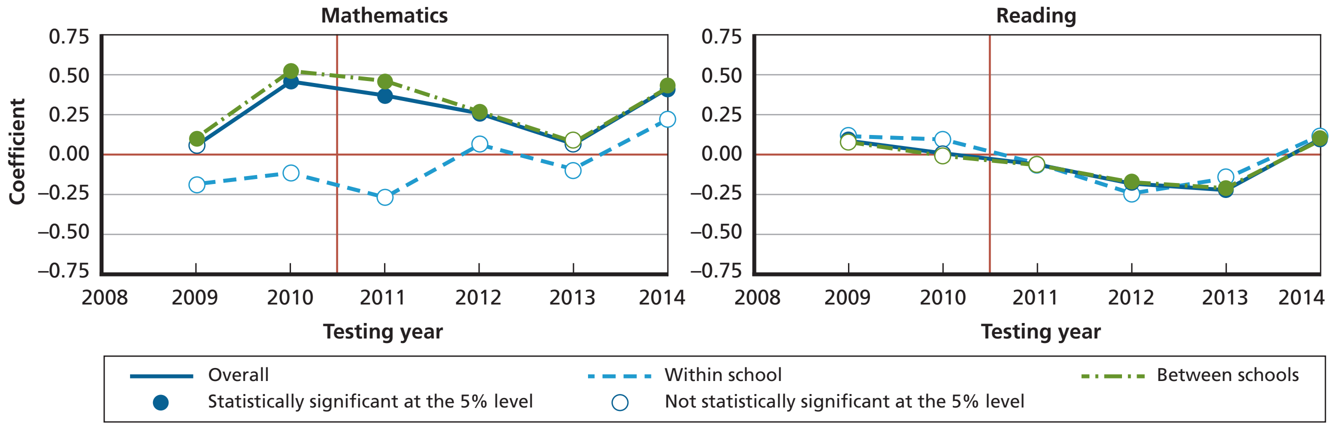
	Beta			Standard Error				
Year	Overall	Between	Within	Overall	Between	Within	Teachers	Schools
Mathematics								
2009	0.052	0.092**	–0.185	0.036	0.037	0.090	958	152
2010	0.457***	0.522***	–0.116	0.046	0.048	0.142	889	139
2011	0.364***	0.456***	–0.270	0.060	0.063	0.160	762	137
2012	0.257***	0.266***	0.065	0.054	0.054	0.159	765	140
2013	0.061	0.079	–0.096	0.065	0.067	0.193	780	135
2014	0.421***	0.416***	0.226	0.059	0.059	0.139	751	137

Table 4.3—Continued

	Beta			Standard Error				
Year	Overall	Between	Within	Overall	Between	Within	Teachers	Schools
Reading								
2009	0.088**	0.075*	0.112	0.039	0.041	0.081	1,121	155
2010	0.007	–0.007	0.098	0.036	0.037	0.097	1,100	140
2011	–0.067*	–0.065*	–0.057	0.036	0.037	0.091	946	136
2012	–0.186***	–0.172***	–0.243	0.033	0.034	0.097	979	140
2013	–0.212***	–0.207***	–0.151	0.034	0.034	0.098	920	137
2014	0.091***	0.084**	0.124	0.033	0.034	0.091	910	137

NOTE: The "Beta" columns show three sorting estimates for each subject and year: the overall coefficient, the between-school coefficient, and the within-school coefficient. The "Standard Error" columns show the standard errors associated with the sorting coefficients. The right two columns present the numbers of teachers and schools in each analysis. *** = $p < 0.01$. ** = $p < 0.05$. * = $p < 0.1$. We used the Benjamini–Hochberg method to adjust significance levels for multiple hypotheses.

Figure 4.4
Relationship Between Teacher Effects and the Percentage of Students Who Are Low-Income Minority Students in Memphis City Schools



NOTE: The red vertical line signifies inception of the Intensive Partnerships initiative. The red horizontal line shows a coefficient of 0, which means that a student's LIM status is unrelated to the effectiveness of the teacher to whom the student is assigned. A positive (i.e., above that line) coefficient means that LIM status is associated with *higher* average teacher value-added estimates; a negative (below the red horizontal) coefficient means that LIM students are being taught by teachers with lower value-added estimates than their non-LIM peers are. The light-blue dashed lines represent the estimated relationship of teacher value-added estimates and students' LIM statuses *within* schools and correspond to parameter $\beta_1^{'}$ in Equation 3.3 in Chapter Three. The green dash-and-dotted lines represent the estimated relationship of teacher value added in each school and the proportion of students in a school who are classified as LIM, or the *between*-school relationship, and correspond to parameter γ_1 in Equation 3.4 in Chapter Three. The solid dark-blue lines represent overall sorting both between and within schools and correspond to parameter $\beta_1^{'}$ in Equation 3.2 in Chapter Three.

RAND *RR1295/4-4.4*

nificant at the 5-percent level, meaning that the relationship between teacher value added and student LIM appears to be nonrandom.

For MCS, we observe the sorting patterns in each subject, mathematics and reading, shown in Figure 4.4.

Mathematics

- Preintervention sorting was strongly favorable for LIM students for mathematics.
- In the first year of the intervention, between-school teacher sorting in mathematics was strongly progressive and statistically significant, with an estimated coefficient of 0.46. Since the first year, between-school sorting has become less favorable to LIM students until 2014, when it increased again to 0.42. In substantive terms, this means that a school with a 10-percentage-point higher LIM enrollment rate than a comparison school was staffed by teachers whose value-added estimates were, on average, 4.2 percent of a standard deviation higher than those of the teachers in the same subject and grade in the comparison school.[4]
- Within-school sorting favored non-LIM students until 2012 but has never been statistically significant.
- The overall sorting initially strongly favored LIM students in math; over time, that has trended toward favoring non-LIM students until 2013–2014, when there was a sharp increase in LIM students' access to effective teachers driven both by within-school and between-school sorting, leading to an overall positive sorting parameter that is higher than in any previous intervention year, at 0.42. That implies that a teacher with 10 percentage points more LIM students will produce achievement that is 4.2 percent of a standard deviation higher than a teacher with otherwise-comparable students would.

[4] Note that, even though MCS schools are mostly more than 80 percent LIM students, there are schools with lower percentages of students who are LIM students, so this comparison is meaningful in the MCS context.

Reading

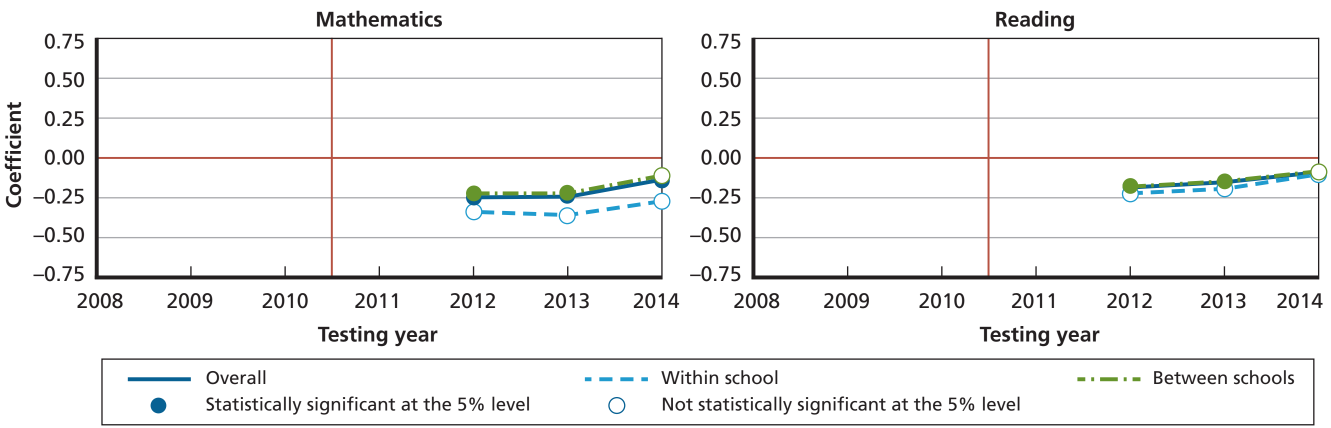
- Preintervention sorting was slightly favorable for LIM students in reading.
- Through 2012–2013, all three sorting trends in reading appear to be less and less advantageous to LIM students every year, with one exception for within-school sorting in 2011–2012. In other words, the downward slope of all three trends—between schools, within school, and overall—runs counter to the aims of the Intensive Partnerships intervention.
- There was a sharp increase in LIM students' access to effective teachers in 2013–2014, leading to an overall positive sorting parameter that is higher than in any previous year, at 0.091. That implies that a teacher with 10 percentage points more LIM students will produce achievement that is 0.9 percent of a standard deviation higher than a teacher with otherwise-comparable students would.

In mathematics, within-school sorting was more negative in middle school grades in the final two years of our analysis than it was for elementary school grades (see Figure A.2 in the appendix). This is a reverse of the relationship between the two grade levels in the early intervention years. For reading, within-school sorting has mostly been more favorable for middle school grades. These patterns do not suggest that academic tracking in later grades is a primary driver for negative within-school sorting.

Sorting of Site-Reported Value-Added Measure Estimates and Composites

For MCS, we have site-generated value-added estimates and composites for the school years 2011–2012 through 2013–2014 but not for the first Intensive Partnerships implementation year, 2010–2011. In Figures 4.5 and 4.6, we present sorting parameters for the sites' value-added estimates and for the sites' teacher effectiveness composites, respectively. Table 4.4 shows the coefficients, standard errors, and sample sizes corresponding to these figures.

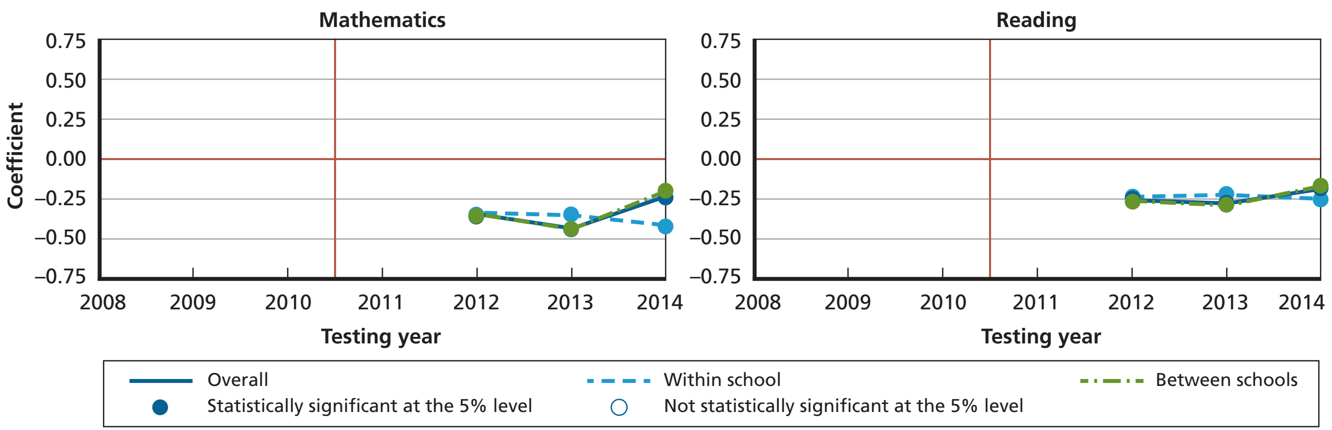
Figure 4.5
Sorting of Site-Provided Value-Added Measure Estimates, by Percentage of Students Who Are Low-Income Minority Students in Memphis City Schools



NOTE: The red vertical line signifies inception of the Intensive Partnerships initiative. The red horizontal line shows a coefficient of 0, which means that a student's LIM status is unrelated to the effectiveness of the teacher to whom the student is assigned. The negative (below the red horizontal) coefficients mean that LIM students are being taught by teachers with lower value-added estimates than their non-LIM peers are. The light-blue dashed lines represent the estimated relationship of teacher value-added estimates and students' LIM statuses *within* schools and correspond to parameter β_1' in Equation 3.3 in Chapter Three. The green dash-and-dotted lines represent the estimated relationship of teacher value added in each school and the proportion of students in a school who are classified as LIM, or the *between*-school relationship, and correspond to parameter γ_1 in Equation 3.4 in Chapter Three. The solid dark-blue lines represent overall sorting both between and within schools and correspond to parameter β_1' in Equation 3.2 in Chapter Three.

RAND *RR1295/4-4.5*

Figure 4.6
Sorting of Site-Provided Composite Estimates, by Percentage of Students Who Are Low-Income Minority Students in Memphis City Schools



NOTE: The red vertical line signifies inception of the Intensive Partnerships initiative. The red horizontal line shows a coefficient of 0, which means that a student's LIM status is unrelated to the effectiveness of the teacher to whom the student is assigned. The negative (below the red horizontal) coefficients mean that LIM students are being taught by teachers with lower value-added estimates than their non-LIM peers are. The light-blue dashed lines represent the estimated relationship of teacher value-added estimates and students' LIM statuses *within* schools and correspond to parameter β'_1 in Equation 3.3 in Chapter Three. The green dash-and-dotted lines represent the estimated relationship of teacher value added in each school and the proportion of students in a school who are classified as LIM, or the *between*-school relationship, and correspond to parameter γ_1 in Equation 3.4 in Chapter Three. The solid dark-blue lines represent overall sorting both between and within schools and correspond to parameter β'_1 in Equation 3.2 in Chapter Three.

RAND *RR1295/4-4.6*

Table 4.4
Memphis City Schools Sorting Parameters, by Model and Year

Year	Beta			Standard Error			Teachers	Schools
	Overall	Between	Within	Overall	Between	Within		
Math, site-based VAM, restricted subsample								
2012	–0.245***	–0.227***	–0.334*	0.052	0.055	0.143	739	140
2013	–0.240***	–0.221***	–0.356*	0.063	0.067	0.177	693	133
2014	–0.135**	–0.112*	–0.268	0.061	0.065	0.165	596	134
Reading, site-based VAM, restricted subsample								
2012	–0.185***	–0.177***	–0.221*	0.033	0.035	0.094	951	140
2013	–0.153***	–0.146***	–0.187	0.038	0.041	0.104	822	136
2014	–0.083**	–0.078*	–0.096	0.040	0.042	0.102	719	136
Math, site-based VAM, all available teachers								
2012	–0.025	–0.210***	–0.340*	0.023	0.049	0.142	3,641	140
2013	–0.202***	–0.201***	–0.191	0.029	0.061	0.164	3,305	133
2014	–0.082***	–0.115*	–0.197	0.023	0.064	0.152	3,208	134

Table 4.4—Continued

	Beta			Standard Error				
Year	**Overall**	**Between**	**Within**	**Overall**	**Between**	**Within**	**Teachers**	**Schools**
Reading, site-based VAM, all available teachers								
2012	–0.016	–0.156***	–0.214*	0.016	0.031	0.090	3,642	140
2013	–0.139***	–0.136***	–0.164	0.019	0.038	0.102	3,304	136
2014	–0.058***	–0.080*	–0.030	0.016	0.042	0.094	3,208	135
Math, site-based composite, restricted subsample								
2012	–0.349***	–0.343***	–0.333**	0.051	0.054	0.140	739	140
2013	–0.434***	–0.440***	–0.355**	0.062	0.065	0.173	693	134
2014	–0.229***	–0.197***	–0.411**	0.061	0.064	0.166	596	134
Reading, site-based composite, restricted subsample								
2012	–0.260***	–0.259***	–0.238**	0.033	0.035	0.092	951	140
2013	–0.281***	–0.287***	–0.223**	0.037	0.040	0.101	822	136
2014	–0.181***	–0.168***	–0.246**	0.039	0.041	0.106	719	136

Table 4.4—Continued

Year	Beta			Standard Error			Teachers	Schools
	Overall	Between	Within	Overall	Between	Within		
Math, site-based composite, all available teachers								
2012	–0.199***	–0.338***	–0.501**	0.023	0.054	0.151	3,642	140
2013	–0.411***	–0.454***	–0.229	0.029	0.067	0.176	3,305	134
2014	–0.180***	–0.228***	–0.488**	0.023	0.072	0.187	3,208	134
Reading, site-based composite, all available teachers								
2012	–0.138***	–0.251***	–0.298**	0.016	0.033	0.097	3,641	140
2013	–0.266***	–0.293***	–0.141	0.019	0.041	0.115	3,305	136
2014	–0.126***	–0.184***	–0.224	0.016	0.044	0.114	3,207	136

NOTE: The "Beta" columns show three sorting estimates for each subject and year: the overall coefficient, the between-school coefficient, and the within-school coefficient. The "Standard Error" columns show the standard errors associated with the sorting coefficients. The right two columns present the numbers of teachers and schools in each analysis. *** = $p < 0.01$. ** = $p < 0.05$. * = $p < 0.1$. We used the Benjamini–Hochberg method to adjust significance levels for multiple hypotheses. We restricted the sample to teachers for whom we can calculate VAM using the common formula in Equation 3.1 in Chapter Three.

With regard to the sorting of teacher effectiveness as MCS estimates, we observe the following:

- Using the site-generated value-added estimates and composites, we find that within-school, between-school, and overall sorting estimates were negative from 2012 to 2014 but with a slight upward trend showing improvements in LIM students' access to effective teachers. Although the upward trend cuts negative sorting in half, the change from 2012 to 2014 is not statistically significant.
- The upward trends of the site's overall sorting during these years are very similar to what we report above using the cross-site value-added estimates in mathematics and reading. However, although the trends are similar, the levels are much lower for the site-estimated sorting patterns. Because MCS did not merge with SCS until July 1, 2013, these data are limited to premerger schools, which means that the merger cannot explain these patterns.
- Sorting in terms of the sites' value-added estimates is similar to the sorting of the composite estimates, except that sorting in terms of the composite estimates appears to put LIM students at a slightly greater disadvantage. In other words, estimates of between-school, within-school, and overall sorting are more negative for composites than for the value-added estimates alone.
- One possible explanation for this difference is that the MCS composites are based not only on students' academic growth (35 percent) but also on their academic levels (15 percent) and classroom observations (40 percent), among other measures, as shown in Figure 2.2 in Chapter Two. Because we have adjusted the value-added estimates for students' prior achievement but not their levels and classroom observations, we would expect to see more-negative sorting of the composites than the value-added estimates by students' LIM statuses (Whitehurst, Chingos, and Lindquist, 2014). In other words, because the composite estimate is less fully adjusted for students' prior learning than the value-added estimate, we would expect it to be more correlated with LIM status insofar as students' achievement levels and LIM status are corre-

lated. Another way of thinking about this is that including student achievement and classroom observation as part of the composite without controlling for prior achievement actually weakens the composite measure's accuracy as an estimate of teacher effectiveness. As a consequence, the estimate of LIM sorting by teacher value added (using either our evaluation model or the site-based model) is likely a better representation than the composite sorting parameter of LIM students' access to effective teachers.

Pittsburgh Public Schools in Pennsylvania

Sorting of Value Added in Mathematics and Reading over Time

PPS is the smallest of the three urban districts in the Intensive Partnerships sample. Like we did for HCPS, we can estimate seven years of teacher sorting patterns with PPS data. One reason the sorting values are more volatile across time in PPS is the smaller sample size.

In the rest of this section, we describe the key sorting patterns we identify for Pittsburgh and illustrate them in Table 4.5 and Figure 4.7.

Mathematics

- In preintervention years, between-school sorting was strongly positive, favoring LIM students, and within-school sorting was strongly negative, favoring non-LIM students. The overall effect favored LIM students by a coefficient of 0.17 in 2007–2008. This means that, even before the intervention began, a teacher with a 10-percentage-point higher LIM proportion than other teachers in the same subject and grade in the district had an average value-added estimate that was 1.7 percent of a student-level standard deviation higher than those other teachers, on average.
- The overall sorting of effective teachers to LIM students declined sharply in 2011–2012 but was positive for all other postintervention years.

Reading

- In reading in PPS, sorting between schools has been variable but favored LIM students in the two most-recent years, and sorting

Table 4.5
Pittsburgh Public Schools Sorting Parameters Using Evaluation Value-Added Measure, by Subject and Year

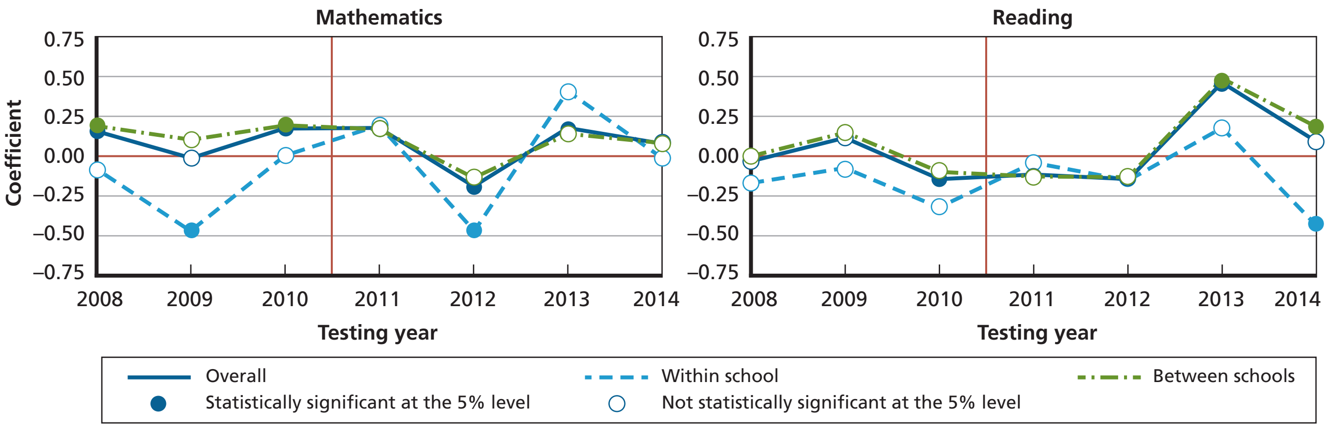
	Beta			Standard Error				
Year	Overall	Between	Within	Overall	Between	Within	Teachers	Schools
Mathematics								
2008	0.169**	0.187**	–0.072	0.070	0.070	0.214	187	52
2009	–0.012	0.107	–0.550**	0.082	0.088	0.187	211	53
2010	0.178**	0.194**	0.010	0.072	0.075	0.215	203	54
2011	0.179*	0.165	0.196	0.086	0.086	0.244	192	53
2012	–0.188**	–0.132	–0.462**	0.069	0.074	0.172	183	51
2013	0.181**	0.143	0.412	0.071	0.073	0.203	161	45
2014	0.080	0.090	–0.008	0.067	0.067	0.164	160	46

Table 4.5—Continued

Year	Beta			Standard Error			Teachers	Schools
	Overall	Between	Within	Overall	Between	Within		
Reading								
2008	–0.028	–0.001	–0.169	0.063	0.065	0.158	251	52
2009	0.119	0.151*	–0.068	0.066	0.067	0.149	264	53
2010	–0.143**	–0.100	–0.324	0.062	0.065	0.148	271	54
2011	–0.115	–0.120	–0.036	0.074	0.075	0.196	238	53
2012	–0.136*	–0.127	–0.154	0.069	0.072	0.171	217	51
2013	0.474***	0.495***	0.187	0.053	0.054	0.164	192	47
2014	0.086	0.178**	–0.443**	0.067	0.067	0.159	196	47

NOTE: The "Beta" columns show three sorting estimates for each subject and year: the overall coefficient, the between-school coefficient, and the within-school coefficient. The "Standard Error" columns show the standard errors associated with the sorting coefficients. The right two columns present the numbers of teachers and schools in each analysis. *** = $p < 0.01$. ** = $p < 0.05$. * = $p < 0.1$. We used the Benjamini–Hochberg method to adjust significance levels for multiple hypotheses.

Figure 4.7
Relationship Between Teacher Effects and the Percentage of Students Who Are Low-Income Minority Students in Pittsburgh Public Schools



NOTE: The red vertical line signifies inception of the Intensive Partnerships initiative. The red horizontal line shows a coefficient of 0, which means that a student's LIM status is unrelated to the effectiveness of the teacher to whom the student is assigned. A positive (i.e., above that line) coefficient means that LIM status is associated with *higher* average teacher value-added estimates; a negative (below the red horizontal) coefficient means that LIM students are being taught by teachers with lower value-added estimates than their non-LIM peers are. The light-blue dashed lines represent the estimated relationship of teacher value-added estimates and students' LIM statuses *within* schools and correspond to parameter β_1' in Equation 3.3 in Chapter Three. The green dash-and-dotted lines represent the estimated relationship of teacher value added in each school and the proportion of students in a school who are classified as LIM, or the *between*-school relationship, and correspond to parameter γ_1 in Equation 3.4 in Chapter Three. The solid dark-blue lines represent overall sorting both between and within schools and correspond to parameter β_1' in Equation 3.2 in Chapter Three.

RAND *RR1295/4-4.7*

within schools has generally disadvantaged them, becoming significantly negative in the most recent year. The weighted average effect has varied from year to year and has been statistically indistinguishable from 0 in many of the years, most recently with a positive but statistically insignificant overall sorting parameter.
- In the years since the Intensive Partnerships intervention began, between-school and overall sorting patterns have been increasingly progressive for the most part, though within-school sorting continues to favor non-LIM students.

One reason the sorting patterns for elementary schools in PPS are fairly erratic is that the sample is so small (Figure A.3 in the appendix). Like we found for all grades together, overall sorting is not significantly different from 0 for either mathematics or reading in the most recent year.

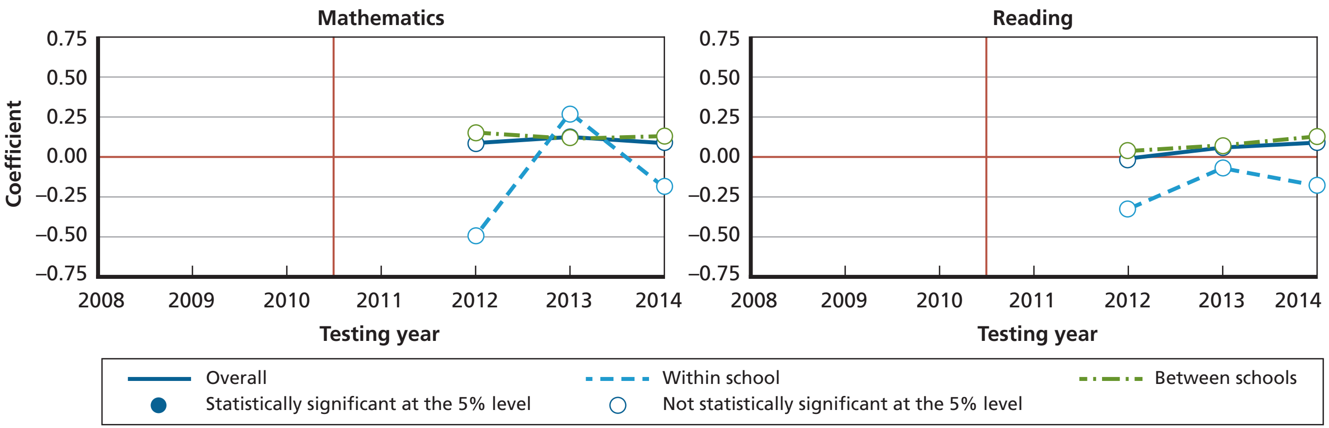
Sorting of Site-Reported Value-Added Measure Estimates and Composites

Like in MCS, the available site-generated value-added estimates and composite measures for PPS pertain to the school years 2011–2012 through 2013–2014. Figures 4.8 and 4.9 display sorting parameters for the sites' value-added estimates and for the sites' teacher effectiveness composites, respectively. The bottom two sections of Table 4.6 show the coefficients, standard errors, and sample sizes corresponding to Figures 4.8 and 4.9.

Examining the sorting of site-reported teacher effectiveness estimates in PPS, we find the following:

- Using the site-reported estimates of teacher effectiveness in PPS, we find that overall and between-school sorting of the value-added estimates favor LIM students but the effectiveness composite sorting favors non-LIM students.
- Within-school sorting disadvantages LIM students, except for in 2012–2013 for the site-provided value added in mathematics. However, the effect was never statistically significant. This

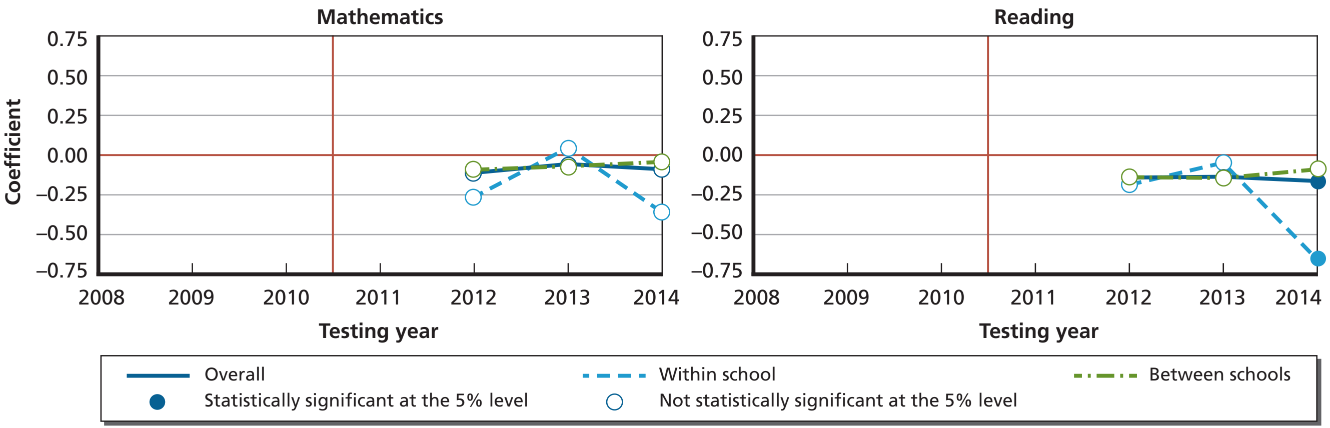
Figure 4.8
Sorting of Site-Provided Value-Added Measure Estimates, by Percentage of Students Who Are Low-Income Minority Students in Pittsburgh Public Schools



NOTE: The red vertical line signifies inception of the Intensive Partnerships initiative. The red horizontal line shows a coefficient of 0, which means that a student's LIM status is unrelated to the effectiveness of the teacher to whom the student is assigned. A positive (i.e., above that line) coefficient means that LIM status is associated with *higher* average teacher value-added estimates; a negative (below the red horizontal) coefficient means that LIM students are being taught by teachers with lower value-added estimates than their non-LIM peers are. The light-blue dashed lines represent the estimated relationship of teacher value-added estimates and students' LIM statuses *within* schools and correspond to parameter β'_1 in Equation 3.3 in Chapter Three. The green dash-and-dotted lines represent the estimated relationship of teacher value added in each school and the proportion of students in a school who are classified as LIM, or the between-school relationship, and correspond to parameter γ_1 in Equation 3.4 in Chapter Three. The solid dark-blue lines represent overall sorting both between and within schools and correspond to parameter β'_1 in Equation 3.2 in Chapter Three.

RAND *RR1295/4-4.8*

Figure 4.9
Sorting of Site-Provided Composite, by Percentage of Students Who Are Low-Income Minority Students in Pittsburgh Public Schools



The red vertical line signifies inception of the Intensive Partnerships initiative. The red horizontal line shows a coefficient of 0, which means that a student's LIM status is unrelated to the effectiveness of the teacher to whom the student is assigned. A positive (i.e., above that line) coefficient means that LIM status is associated with *higher* average teacher value-added estimates; a negative (below the red horizontal) coefficient means that LIM students are being taught by teachers with lower value-added estimates than their non-LIM peers are. The light-blue dashed lines represent the estimated relationship of teacher value-added estimates and students' LIM statuses *within* schools and correspond to parameter β'_1 in Equation 3.3 in Chapter Three. The green dash-and-dotted lines represent the estimated relationship of teacher value added in each school and the proportion of students in a school who are classified as LIM, or the *between*-school relationship, and correspond to parameter γ_1 in Equation 3.4 in Chapter Three. The solid dark-blue lines represent overall sorting both between and within schools and correspond to parameter β'_1 in Equation 3.2 in Chapter Three.

RAND *RR1295/4-4.9*

Table 4.6
Pittsburgh Public Schools Sorting Parameters, by Model and Year

	Beta			Standard Error				
Year	**Overall**	**Between**	**Within**	**Overall**	**Between**	**Within**	**Teachers**	**Schools**
Math, site-based VAM, restricted subsample								
2012	0.086	0.151	–0.488	0.077	0.079	0.229	113	44
2013	0.122	0.112	0.271	0.078	0.083	0.284	104	41
2014	0.088	0.131	–0.195	0.061	0.066	0.169	132	44
Reading, site-based VAM, restricted subsample								
2012	–0.009	0.038	–0.326	0.068	0.073	0.188	134	48
2013	0.060	0.073	–0.073	0.071	0.075	0.235	125	45
2014	0.093	0.130	–0.179	0.056	0.059	0.162	157	46
Math, site-based VAM, all available teachers								
2012	0.011	0.135	–0.410	0.046	0.071	0.198	350	44
2013	0.047	0.116	0.171	0.050	0.086	0.321	307	41
2014	0.065*	0.137	–0.097	0.033	0.068	0.170	507	44

Table 4.6—Continued

Year	Beta			Standard Error			Teachers	Schools
	Overall	Between	Within	Overall	Between	Within		
Reading, site-based VAM, all available teachers								
2012	0.011	0.035	–0.257	0.045	0.068	0.170	350	48
2013	0.047	0.072	–0.162	0.050	0.073	0.243	307	45
2014	0.066*	0.119	–0.113	0.033	0.054	0.148	507	46
Math, site-based composite, restricted subsample								
2012	–0.109	–0.092	–0.263	0.070	0.075	0.225	144	49
2013	–0.057	–0.068	0.049	0.066	0.068	0.209	149	44
2014	–0.086	–0.042	–0.355*	0.059	0.064	0.155	146	45
Reading, site-based composite, restricted subsample								
2012	–0.149**	–0.144*	–0.190	0.056	0.059	0.162	172	50
2013	–0.140**	–0.147*	–0.053	0.062	0.064	0.207	180	46
2014	–0.166**	–0.096	–0.661***	0.061	0.066	0.168	176	46

Table 4.6—Continued

Year	Beta			Standard Error			Teachers	Schools
	Overall	Between	Within	Overall	Between	Within		
Math, site-based composite, all available teachers								
2012	–0.122***	–0.083	–0.418	0.031	0.067	0.196	803	49
2013	–0.149***	–0.070	0.102	0.027	0.070	0.192	871	44
2014	–0.149***	–0.045	–0.203	0.026	0.069	0.171	807	45
Reading, site-based composite, all available teachers								
2012	–0.105***	–0.122*	–0.184	0.027	0.050	0.138	803	50
2013	–0.161***	–0.126*	0.020	0.029	0.055	0.171	871	46
2014	–0.173***	–0.096	–0.670***	0.030	0.066	0.169	807	46

NOTE: The "Beta" columns show three sorting estimates for each subject and year: the overall coefficient, the between-school coefficient, and the within-school coefficient. The "Standard Error" columns show the standard errors associated with the sorting coefficients. The right two columns present the numbers of teachers and schools in each analysis. *** = $p < 0.01$. ** = $p < 0.05$. * = $p < 0.1$. We used the Benjamini–Hochberg method to adjust significance levels for multiple hypotheses. We restricted the sample to teachers for whom we can calculate VAM using the common formula in Equation 3.1 in Chapter Three.

is slightly different from what we find with our own value-added estimates for PPS.

- Like in HCPS and MCS, PPS's teacher effectiveness composites show stronger negative sorting patterns than their site-generated value-added estimates, as seen by contrasting Figure 4.8 with Figure 4.9. The composite components in PPS that are not directly related to student growth are observations (50 percent) and student surveys (15 percent).[5] The fact that sorting appears less favorable to LIM students in terms of the composites than in terms of the sites' value-added estimates alone suggests that LIM students have better access to teachers with strong value added than to teachers with strong scores on these other measures. This finding is consistent with other evidence that classroom observations place teachers of LIM students at a disadvantage (Whitehurst, Chingos, and Lindquist, 2014).

Cross-Site Discussion of Sorting

Overall, Between-School, and Within-School Sorting

Both during and after the intervention in most sites, *between-school* sorting of teacher value added has been somewhat more progressive than *within-school* sorting, and this discrepancy appears slightly more common in mathematics than in reading. It suggests that efforts to improve students' access to highly effective teachers should examine student assignment policies within schools, which has a more detrimental effect than the current distribution of effective teachers between schools. The process by which this within-school sorting likely occurs is not clear from these data. Still, one might imagine that greater advocacy among non-LIM parents, as well as principals' preference for matching the more-skilled teachers to the more–academically rigorous curriculum tracks, might play some role. If these current practices reflect the preferences of the more-influential teachers and community

[5] Five percent is school-level student growth, and the remaining 30 percent is teacher-level student growth.

members, changing within-school sorting without diminishing overall teacher effectiveness and community support could be very challenging. Furthermore, our analysis of the between- and within-school variation in teachers' percentage of LIM students (see Table 2.2 in Chapter Two) shows that there is much less within-school variation, which will also make it difficult to change overall sorting by limiting redistribution to within-school changes. These considerations make it important that both between- and within-school sorting be considered when searching for avenues to further increase access.

Although sorting patterns vary among sites and years, the general trends are a decrease in access immediately following the intervention, followed up with an upward trajectory over time.

Our finding of nonnegative sorting except within schools is at odds with a recent multistate examination of sorting conducted for the National Center for Education Evaluation and Regional Assistance (NCEE) (Isenberg et al., 2013). The authors found negative sorting in most districts and found that between-school sorting was more negative than within-school sorting. However, there are several important differences between the model, analysis, and sample used in the NCEE report and those in the present study. First, we use a value-added model that includes classroom characteristics, such as average prior score and average demographic composition, whereas the primary analysis in the NCEE report does not. However, in the NCEE sensitivity analysis that includes classroom characteristics in the value-added model, the authors no longer found negative sorting (Isenberg et al., 2013, p. C.21, Table C.7).

Also of importance is the fact that we define within-school sorting differently from Isenberg et al., 2013. We measure within-school sorting with a regression of teacher value added on teacher percentage LIM and school-level dummies, so that it reflects the average difference in value added in the district between any two teachers *within the same school* but with differing percentages of LIM students. Following Raudenbush and Bryk, 2002, this permits a decomposition of the overall association of teacher value added and percentage LIM into a weighted average of the between- and within-school associations. The NCEE report uses similar measures to ours of overall and between-school sorting but calculates within-school sorting as the difference between overall and between-school sorting (Isenberg et al., 2013, p. 12). This less familiar formulation of within-school sorting does not have a straightforward interpretation.

Finally, our finding of less negative sorting than Isenberg et al., 2013, reported might be due in part to the select nature of the districts in our sample. The NCEE sample included large districts from across the country, a sample intentionally including a wide array of policies and procedures (Isenberg et al., 2013, p. 19). The Bill & Melinda Gates Foundation intentionally selected the Intensive Partnership sites, on the other hand, in part because of those sites' existing efforts and willingness to improve teaching and more equitably distribute teaching to LIM students.

Sorting by Composite Measures

When we compare sorting patterns using the cross-site evaluation value-added model and the value-added and student growth models that the sites used, we find many similarities but also a few cases in which the site-generated estimates show sorting that is modestly more disadvantageous to LIM students than those shown by our value-added model. Our sorting estimates are modestly more positive than HCPS's and MCS's for mathematics and varies modestly more than PPS. In terms of overall sorting, they are actually quite similar to those for reading in MCS. (Site-generated estimates from MCS show more volatility than our estimates in terms of between- and within-school sorting, but these within-school and between-school estimates are noisily measured.)

Another finding that emerges from the site-generated estimates is that sorting in terms of the effectiveness composites is almost always more negative—that is, less favorable to LIM students—than sorting in terms of our value-added estimates or the sites' achievement growth estimates alone. This suggests that teachers with more LIM students differ less from other teachers in terms of their value added than in terms of other factors used in the composites, such as observation or student survey data. This raises a few questions about possible factors that could favor teachers in more-advantaged settings: First, to what extent are classroom observers consistent across schools and classrooms, and how well do the observation instruments accommodate diverse instructional styles and classroom compositions? In particular, can observers control for differences in demographic makeup of the classrooms they evaluate? Evidence from prior research suggests that this is likely to bias the estimates of sorting that use observation data (Whitehurst, Chingos, and Lindquist, 2014). Second, to what extent do students' and parents' survey data vary by students' prior performance? In other words, do students' and parents' ratings of their teachers depend to some extent on how well their students have performed in school previously? Insofar as the composites offer accurate and reliable estimates of teachers' effectiveness, the fact that sorting appears more negative on these estimates than on value added is troubling. However, if value added is a less biased measure, the fact that value added appears less negatively sorted means that defining sorting in terms of the composites might overstate the problem.

CHAPTER FIVE

Mechanisms That Each Site Used to Change Distribution

The sorting parameters estimated using our own cross-site value-added model suggest that improving LIM students' access to high value–added teachers is challenging because, even before the Intensive Partnerships intervention, teachers were distributed across—and sometimes within—schools in ways that already favored LIM students to a small degree.

Up to this point, we have focused on the distribution of teachers among students and how it changes from year to year, but we have not examined how these changes come about. In this chapter, we first explore whether the LIM composition of a teacher's class changes from one year to the next in a manner that depends on the teacher's prior performance. That is, are the sites assigning teachers and students in a manner that would improve LIM students' access to effective teaching? We do this analysis first because it reflects knowledge that sites could have when doing the assignments: prior performance and the LIM status of students in the prior and upcoming years. Next, we look beyond reassignment and examine whether value added appears to be improving more for teachers who have high percentages of students in their classes who are LIM students. Although sites do not know value added for the upcoming year when making class assignments, the sites can have an effect on the teachers' value added in the coming year by directing PD resources in a targeted fashion.

Changes in Low-Income Minority Composition, by Teacher Performance Level

First, we divide grade 4–8 math and reading teachers into three categories of performance and then compare the LIM composition in their classrooms the following year and the composition in the current year. We do this separately for three time periods: prereform, early reform, and recent reform.[1] Following the reforms, we expect to see more high-performing teachers moving to classes with more LIM students and more low-performing teachers moving to classes with fewer LIM students or moving out of teaching.

This analysis does not show a clear pattern of a differential shift to having more high-LIM classes taught by high-performing teachers or to more low-LIM classes taught by low-performing teachers in the early or recent periods of the reform in any of the sites. Although a few sites exhibit improving shifts in one subject or the other, no site shows access-increasing shifts that are not offset by access-diminishing shifts. For example, after the beginning of the reforms, the fraction of reading teachers in the bottom 20 percent of value added in HCPS who teach fewer LIM students decreases, thereby implying that LIM students have increased access to higher-performing teachers. However, the fraction of higher-performing reading teachers who teach fewer LIM students also increases, offsetting the gain in access. Apparently, all types of returning teachers were more likely to teach fewer LIM students.[2] To more fully understand the changes in access, we now turn

---

[1] For all sites except MCS, prereform is the average of changes from 2007–2008 to 2008–2009 and from 2008–2009 to 2009–2010; early reform is the average of changes from 2009–2010 to 2010–2011 and from 2010–2011 to 2011–2012; and recent reform is the average of changes from 2011–2012 to 2012–2013 and 2012–2013 to 2013–2014. For MCS, prereform is the average of changes from 2008–2009 to 2009–2010 and 2009–2010 to 2010–2011; early reform is the average of changes from 2010–2011 to 2011–2012; and recent reform is the average of changes from 2011–2012 to 2012–2013 and 2012–2013 to 2013–2014.

[2] In the appendix, Figures A.4 through A.12 present our findings for each site and for the various performance measures. For each site, we present the findings separately for our estimates of reading and math VAM in three periods. We present the findings for the sites'

to an analysis that accounts for all possible ways in which access can change.

Decomposing Changes in Access

We examine all the possible changes from one year to the next that can lead to an increase or a decrease in the association between teacher performance and the percentage of students they teach who are LIM students. The solid lines in Figures 4.1, 4.4, and 4.7 (and Figure A.22 in the appendix) show whether the overall sorting (i.e., the weighted average of between and within) increases or decreases from one year to the next. We decompose this change into three possible sources: (1) changes in value added that differ between teachers with many and few LIM students, (2) changes in LIM enrollment that differ between teachers with high and low value-added scores, and (3) changes in teaching staff that change the association between value added and percentage LIM. We refer to the actions behind these three sources as *improve*, *reassign*, and *replace*, respectively. Each of these site actions can change the access that LIM students have to effective teachers. In the appendix, we provide a detailed explanation of how we calculated this decomposition.

We find very little evidence of changes that are consistent with any of these actions. After correcting for multiple-hypothesis testing, we find that MCS reading is one case in which we see changes that are consistent with specific actions to improve access (see Figure A.16 and Table A.13 in the appendix). In recent years, the district is both replacing ineffective teachers in high-LIM classrooms with more-effective teachers and improving the value added of returning teachers in high-LIM classrooms. These actions explain much more of the increase in access than the third action of reassigning high value–added teachers to high-LIM classrooms does.

PPS also shows changes that reflect improvement in some actions in recent years (see Figure A.19 and Table A.14 in the appendix). In math, all three changes reflect actions to increase access from the early

---

performance measures (i.e., their VAMs and their composite measures) for the periods for which we have data. Tables A.3 through A.19 present percentages and standard errors.

phase of the intervention to the recent years. The effect of replacing departing math teachers with more-effective teachers in high-LIM classrooms is particularly large, although not quite significant. Also, in the early phase, the improvement in value added was significantly lower for teachers in high-LIM classrooms than in low-LIM classrooms, but that is no longer the case. In reading, there also is a positive trend in the effect of replacement, but it is not significant.

The estimated effects of the actions in HCPS are very small and never significant in the desired direction. However, they are replacing teachers in high-LIM classrooms with lower value–added teachers.

We also examine whether the two site-provided effectiveness measures are associated with these mechanisms that change access. The two measures are available only following the reforms and, in some cases, only in the latter years of the reform. The largest significant effect in the recent period is a 0.05 increase in the correlation between the site-provided VAM and percentage LIM for mathematics teachers in HCPS attributable to a differential improvement in this VAM for teachers of high-LIM classes.

Although we have corrected for multiple-hypothesis testing, we emphasize that one must be cautious when interpreting the few significant findings in the correct direction out of so many parameter estimates.

CHAPTER SIX

Conclusion

First, given our cross-site value-added estimates in mathematics and reading, we find that, in the years before and after Intensive Partnerships implementation, LIM and non-LIM students experienced similar levels of teacher quality in most of the sites. However, there are small but significant advantages for LIM students in some sites, subjects, and years. In general, across sites, between-school sorting of teacher effectiveness has been more favorable to LIM students than sorting within schools. These findings differ from those in a recent NCEE report on differential access to effective teaching (Isenberg et al., 2013). As we explained earlier, this difference is not unexpected because of differences in value-added modeling choices and definitions of sorting parameters and because of the Bill & Melinda Gates Foundation's intentional choice of innovative sites.

Since Intensive Partnerships implementation began, LIM students' access to effective teachers has shown little change in HCPS; has increased slightly in PPS (especially between 2012 and 2013); and initially decreased in MCS but, in the past year, has experienced an upward movement leading to favorable sorting.

Moreover, although our estimates of overall teacher sorting by student LIM status tend to be slightly to moderately more positive than those based on site-generated values, the variations in sorting from year to year that we estimate are strikingly similar to the short-term variations we can estimate with site-generated estimates. A possible explanation for our estimates being more positive and variable than estimates from the sites is that we are controlling for student covariates at both the student and classroom levels (facilitated by our multistage model) and that our estimates are based only on a teacher's current-year students. This allows us to capture short-term changes in the association between effectiveness and student LIM composition.

We also find that the sites' composite effectiveness measures show sorting patterns that are more negative than either our own value-added models or those of the sites. This raises questions about whether the other elements of the composites are capturing true differences that are unequally distributed by students' LIM status or whether the LIM status of students in the classroom inappropriately affects the other components (such as classroom-observation scores). The latter interpretation is consistent with other recent research (Whitehurst, Chingos, and Lindquist, 2014).

We also examine the possible mechanisms by which districts are making changes in the association of effective teaching and LIM status. First, we categorize teachers by their current VAM (or other performance measure) and examine whether they teach a higher or lower percentage of LIM students in the following year. Even though the sites do not have access to our value-added calculations for each of their teachers, we would expect these percentages to change if the sites were making assignments based on a similar measure of teaching performance. We examine the trend in this association from before the reforms up through the most-recent years of our data. In general, we do not find any evidence that performance level is positively associated with the following year's LIM percentage, as would be the case if sites were making assignments based on performance.

Second, we decompose the change in the association between performance and LIM percentage into three components, each reflecting a possible mechanism by which the sites could affect a change in the equity of access: (1) replace some teachers so that LIM students have better teachers, (2) improve the performance of LIM students' teachers perhaps through targeted PD, and (3) reassign LIM students to higher-performing teachers. (This third component was the only mechanism that the prior analysis captured.) In several cases, we see significant evidence of one or more mechanisms being used in recent years. In other cases, we see that mechanisms that were reducing LIM students' access to highly effective teachers are no longer evident. For example, in recent years, LIM students in MCS are being taught by more-effective reading teachers and reading teachers of LIM students are improving more than those of non-LIM students. LIM students in PPS are being taught by higher-performing teachers. The only mechanism that is having a negative and significant effect on equity in recent years is the replacement of math teachers in HCPS, where the performance difference between incoming and outgoing teachers favors non-LIM students over LIM students. Although these significant findings account for false discovery rates caused by testing multiple hypotheses, we still encourage caution when such a low proportion of our estimates are statistically significant.

This report has demonstrated that there is little evidence of increases in LIM students' access to more-effective teachers in the districts that are participating in the Intensive Partnerships initiative. We find that LIM students have access to effective teaching that is roughly equal to what they had before the initiative began; efforts to improve access have not produced significant gains. Identifying the possible reasons that gains have not been achieved is beyond the scope of this report. However, we have shown that gains in access have not occurred either by assigning more-effective teachers to classes with higher percentages of LIM students within schools, nor have gains been realized by moving more-effective teachers to schools with higher percentages of LIM students. With few exceptions, districts have not improved access by strategically removing and replacing teachers, by strategically improving teachers, or by strategically reassigning teachers or students.

There are many reasons that improving access might be difficult. For example, in some sites, collective bargaining agreements regulate the movement of teachers among schools and the removal of teachers. In some sites, either contract or convention makes it difficult or impossible to provide incentives to move effective teachers to classes that are more challenging to teach, which would likely include more students from less affluent backgrounds. In some cases, a teacher has the option of transferring to a different school if a principal asked that teacher to teach a more challenging class. Above all, only recently have district and school leaders obtained detailed information on teacher effectiveness that would allow them to remove, improve, or reassign teachers in a way that improves access. However, two years remain in the initiative during which the sites can demonstrate that they can use teacher effectiveness information to improve LIM students' access to more-effective teachers.

APPENDIX

Additional Details and Results

Distribution of Sections Among Teachers

Tables A.1 and A.2 show the mean number of sections per teacher per year in mathematics and reading, respectively, by site and grade. They also show the proportion of teachers in each site and grade who teach more than one section in a given subject each year. As the tables demonstrate, a substantial number of teachers have multiple sections in a

Table A.1
Sections per Math Teacher and Share with More Than One Section, 2013–2014

Grade	Variable	HCPS	MCS	PPS	Aspire
4	Mean sections	1.90	1.92	1.75	1.08
	Share with >1 section	0.69	0.50	0.62	0.08
5	Mean sections	2.05	2.09	1.89	1.08
	Share with >1 section	0.77	0.57	0.68	0.08
6	Mean sections	11.61	3.94	2.60	4.09
	Share with >1 section	1.00	0.90	0.85	0.91
7	Mean sections	12.05	4.06	2.52	5.73
	Share with >1 section	1.00	0.89	0.84	0.98
8	Mean sections	10.94	4.34	2.42	3.98
	Share with >1 section	1.00	0.91	0.77	0.96

Table A.2
Sections per Reading Teacher and Share with More Than One Section, 2013–2014

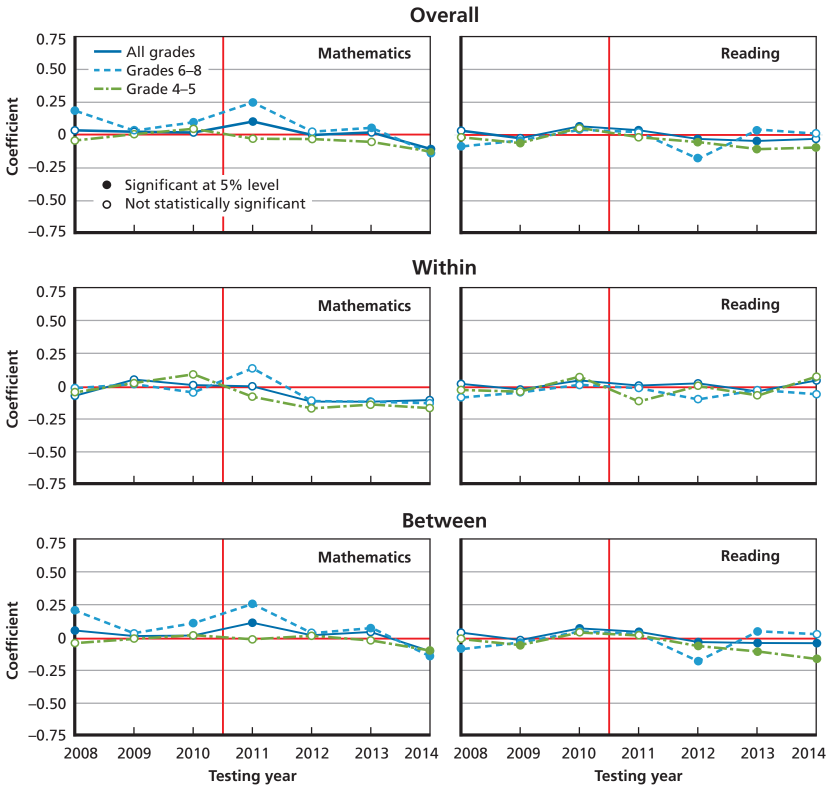
Grade	Variable	HCPS	MCS	PPS	Aspire
4	Mean sections	2.04	1.97	1.85	1.08
	Share with >1 section	0.68	0.50	0.65	0.08
5	Mean sections	2.08	2.17	2.06	1.08
	Share with >1 section	0.74	0.57	0.77	0.08
6	Mean sections	11.01	4.03	2.28	3.71
	Share with >1 section	1.00	0.89	0.85	0.90
7	Mean sections	11.02	4.10	2.41	4.09
	Share with >1 section	1.00	0.91	0.85	1.00
8	Mean sections	10.96	4.13	2.38	4.11
	Share with >1 section	1.00	0.89	0.90	1.00

given year, which allows us to separate the average effects of classroom-level covariates across classrooms from teachers' value-added estimates. Given that we estimate teachers' value added net of these classroom-level characteristics, we can then fit subsequent regression models that examine how teachers' value added is related to the demographic attributes of the students they teach.

Analysis of Sorting, Separately by Grade Level

Here, we separately estimate the sorting coefficients by grade level. As shown in Figure A.1, in mathematics, sorting is slightly more progressive for grades 6 through 8 than for younger grades, especially in 2011. In reading, there are no consistent differences between the younger and older grades. These results are at odds with the hypothesis that tracking in higher grades drives worse sorting for LIM students.

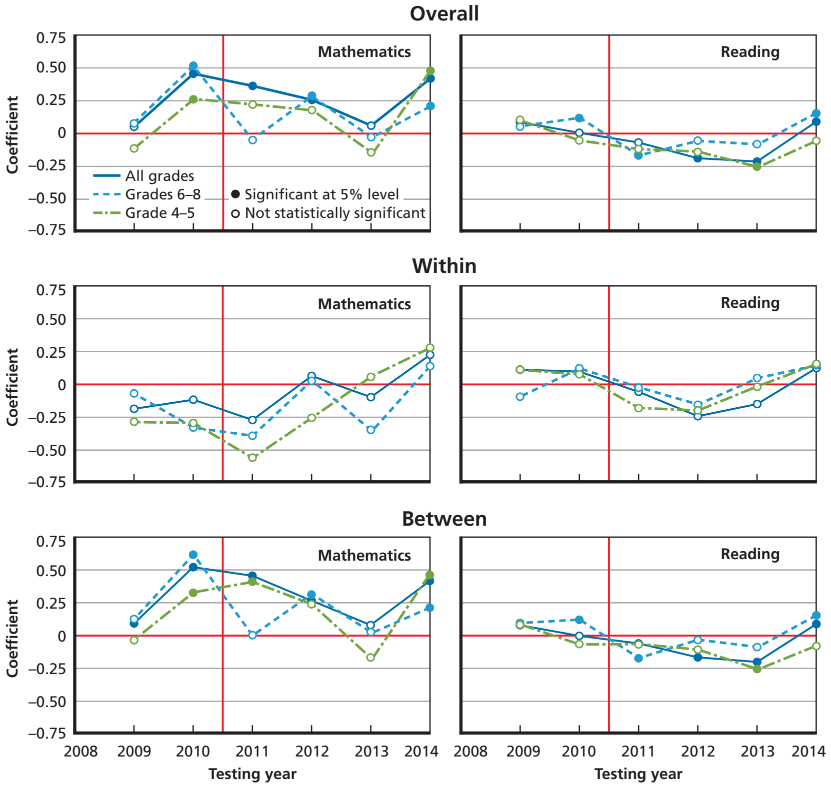
Figure A.1
Relationship Between Teacher Effects and the Percentage of Students Who Are Low-Income Minority Students, by Grade Level in Hillsborough County Public Schools



NOTE: The red vertical line signifies inception of the Intensive Partnerships initiative. The red horizontal line shows a coefficient of 0, which means that a student's LIM status is unrelated to the effectiveness of the teacher to whom the student is assigned. A positive (i.e., above that line) coefficient means that LIM status is associated with *higher* average teacher value-added estimates; a negative (below the red horizontal) coefficient means that LIM students are being taught by teachers with lower value-added estimates than their non-LIM peers are. The light-blue dashed lines represent the estimated relationship of teacher value-added estimates and students' LIM statuses *within* schools and correspond to parameter β'_1 in Equation 3.3 in Chapter Three. The green dash-and-dotted lines represent the estimated relationship of teacher value added in each school and the proportion of students in a school who are classified as LIM, or the *between*-school relationship, and correspond to parameter γ_1 in Equation 3.4 in Chapter Three. The solid dark-blue lines represent overall sorting both between and within schools and correspond to parameter β_1 in Equation 3.2 in Chapter Three.

RAND *RR1295/4-A.1*

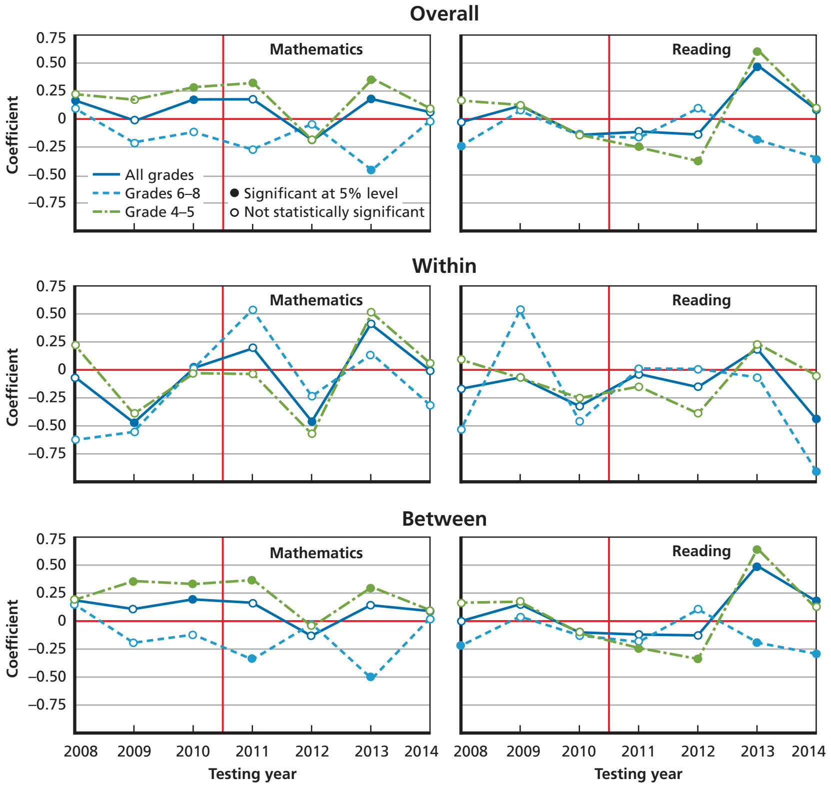
Figure A.2
Relationship Between Teacher Effects and the Percentage of Students Who Are Low-Income Minority Students, by Grade Level in Memphis City Schools



NOTE: The red vertical line signifies inception of the Intensive Partnerships initiative. The red horizontal line shows a coefficient of 0, which means that a student's LIM status is unrelated to the effectiveness of the teacher to whom the student is assigned. A positive (i.e., above that line) coefficient means that LIM status is associated with *higher* average teacher value-added estimates; a negative (below the red horizontal) coefficient means that LIM students are being taught by teachers with lower value-added estimates than their non-LIM peers are. The light-blue dashed lines represent the estimated relationship of teacher value-added estimates and students' LIM statuses *within* schools and correspond to parameter β'_1 in Equation 3.3 in Chapter Three. The green dash-and-dotted lines represent the estimated relationship of teacher value added in each school and the proportion of students in a school who are classified as LIM, or the *between*-school relationship, and correspond to parameter γ_1 in Equation 3.4 in Chapter Three. The solid dark-blue lines represent overall sorting both between and within schools and correspond to parameter β_1 in Equation 3.2 in Chapter Three.

RAND *RR1295/4-A.2*

Figure A.3
Relationship Between Teacher Effects and the Percentage of Students Who Are Low-Income Minority Students, by Grade Level in Pittsburgh Public Schools



NOTE: The red vertical line signifies inception of the Intensive Partnerships initiative. The red horizontal line shows a coefficient of 0, which means that a student's LIM status is unrelated to the effectiveness of the teacher to whom the student is assigned. A positive (i.e., above that line) coefficient means that LIM status is associated with *higher* average teacher value-added estimates; a negative (below the red horizontal) coefficient means that LIM students are being taught by teachers with lower value-added estimates than their non-LIM peers are. The light-blue dashed lines represent the estimated relationship of teacher value-added estimates and students' LIM statuses *within* schools and correspond to parameter β'_1 in Equation 3.3 in Chapter Three. The green dash-and-dotted lines represent the estimated relationship of teacher value added in each school and the proportion of students in a school who are classified as LIM, or the *between*-school relationship, and correspond to parameter γ_1 in Equation 3.4 in Chapter Three. The solid dark-blue lines represent overall sorting both between and within schools and correspond to parameter β_1 in Equation 3.2 in Chapter Three.

RAND *RR1295/4-A.3*

Changes in Low-Income Minority Composition

Table A.3
Hillsborough County Public Schools Changes in Low-Income Minority Composition, by Performance Level: Value-Added Measure Tercile

	Prereform				Early Reform				Recent Reform			
Change	**1**	**2**	**3**	**Total**	**1**	**2**	**3**	**Total**	**1**	**2**	**3**	**Total**
Mathematics												
Increase LIM, same school	0.360	0.412	0.364	0.392	0.305	0.361	0.325	0.342	0.270	0.350	0.332	0.330
	(0.031)	(0.017)	(0.031)	(0.013)	(0.032)	(0.018)	(0.032)	(0.014)	(0.035)	(0.019)	(0.033)	(0.015)
Decrease LIM, same school	0.200	0.271	0.287	0.260	0.271	0.308	0.330	0.305	0.262	0.295	0.324	0.294
	(0.034)	(0.019)	(0.032)	(0.015)	(0.033)	(0.019)	(0.032)	(0.014)	(0.035)	(0.020)	(0.034)	(0.015)
Increase LIM, new school	0.028	0.028	0.028	0.028	0.012	0.018	0.015	0.016	0.026	0.019	0.015	0.020
	(0.038)	(0.022)	(0.038)	(0.017)	(0.039)	(0.022)	(0.039)	(0.017)	(0.040)	(0.023)	(0.040)	(0.018)
Decrease LIM, new school	0.025	0.024	0.024	0.024	0.015	0.026	0.026	0.024	0.030	0.023	0.018	0.023
	(0.038)	(0.022)	(0.038)	(0.017)	(0.039)	(0.022)	(0.038)	(0.017)	(0.040)	(0.023)	(0.040)	(0.018)
Out of tested	0.154	0.107	0.109	0.117	0.195	0.143	0.145	0.154	0.235	0.161	0.161	0.176
	(0.035)	(0.021)	(0.036)	(0.016)	(0.035)	(0.021)	(0.036)	(0.016)	(0.036)	(0.022)	(0.037)	(0.017)
Out of district	0.232	0.159	0.189	0.180	0.202	0.144	0.159	0.159	0.177	0.152	0.150	0.157
	(0.034)	(0.020)	(0.035)	(0.016)	(0.035)	(0.021)	(0.036)	(0.016)	(0.037)	(0.022)	(0.038)	(0.017)

Table A.3—Continued

	Prereform				Early Reform				Recent Reform			
Change	**1**	**2**	**3**	**Total**	**1**	**2**	**3**	**Total**	**1**	**2**	**3**	**Total**
Total proportion	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Total number of schools	680	2,039	679	3,398	663	1,987	661	3,311	604	1,809	602	3,015
Reading												
Increase LIM, same school	0.354	0.434	0.419	0.415	0.318	0.378	0.336	0.358	0.275	0.334	0.330	0.321
	(0.028)	(0.015)	(0.027)	(0.012)	(0.029)	(0.016)	(0.028)	(0.012)	(0.031)	(0.017)	(0.030)	(0.013)
Decrease LIM, same school	0.252	0.258	0.273	0.260	0.301	0.292	0.333	0.302	0.260	0.301	0.299	0.292
	(0.030)	(0.017)	(0.030)	(0.014)	(0.029)	(0.017)	(0.028)	(0.013)	(0.031)	(0.017)	(0.030)	(0.014)
Increase LIM, new school	0.025	0.031	0.015	0.026	0.024	0.013	0.013	0.016	0.025	0.020	0.013	0.019
	(0.035)	(0.020)	(0.035)	(0.016)	(0.034)	(0.020)	(0.035)	(0.015)	(0.036)	(0.021)	(0.036)	(0.016)
Decrease LIM, new school	0.038	0.025	0.015	0.026	0.021	0.033	0.029	0.030	0.034	0.024	0.026	0.026
	(0.034)	(0.020)	(0.035)	(0.016)	(0.034)	(0.020)	(0.034)	(0.015)	(0.036)	(0.021)	(0.036)	(0.016)
Out of tested	0.149	0.116	0.128	0.125	0.181	0.148	0.150	0.155	0.202	0.155	0.152	0.164
	(0.032)	(0.019)	(0.033)	(0.015)	(0.032)	(0.019)	(0.032)	(0.014)	(0.032)	(0.019)	(0.033)	(0.015)

Table A.3—Continued

	Prereform				Early Reform				Recent Reform			
Change	**1**	**2**	**3**	**Total**	**1**	**2**	**3**	**Total**	**1**	**2**	**3**	**Total**
Out of district	0.182	0.136	0.151	0.148	0.154	0.135	0.139	0.140	0.204	0.168	0.180	0.177
	(0.032)	(0.019)	(0.032)	(0.015)	(0.032)	(0.019)	(0.032)	(0.014)	(0.032)	(0.019)	(0.033)	(0.015)
Total proportion	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Total number of schools	808	2,426	807	4,041	823	2,467	822	4,112	766	2,296	763	3,825

NOTE: The numbers in parentheses are standard errors.

Table A.4
Memphis City Schools Changes in Low-Income Minority Composition, by Performance Level: Value-Added Measure Tercile

	Prereform				Early Reform				Recent Reform			
Change	**1**	**2**	**3**	**Total**	**1**	**2**	**3**	**Total**	**1**	**2**	**3**	**Total**
Mathematics												
Increase LIM, same school	0.297	0.288	0.288	0.290	0.170	0.247	0.283	0.239	0.143	0.217	0.236	0.206
	(0.044)	(0.025)	(0.044)	(0.020)	(0.074)	(0.041)	(0.069)	(0.032)	(0.053)	(0.029)	(0.050)	(0.023)
Decrease LIM, same school	0.224	0.271	0.310	0.269	0.327	0.309	0.408	0.332	0.153	0.219	0.243	0.211
	(0.046)	(0.026)	(0.043)	(0.020)	(0.066)	(0.039)	(0.062)	(0.030)	(0.052)	(0.029)	(0.050)	(0.023)
Increase LIM, new school	0.027	0.043	0.060	0.043	0.013	0.035	0.013	0.026	0.094	0.091	0.129	0.099
	(0.051)	(0.029)	(0.051)	(0.023)	(0.080)	(0.046)	(0.081)	(0.036)	(0.054)	(0.031)	(0.053)	(0.024)
Decrease LIM, new school	0.030	0.070	0.041	0.056	0.046	0.072	0.053	0.063	0.068	0.146	0.142	0.130
	(0.051)	(0.029)	(0.051)	(0.023)	(0.079)	(0.045)	(0.079)	(0.035)	(0.055)	(0.030)	(0.053)	(0.024)
Out of tested	0.197	0.138	0.128	0.148	0.216	0.147	0.118	0.155	0.256	0.141	0.100	0.156
	(0.047)	(0.028)	(0.049)	(0.021)	(0.072)	(0.043)	(0.076)	(0.033)	(0.049)	(0.030)	(0.054)	(0.023)
Out of district	0.224	0.191	0.174	0.194	0.229	0.190	0.125	0.185	0.286	0.186	0.149	0.198
	(0.046)	(0.027)	(0.047)	(0.021)	(0.071)	(0.042)	(0.076)	(0.033)	(0.048)	(0.030)	(0.052)	(0.023)

Table A.4—Continued

Change	Prereform 1	Prereform 2	Prereform 3	Prereform Total	Early Reform 1	Early Reform 2	Early Reform 3	Early Reform Total	Recent Reform 1	Recent Reform 2	Recent Reform 3	Recent Reform Total
Total proportion	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Total number of schools	370	1,105	368	1,843	153	457	152	762	308	926	309	1,543
Reading												
Increase LIM, same school	0.308	0.334	0.295	0.321	0.221	0.252	0.270	0.249	0.134	0.200	0.272	0.201
	(0.039)	(0.022)	(0.040)	(0.017)	(0.064)	(0.036)	(0.062)	(0.028)	(0.048)	(0.026)	(0.044)	(0.021)
Decrease LIM, same school	0.247	0.287	0.356	0.293	0.347	0.349	0.397	0.358	0.166	0.197	0.201	0.191
	(0.041)	(0.023)	(0.038)	(0.018)	(0.059)	(0.034)	(0.056)	(0.026)	(0.047)	(0.027)	(0.046)	(0.021)
Increase LIM, new school	0.016	0.034	0.018	0.027	0.016	0.025	0.011	0.020	0.097	0.093	0.092	0.094
	(0.047)	(0.027)	(0.047)	(0.021)	(0.072)	(0.041)	(0.072)	(0.032)	(0.049)	(0.028)	(0.049)	(0.022)
Decrease LIM, new school	0.029	0.041	0.023	0.035	0.042	0.041	0.032	0.039	0.113	0.126	0.132	0.124
	(0.047)	(0.027)	(0.047)	(0.021)	(0.071)	(0.041)	(0.072)	(0.032)	(0.048)	(0.028)	(0.048)	(0.021)
Out of tested	0.211	0.145	0.131	0.155	0.221	0.164	0.138	0.170	0.253	0.173	0.124	0.179
	(0.042)	(0.025)	(0.044)	(0.020)	(0.064)	(0.038)	(0.068)	(0.030)	(0.044)	(0.027)	(0.048)	(0.021)

Table A.4—Continued

	Prereform				Early Reform				Recent Reform			
Change	**1**	**2**	**3**	**Total**	**1**	**2**	**3**	**Total**	**1**	**2**	**3**	**Total**
Out of district	0.189	0.159	0.178	0.169	0.153	0.169	0.153	0.163	0.237	0.212	0.179	0.210
	(0.043)	(0.025)	(0.043)	(0.019)	(0.067)	(0.038)	(0.067)	(0.030)	(0.045)	(0.026)	(0.047)	(0.020)
Total proportion	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Total number of schools	445	1,332	444	2,221	190	567	189	946	380	1,139	379	1,898

NOTE: The numbers in parentheses are standard errors.

Table A.5
Pittsburgh Public Schools Changes in Low-Income Minority Composition, by Performance Level: Value-Added Measure Tercile

	Prereform				Early Reform				Recent Reform			
Change	**1**	**2**	**3**	**Total**	**1**	**2**	**3**	**Total**	**1**	**2**	**3**	**Total**
Mathematics												
Increase LIM, same school	0.259	0.308	0.443	0.325	0.263	0.278	0.295	0.278	0.243	0.340	0.382	0.328
	(0.096)	(0.054)	(0.084)	(0.041)	(0.096)	(0.055)	(0.095)	(0.043)	(0.104)	(0.057)	(0.095)	(0.044)
Decrease LIM, same school	0.321	0.312	0.291	0.310	0.175	0.371	0.410	0.339	0.286	0.296	0.397	0.314
	(0.092)	(0.054)	(0.095)	(0.042)	(0.102)	(0.052)	(0.087)	(0.041)	(0.101)	(0.058)	(0.094)	(0.045)
Increase LIM, new school	0.012	0.038	0.013	0.028	0.025	0.021	0.000	0.018	0.000	0.034	0.029	0.026
	(0.110)	(0.064)	(0.112)	(0.049)	(0.110)	(0.064)	(0.000)	(0.050)	(0.000)	(0.068)	(0.119)	(0.053)
Decrease LIM, new school	0.049	0.034	0.038	0.038	0.075	0.068	0.026	0.061	0.029	0.029	0.029	0.029
	(0.108)	(0.064)	(0.110)	(0.049)	(0.108)	(0.063)	(0.112)	(0.049)	(0.118)	(0.069)	(0.119)	(0.053)
Out of tested	0.062	0.118	0.076	0.098	0.138	0.076	0.077	0.089	0.157	0.126	0.059	0.119
	(0.108)	(0.061)	(0.108)	(0.048)	(0.104)	(0.062)	(0.109)	(0.048)	(0.110)	(0.065)	(0.118)	(0.051)
Out of district	0.296	0.190	0.139	0.202	0.325	0.186	0.192	0.215	0.286	0.175	0.103	0.183
	(0.093)	(0.058)	(0.104)	(0.045)	(0.092)	(0.059)	(0.102)	(0.045)	(0.101)	(0.063)	(0.115)	(0.049)

Table A.5—Continued

	Prereform				Early Reform				Recent Reform			
Change	**1**	**2**	**3**	**Total**	**1**	**2**	**3**	**Total**	**1**	**2**	**3**	**Total**
Total proportion	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Total number of schools	81	237	79	397	80	237	78	395	70	206	68	344
Reading												
Increase LIM, same school	0.212	0.351	0.324	0.317	0.204	0.331	0.366	0.312	0.313	0.318	0.383	0.330
	(0.087)	(0.046)	(0.081)	(0.036)	(0.088)	(0.047)	(0.079)	(0.037)	(0.091)	(0.053)	(0.087)	(0.040)
Decrease LIM, same school	0.327	0.286	0.333	0.304	0.223	0.331	0.376	0.318	0.277	0.294	0.383	0.308
	(0.080)	(0.048)	(0.081)	(0.037)	(0.087)	(0.047)	(0.079)	(0.037)	(0.093)	(0.054)	(0.087)	(0.041)
Increase LIM, new school	0.010	0.039	0.020	0.029	0.049	0.020	0.020	0.026	0.024	0.033	0.037	0.032
	(0.098)	(0.056)	(0.098)	(0.043)	(0.096)	(0.057)	(0.099)	(0.044)	(0.108)	(0.063)	(0.109)	(0.049)
Decrease LIM, new school	0.048	0.026	0.029	0.031	0.049	0.026	0.000	0.026	0.012	0.033	0.025	0.027
	(0.096)	(0.056)	(0.098)	(0.043)	(0.096)	(0.057)	(0.000)	(0.044)	(0.109)	(0.063)	(0.110)	(0.049)
Out of tested	0.135	0.094	0.078	0.099	0.272	0.138	0.079	0.153	0.169	0.180	0.037	0.149
	(0.091)	(0.054)	(0.095)	(0.042)	(0.084)	(0.053)	(0.095)	(0.041)	(0.100)	(0.058)	(0.109)	(0.046)

Table A.5—Continued

	Prereform				Early Reform				Recent Reform			
Change	**1**	**2**	**3**	**Total**	**1**	**2**	**3**	**Total**	**1**	**2**	**3**	**Total**
Out of district	0.269	0.205	0.216	0.220	0.204	0.154	0.158	0.165	0.205	0.143	0.136	0.154
	(0.084)	(0.051)	(0.088)	(0.039)	(0.088)	(0.053)	(0.091)	(0.041)	(0.098)	(0.059)	(0.103)	(0.045)
Total proportion	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Total number of schools	104	308	102	514	103	305	101	509	83	245	81	409

NOTE: The numbers in parentheses are standard errors.

Table A.6
Hillsborough County Public Schools Changes in Low-Income Minority Composition, by Performance Level: Site Value-Added Measure Tercile

	Early Reform				Recent Reform			
Change	**1**	**2**	**3**	**Total**	**1**	**2**	**3**	**Total**
Mathematics								
Increase LIM, same school	0.289	0.321	0.404	0.331	0.322	0.323	0.362	0.331
	(0.047)	(0.026)	(0.043)	(0.020)	(0.034)	(0.019)	(0.033)	(0.015)
Decrease LIM, same school	0.289	0.303	0.327	0.305	0.257	0.290	0.346	0.295
	(0.047)	(0.027)	(0.046)	(0.021)	(0.035)	(0.020)	(0.033)	(0.015)
Increase LIM, new school	0.015	0.013	0.009	0.013	0.018	0.023	0.012	0.020
	(0.055)	(0.032)	(0.055)	(0.025)	(0.040)	(0.023)	(0.041)	(0.018)
Decrease LIM, new school	0.025	0.026	0.019	0.024	0.023	0.025	0.018	0.023
	(0.055)	(0.032)	(0.055)	(0.025)	(0.040)	(0.023)	(0.040)	(0.018)
Out of tested	0.218	0.182	0.142	0.181	0.202	0.181	0.132	0.175
	(0.049)	(0.029)	(0.051)	(0.022)	(0.036)	(0.021)	(0.038)	(0.017)
Out of district	0.163	0.155	0.099	0.145	0.177	0.158	0.130	0.156
	(0.051)	(0.029)	(0.053)	(0.023)	(0.037)	(0.022)	(0.038)	(0.017)

Table A.6—Continued

Change	Early Reform 1	Early Reform 2	Early Reform 3	Early Reform Total	Recent Reform 1	Recent Reform 2	Recent Reform 3	Recent Reform Total
Total proportion	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Total number of schools	325	974	324	1,623	603	1,806	599	3,008
Reading								
Increase LIM, same school	0.331	0.367	0.383	0.363	0.314	0.319	0.342	0.323
	(0.040)	(0.023)	(0.039)	(0.018)	(0.030)	(0.017)	(0.029)	(0.013)
Decrease LIM, same school	0.258	0.278	0.329	0.284	0.273	0.291	0.321	0.293
	(0.042)	(0.024)	(0.040)	(0.019)	(0.031)	(0.018)	(0.030)	(0.014)
Increase LIM, new school	0.024	0.015	0.010	0.015	0.026	0.021	0.007	0.019
	(0.049)	(0.028)	(0.049)	(0.022)	(0.036)	(0.021)	(0.036)	(0.016)
Decrease LIM, new school	0.053	0.032	0.019	0.034	0.026	0.028	0.021	0.026
	(0.048)	(0.028)	(0.049)	(0.022)	(0.036)	(0.021)	(0.036)	(0.016)
Out of tested	0.196	0.175	0.131	0.170	0.185	0.163	0.142	0.163
	(0.044)	(0.026)	(0.046)	(0.020)	(0.033)	(0.019)	(0.034)	(0.015)
Out of district	0.138	0.134	0.128	0.133	0.176	0.178	0.167	0.175
	(0.046)	(0.026)	(0.046)	(0.020)	(0.033)	(0.019)	(0.033)	(0.015)

Table A.6—Continued

	Early Reform				Recent Reform			
Change	**1**	**2**	**3**	**Total**	**1**	**2**	**3**	**Total**
Total proportion	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Total number of schools	414	1,241	413	2,068	762	2,283	760	3,805

NOTE: The numbers in parentheses are standard errors.

Table A.7
Hillsborough County Public Schools Changes in Low-Income Minority Composition, by Performance Level: Site Composite Tercile

	Early Reform				Recent Reform			
Change	**1**	**2**	**3**	**Total**	**1**	**2**	**3**	**Total**
Mathematics								
Increase LIM, same school	0.283	0.335	0.370	0.331	0.282	0.337	0.359	0.331
	(0.047)	(0.026)	(0.044)	(0.020)	(0.035)	(0.019)	(0.033)	(0.015)
Decrease LIM, same school	0.249	0.309	0.349	0.305	0.235	0.286	0.377	0.294
	(0.048)	(0.027)	(0.045)	(0.021)	(0.036)	(0.020)	(0.032)	(0.015)
Increase LIM, new school	0.025	0.010	0.009	0.013	0.020	0.024	0.008	0.020
	(0.055)	(0.032)	(0.055)	(0.025)	(0.040)	(0.023)	(0.041)	(0.018)
Decrease LIM, new school	0.018	0.025	0.028	0.024	0.017	0.029	0.012	0.023
	(0.055)	(0.032)	(0.055)	(0.025)	(0.040)	(0.023)	(0.041)	(0.018)
Out of tested	0.228	0.180	0.139	0.181	0.279	0.163	0.111	0.176
	(0.049)	(0.029)	(0.052)	(0.022)	(0.035)	(0.022)	(0.038)	(0.017)
Out of district	0.197	0.142	0.105	0.145	0.167	0.160	0.133	0.156
	(0.050)	(0.030)	(0.053)	(0.023)	(0.037)	(0.022)	(0.038)	(0.017)

Table A.7—Continued

	Early Reform				Recent Reform			
Change	**1**	**2**	**3**	**Total**	**1**	**2**	**3**	**Total**
Total proportion	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Total number of schools	325	974	324	1,623	603	1,805	602	3,010
Reading								
Increase LIM, same school	0.324	0.377	0.358	0.363	0.271	0.332	0.343	0.322
	(0.040)	(0.022)	(0.039)	(0.018)	(0.031)	(0.017)	(0.029)	(0.013)
Decrease LIM, same school	0.222	0.287	0.339	0.284	0.235	0.303	0.323	0.293
	(0.043)	(0.024)	(0.040)	(0.019)	(0.032)	(0.017)	(0.030)	(0.014)
Increase LIM, new school	0.029	0.014	0.007	0.015	0.022	0.023	0.007	0.019
	(0.048)	(0.028)	(0.049)	(0.022)	(0.036)	(0.021)	(0.036)	(0.016)
Decrease LIM, new school	0.041	0.035	0.022	0.034	0.028	0.031	0.012	0.026
	(0.048)	(0.028)	(0.049)	(0.022)	(0.036)	(0.021)	(0.036)	(0.016)
Out of tested	0.227	0.167	0.123	0.170	0.246	0.145	0.135	0.163
	(0.043)	(0.026)	(0.046)	(0.020)	(0.031)	(0.019)	(0.034)	(0.015)
Out of district	0.157	0.120	0.150	0.133	0.198	0.167	0.180	0.176
	(0.045)	(0.027)	(0.045)	(0.020)	(0.032)	(0.019)	(0.033)	(0.015)

Table A.7—Continued

Change	Early Reform 1	Early Reform 2	Early Reform 3	Early Reform Total	Recent Reform 1	Recent Reform 2	Recent Reform 3	Recent Reform Total
Total proportion	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Total number of schools	414	1,241	413	2,068	763	2,287	761	3,811

NOTE: The numbers in parentheses are standard errors.

Table A.8
Memphis City Schools Changes in Low-Income Minority Composition, by Performance Level: Site Value-Added Measure Tercile

	Recent Reform			
Change	**1**	**2**	**3**	**Total**
Mathematics				
Increase LIM, same school	0.174	0.237	0.000	0.216
	(0.042)	(0.028)	(0.000)	(0.023)
Decrease LIM, same school	0.183	0.240	0.000	0.221
	(0.041)	(0.028)	(0.000)	(0.023)
Increase LIM, new school	0.088	0.112	0.000	0.104
	(0.044)	(0.031)	(0.000)	(0.025)
Decrease LIM, new school	0.118	0.147	0.000	0.137
	(0.043)	(0.030)	(0.000)	(0.025)
Out of tested	0.227	0.119	0.000	0.155
	(0.040)	(0.030)	(0.000)	(0.024)
Out of district	0.210	0.145	0.000	0.166
	(0.041)	(0.030)	(0.000)	(0.024)
Total proportion	1.000	1.000	0.000	1.000
Total number of schools	476	954	0	1,430
Reading				
Increase LIM, same school	0.158	0.229	0.000	0.213
	(0.047)	(0.024)	(0.000)	(0.021)
Decrease LIM, same school	0.166	0.209	0.000	0.200
	(0.046)	(0.024)	(0.000)	(0.021)
Increase LIM, new school	0.091	0.098	0.000	0.097
	(0.049)	(0.026)	(0.000)	(0.023)

Table A.8—Continued

	Recent Reform			
Change	**1**	**2**	**3**	**Total**
Decrease LIM, new school	0.140	0.124	0.000	0.128
	(0.047)	(0.025)	(0.000)	(0.022)
Out of tested	0.249	0.164	0.000	0.182
	(0.044)	(0.025)	(0.000)	(0.021)
Out of district	0.197	0.176	0.000	0.181
	(0.046)	(0.024)	(0.000)	(0.022)
Total proportion	1.000	1.000	0.000	1.000
Total number of schools	386	1,386	0	1,772

NOTE: The numbers in parentheses are standard errors.

Table A.9
Memphis City Schools Changes in Low-Income Minority Composition, by Performance Level: Site Composite Tercile

	Recent Reform			
Change	**1**	**2**	**3**	**Total**
Mathematics				
Increase LIM, same school	0.140	0.223	0.272	0.216
	(0.055)	(0.030)	(0.051)	(0.023)
Decrease LIM, same school	0.196	0.215	0.265	0.221
	(0.053)	(0.030)	(0.051)	(0.023)
Increase LIM, new school	0.049	0.107	0.152	0.104
	(0.058)	(0.032)	(0.055)	(0.025)
Decrease LIM, new school	0.095	0.145	0.155	0.137
	(0.056)	(0.031)	(0.055)	(0.025)
Out of tested	0.277	0.157	0.028	0.155
	(0.050)	(0.031)	(0.059)	(0.024)

Table A.9—Continued

Change	Recent Reform 1	2	3	Total
Out of district	0.242	0.154	0.127	0.166
	(0.052)	(0.031)	(0.056)	(0.024)
Total proportion	1.000	1.000	1.000	1.000
Total number of schools	285	862	283	1,430
Reading				
Increase LIM, same school	0.132	0.217	0.285	0.213
	(0.049)	(0.027)	(0.045)	(0.021)
Decrease LIM, same school	0.169	0.199	0.234	0.200
	(0.048)	(0.027)	(0.047)	(0.021)
Increase LIM, new school	0.087	0.100	0.096	0.097
	(0.051)	(0.029)	(0.051)	(0.023)
Decrease LIM, new school	0.132	0.130	0.116	0.128
	(0.049)	(0.029)	(0.050)	(0.022)
Out of tested	0.275	0.186	0.076	0.182
	(0.045)	(0.028)	(0.051)	(0.021)
Out of district	0.205	0.169	0.192	0.181
	(0.047)	(0.028)	(0.048)	(0.022)
Total proportion	1.000	1.000	1.000	1.000
Total number of schools	356	1,062	354	1,772

NOTE: The numbers in parentheses are standard errors.

Table A.10
Pittsburgh Public Schools Changes in Low-Income Minority Composition, by Performance Level: Site Value-Added Measure Tercile

	Recent Reform			
Change	**1**	**2**	**3**	**Total**
Mathematics				
Increase LIM, same school	0.326	0.388	0.286	0.355
	(0.121)	(0.069)	(0.130)	(0.055)
Decrease LIM, same school	0.370	0.349	0.524	0.387
	(0.117)	(0.071)	(0.106)	(0.053)
Increase LIM, new school	0.043	0.031	0.024	0.032
	(0.144)	(0.087)	(0.152)	(0.067)
Decrease LIM, new school	0.000	0.054	0.024	0.037
	(0.000)	(0.086)	(0.152)	(0.067)
Out of tested	0.109	0.078	0.024	0.074
	(0.139)	(0.085)	(0.152)	(0.065)
Out of district	0.152	0.101	0.119	0.115
	(0.136)	(0.083)	(0.145)	(0.064)
Total proportion	1.000	1.000	1.000	1.000
Total number of schools	46	129	42	217
Reading				
Increase LIM, same school	0.321	0.369	0.367	0.359
	(0.113)	(0.063)	(0.114)	(0.050)
Decrease LIM, same school	0.264	0.427	0.367	0.382
	(0.118)	(0.060)	(0.114)	(0.049)
Increase LIM, new school	0.057	0.051	0.000	0.042
	(0.133)	(0.078)	(0.000)	(0.061)

Table A.10—Continued

	Recent Reform			
Change	**1**	**2**	**3**	**Total**
Decrease LIM, new school	0.019	0.032	0.000	0.023
	(0.136)	(0.079)	(0.000)	(0.061)
Out of tested	0.170	0.051	0.163	0.097
	(0.125)	(0.078)	(0.131)	(0.059)
Out of district	0.170	0.070	0.102	0.097
	(0.125)	(0.077)	(0.135)	(0.059)
Total proportion	1.000	1.000	1.000	1.000
Total number of schools	53	157	49	259

NOTE: The numbers in parentheses are standard errors.

Table A.11
Pittsburgh Public Schools Changes in Low-Income Minority Composition, by Performance Level: Site Composite Tercile

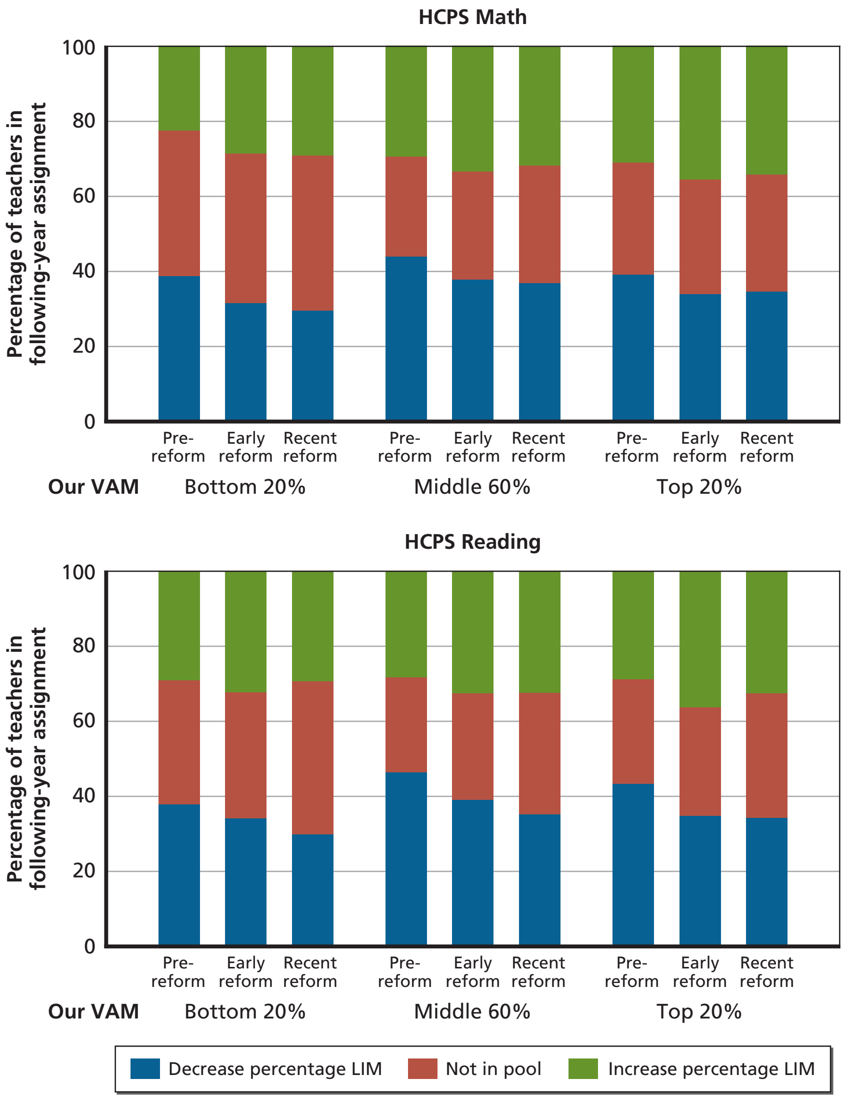
	Recent Reform			
Change	**1**	**2**	**3**	**Total**
Mathematics				
Increase LIM, same school	0.175	0.379	0.321	0.324
	(0.114)	(0.060)	(0.110)	(0.048)
Decrease LIM, same school	0.254	0.356	0.464	0.355
	(0.109)	(0.061)	(0.098)	(0.047)
Increase LIM, new school	0.048	0.029	0.018	0.031
	(0.123)	(0.075)	(0.132)	(0.058)
Decrease LIM, new school	0.063	0.029	0.000	0.031
	(0.122)	(0.075)	(0.000)	(0.058)
Out of tested	0.159	0.086	0.089	0.102
	(0.116)	(0.072)	(0.128)	(0.055)

Table A.11—Continued

	Recent Reform			
Change	**1**	**2**	**3**	**Total**
Out of district	0.302	0.121	0.107	0.157
	(0.105)	(0.071)	(0.126)	(0.054)
Total proportion	1.000	1.000	1.000	1.000
Total number of schools	63	174	56	293
Reading				
Increase LIM, same school	0.236	0.351	0.348	0.327
	(0.103)	(0.055)	(0.097)	(0.044)
Decrease LIM, same school	0.250	0.341	0.420	0.338
	(0.102)	(0.056)	(0.092)	(0.043)
Increase LIM, new school	0.056	0.038	0.000	0.034
	(0.115)	(0.068)	(0.000)	(0.052)
Decrease LIM, new school	0.042	0.024	0.014	0.026
	(0.115)	(0.068)	(0.120)	(0.053)
Out of tested	0.222	0.142	0.116	0.153
	(0.104)	(0.064)	(0.113)	(0.049)
Out of district	0.194	0.104	0.101	0.122
	(0.106)	(0.065)	(0.114)	(0.050)
Total proportion	1.000	1.000	1.000	1.000
Total number of schools	72	211	69	352

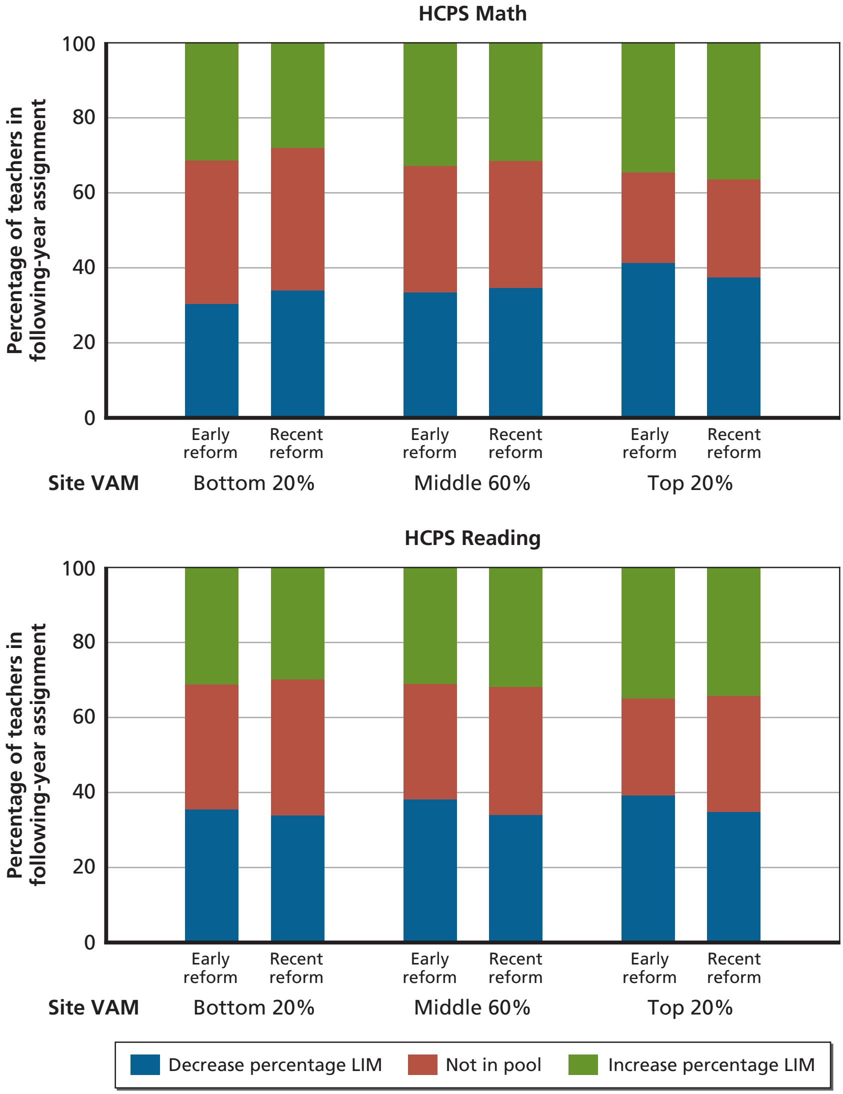
NOTE: The numbers in parentheses are standard errors.

Figure A.4
Changes in Low-Income Minority Composition, by Teacher Performance Level, Hillsborough County Public Schools, Our Value-Added Measure



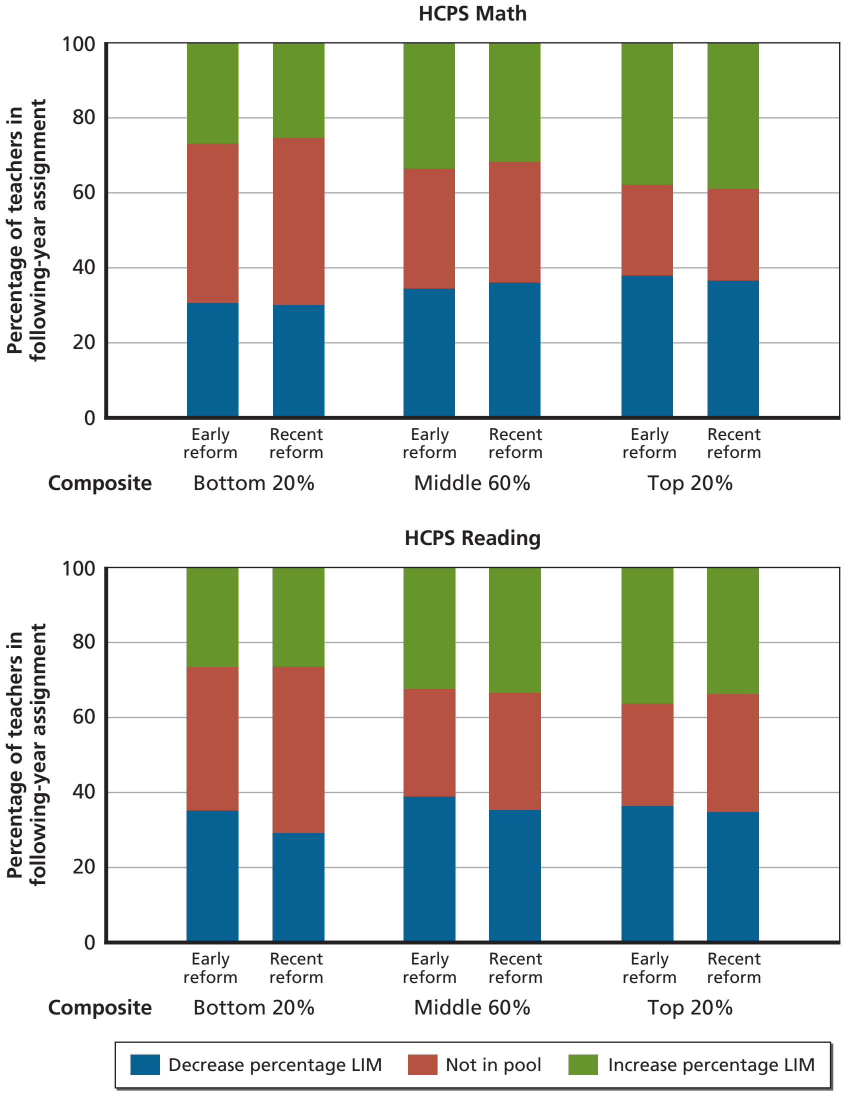
RAND *RR1295/4-A.4*

Figure A.5
Changes in Low-Income Minority Composition, by Teacher Performance Level, Hillsborough County Public Schools, Site Value-Added Measure



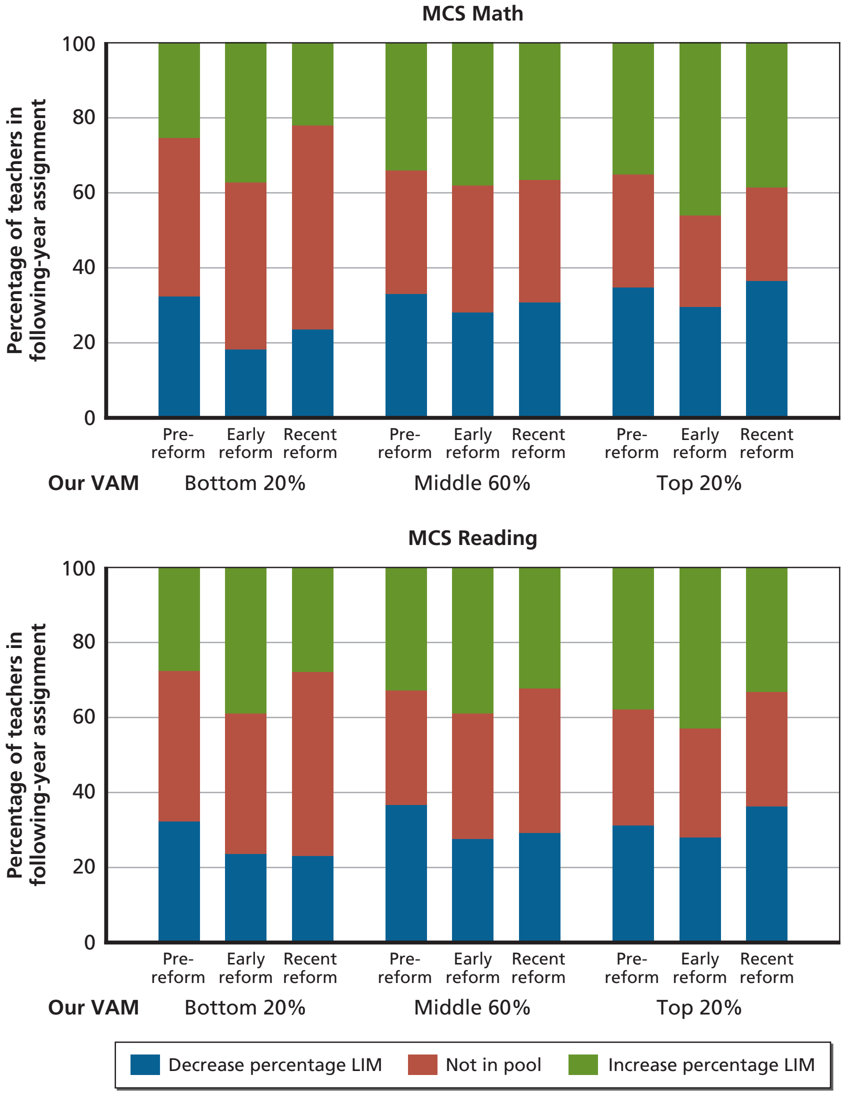
RAND *RR1295/4-A.5*

Figure A.6
Changes in Low-Income Minority Composition, by Teacher Performance Level, Hillsborough County Public Schools, Composite



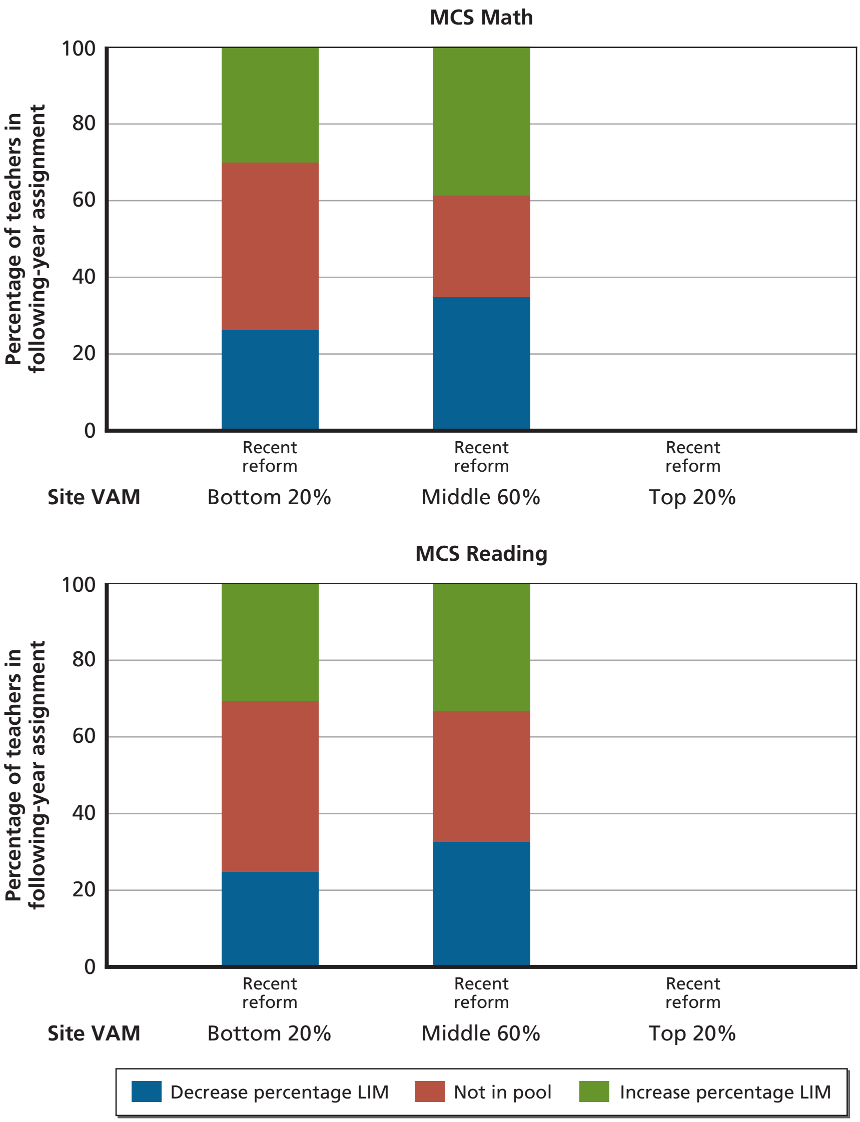
RAND *RR1295/4-A.6*

Figure A.7
Changes in Low-Income Minority Composition, by Teacher Performance Level, Memphis City Schools, Our Value-Added Measure



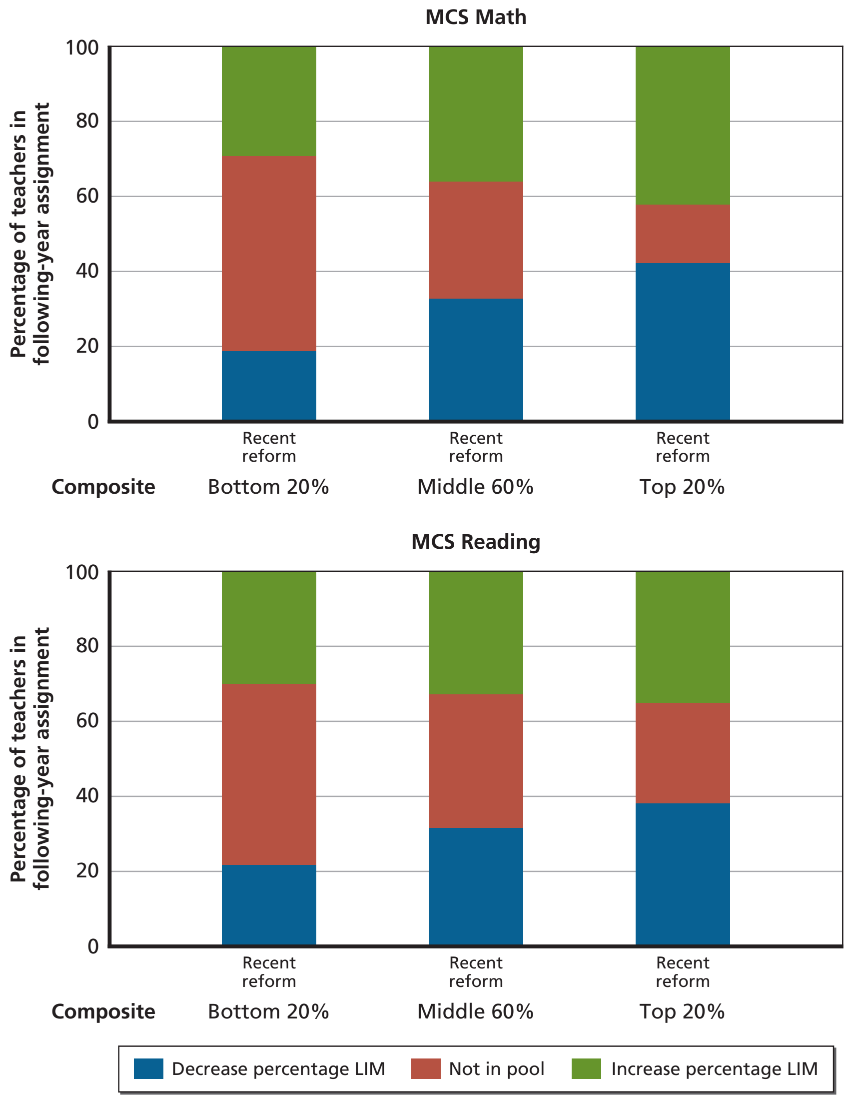
RAND *RR1295/4-A.7*

Figure A.8
Changes in Low-Income Minority Composition, by Teacher Performance Level, Memphis City Schools, Site Value-Added Measure



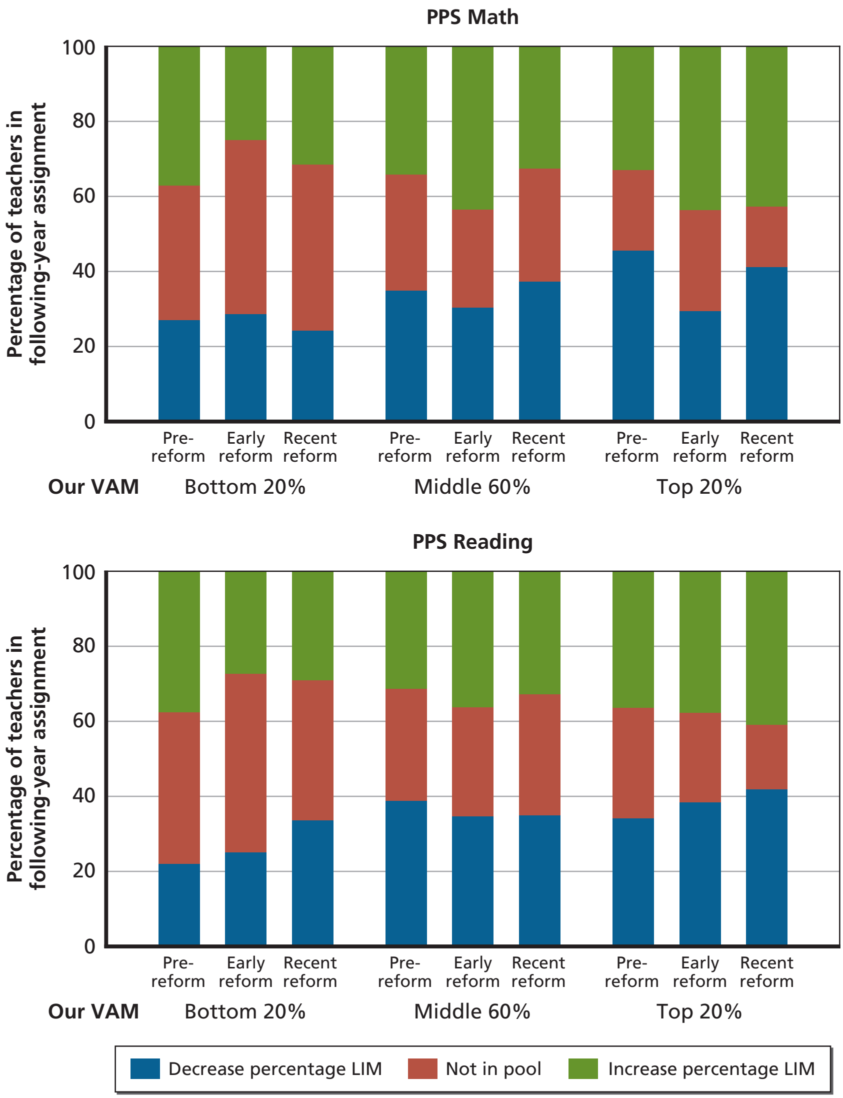
RAND *RR1295/4-A.8*

Figure A.9
Changes in Low-Income Minority Composition, by Teacher Performance Level, Memphis City Schools, Composite



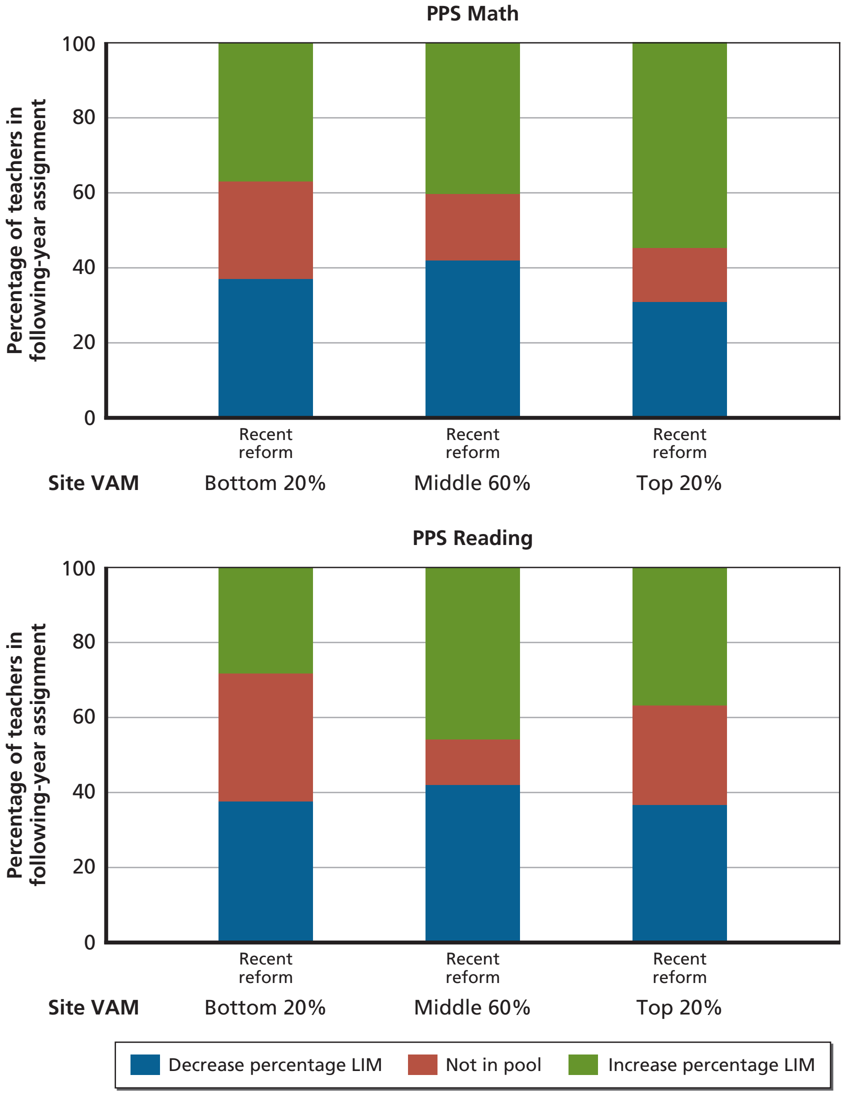
RAND *RR1295/4-A.9*

Figure A.10
Changes in Low-Income Minority Composition, by Teacher Performance Level, Pittsburgh Public Schools, Our Value-Added Measure



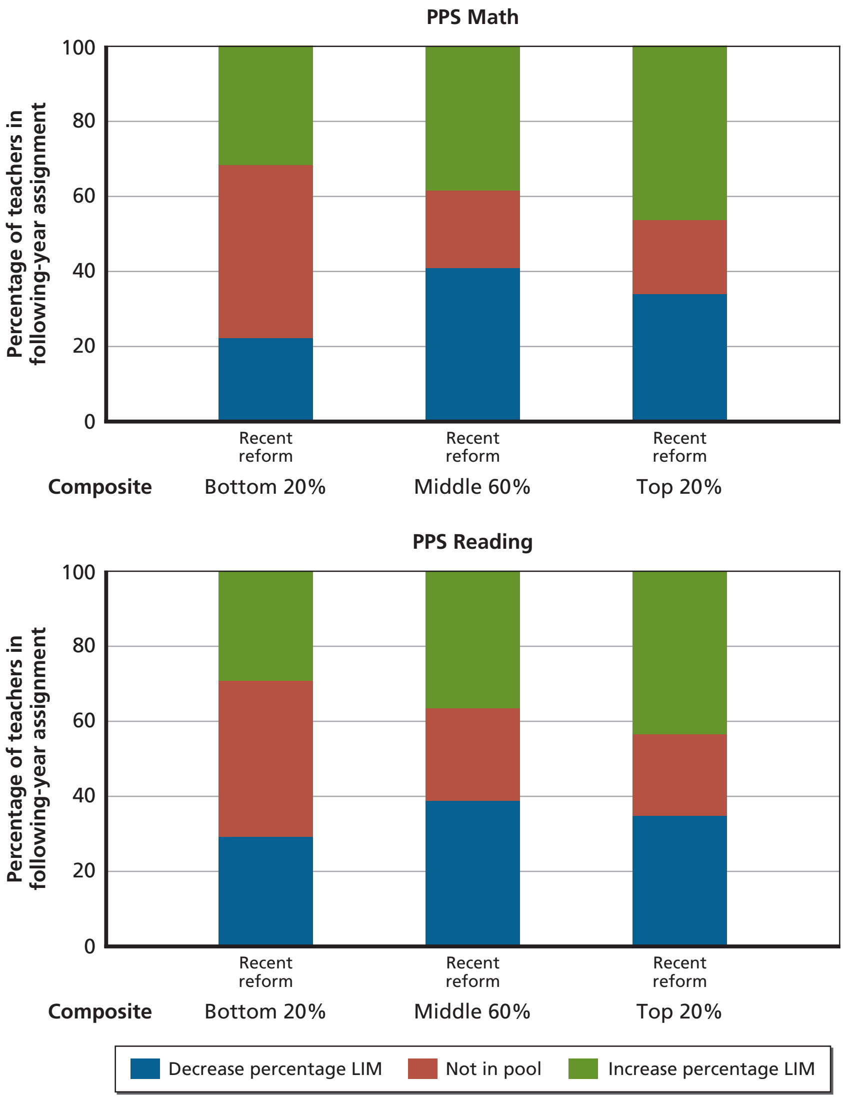
RAND *RR1295/4-A.10*

Figure A.11
Changes in Low-Income Minority Composition, by Teacher Performance Level, Pittsburgh Public Schools, Site Value-Added Measure



RAND *RR1295/4-A.11*

Figure A.12
Changes in Low-Income Minority Composition, by Teacher Performance Level, Pittsburgh Public Schools, Composite



RAND *RR1295/4-A.12*

Figures A.13 through A.30 show our estimates of the three components. We provide the estimates for the prereform, early reform, and recent periods for each subject in each site.

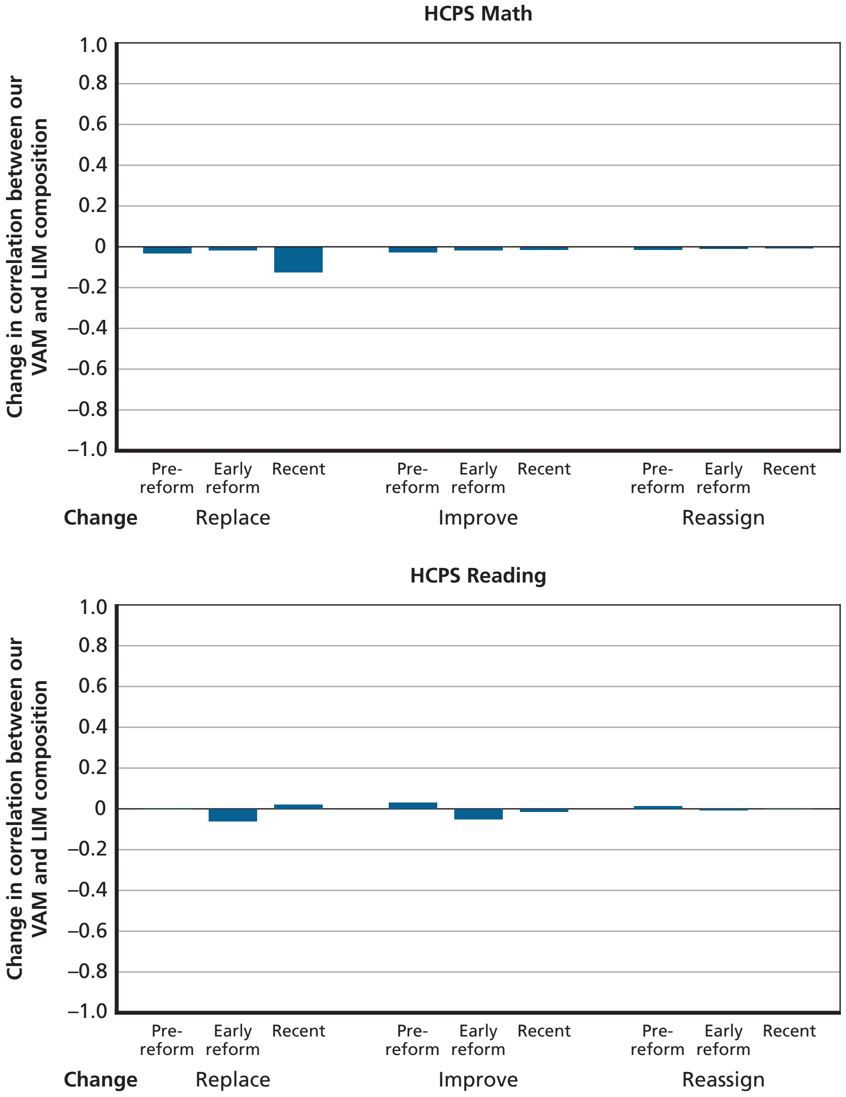
Analysis of Mechanisms Used to Change Access

Although the intended mechanisms discussed in Chapter Five shed light on districts' motives, that analysis does not decompose the actual end sorting of each year through the mechanisms, and it holds changes in teacher effectiveness (as measured by VAM) constant. However, teacher VAM might be changing differentially for teachers assigned to higher- or lower-fraction LIM classrooms. A teacher being assigned a higher-LIM classroom might view this as a signal of trust, or feel a greater obligation to the students, and so might experience an increase in VAM, which would lead to a higher sorting coefficient. Or, a teacher might view the higher-LIM classroom as an easier assignment or a challenge, either of which could lead to complacency and lower VAM and a lower sorting coefficient. Even if assignments were not changed at all (all teachers had exactly the same fraction of LIM students), it might be that VAM for higher-fraction LIM classroom teachers increases (or decreases) more for any number of reasons, leading to more-progressive (or more-regressive) sorting patterns. To answer that question, we can decompose the change in overall sorting into four components as follows:

$$\begin{aligned}\hat{\beta}_t - \hat{\beta}_{t-1} &= \frac{p_{new} + p_{exit}}{2}\left(\hat{\beta}_{V1L1|new} - \hat{\beta}_{V0L0|exit}\right) \\ &\quad + \frac{p_{stay} + p_{exp}}{2}\left(\hat{\beta}_{V1L0|stay} - \hat{\beta}_{V0L0|stay} + \hat{\beta}_{V0L1|stay} - \hat{\beta}_{V0L0|stay}\right) + R \\ &= (1-p)\Delta_{replace} + p\left(\Delta_{improve} + \Delta_{reassign}\right) + R.\end{aligned}$$

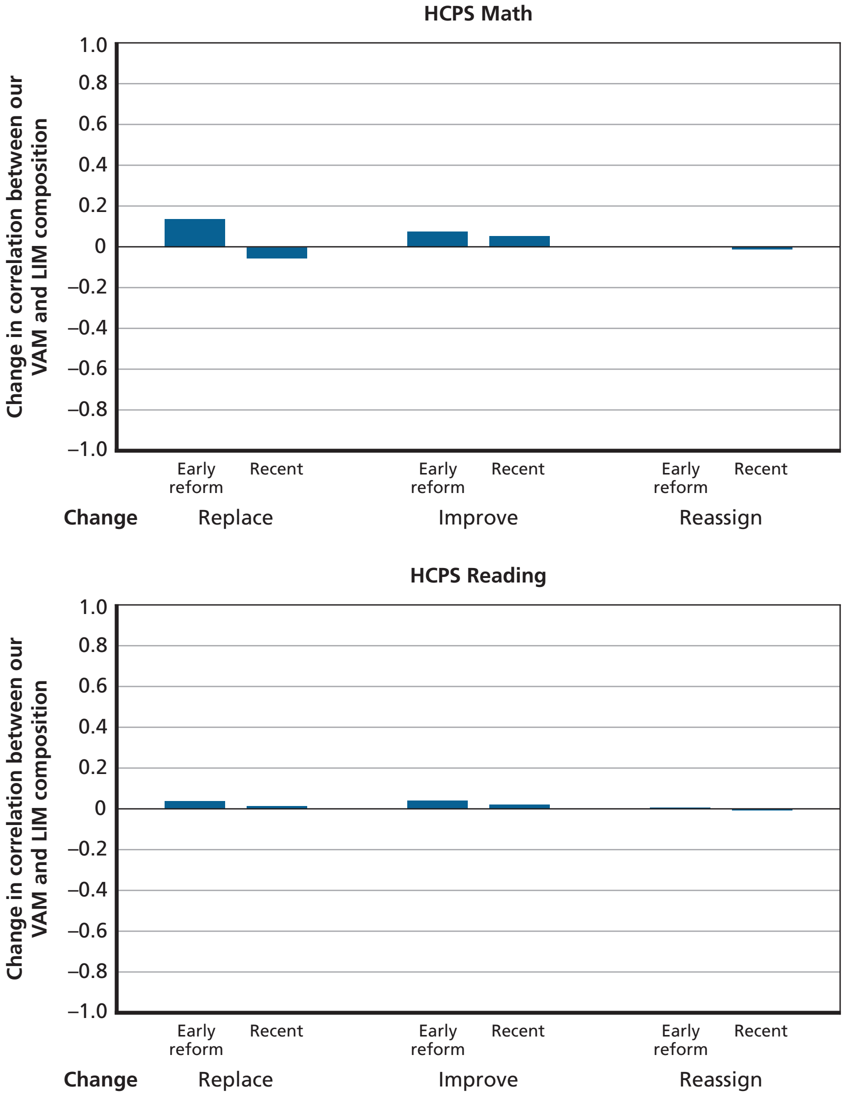
In other words, the change in the overall sorting coefficient can be decomposed into three main elements. $\Delta_{replace} = \beta_{V1L1|new} - \beta_{V0L0|exit}$

Figure A.13
Decomposition of Change in Overall Sorting, Hillsborough County Public Schools, Our Value-Added Measure



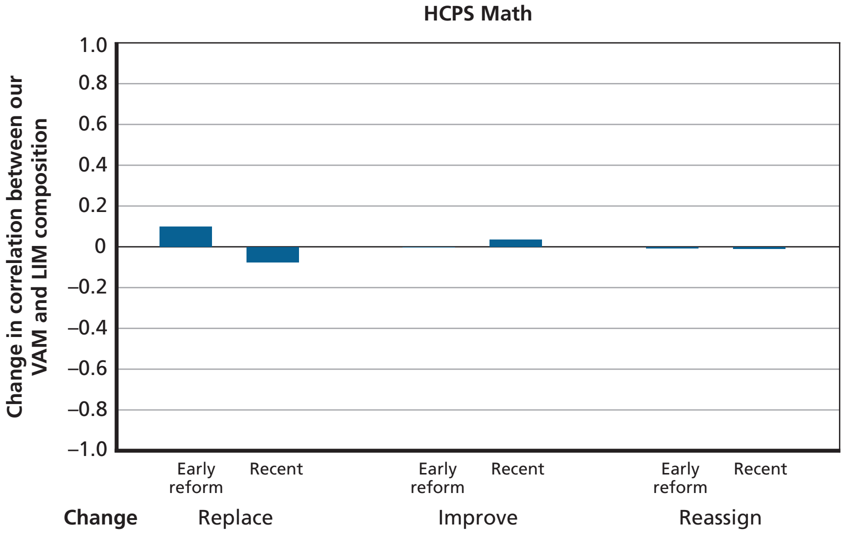
RAND *RR1295/4-A.13*

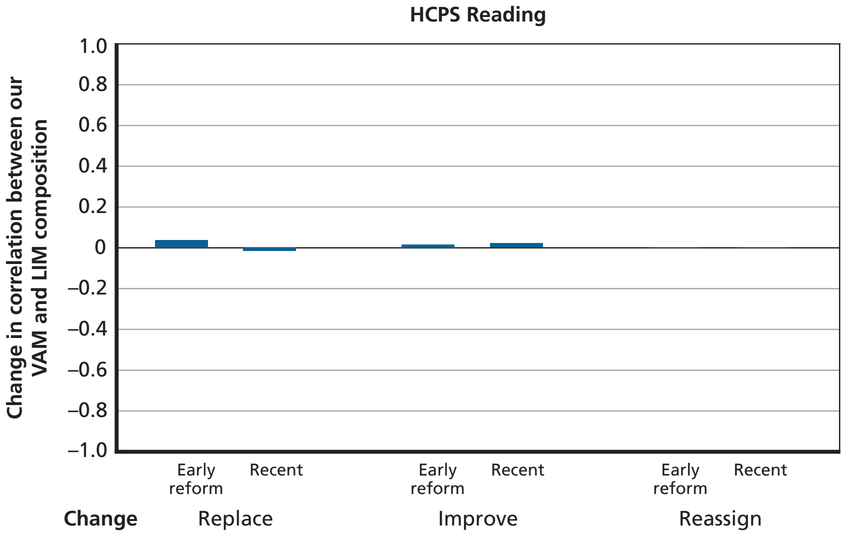
Figure A.14
Decomposition of Change in Overall Sorting, Hillsborough County Public Schools, Site Value-Added Measure



RAND *RR1295/4-A.14*

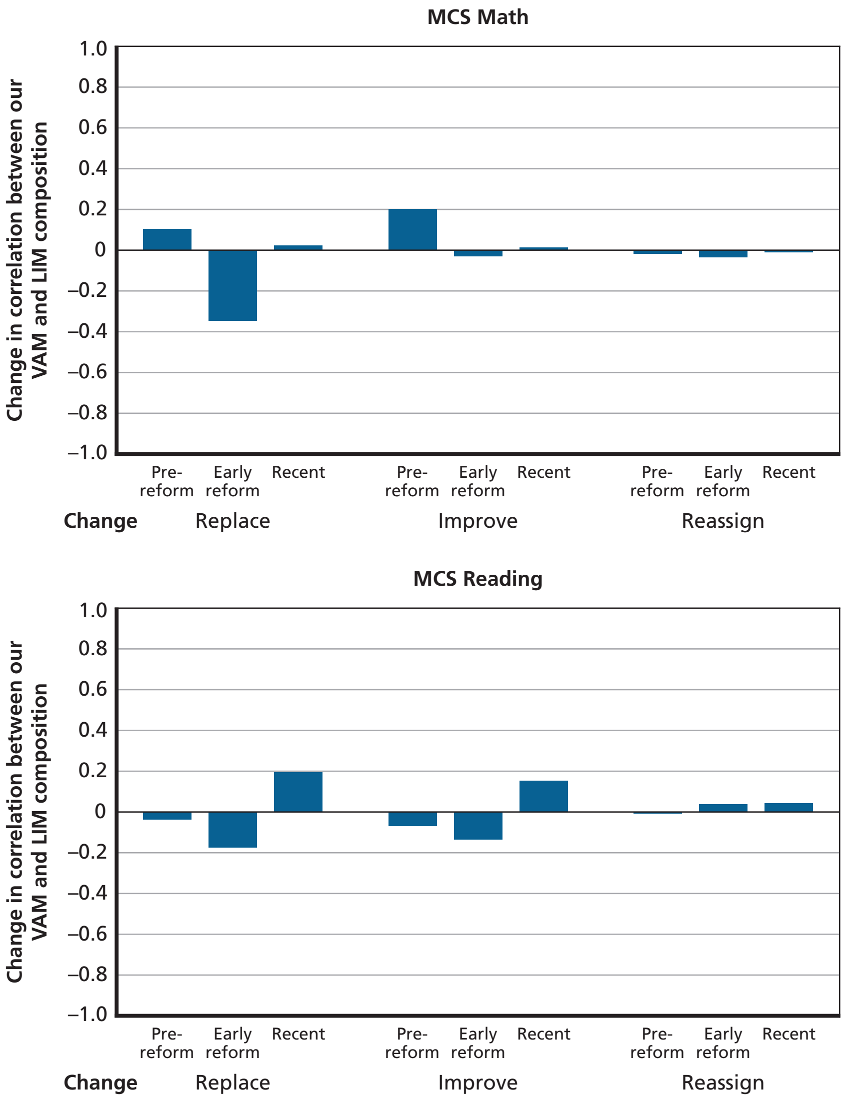
Figure A.15
Decomposition of Change in Overall Sorting, Hillsborough County Public Schools, Composite





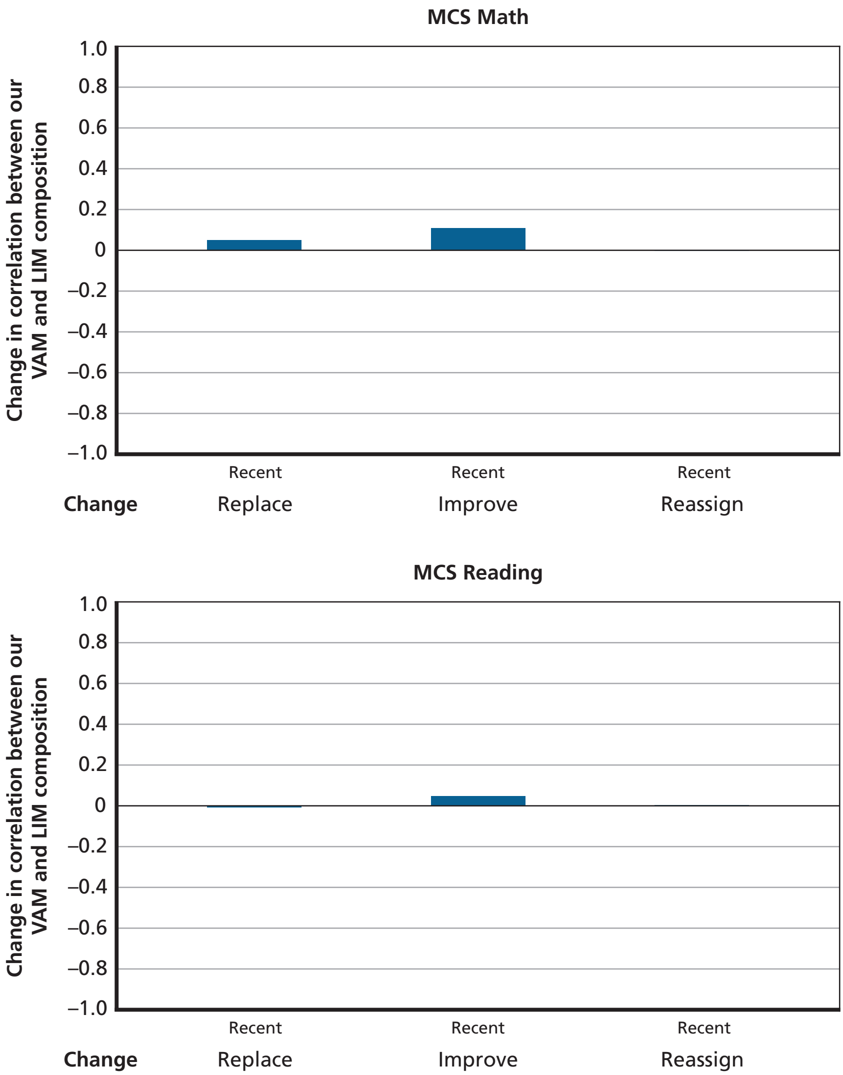
RAND *RR1295/4-A.15*

Figure A.16
Decomposition of Change in Overall Sorting, Memphis City Schools, Our Value-Added Measure



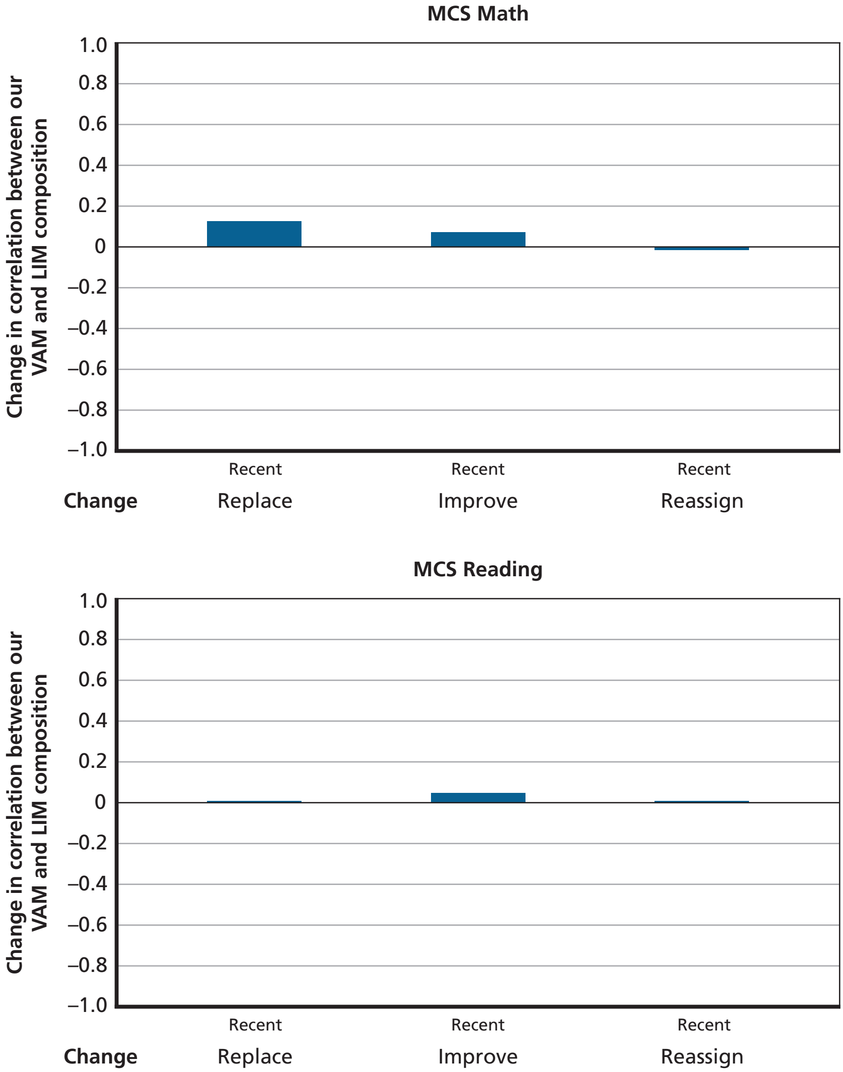
RAND *RR1295/4-A.16*

Figure A.17
Decomposition of Change in Overall Sorting, Memphis City Schools, Site Value-Added Measure



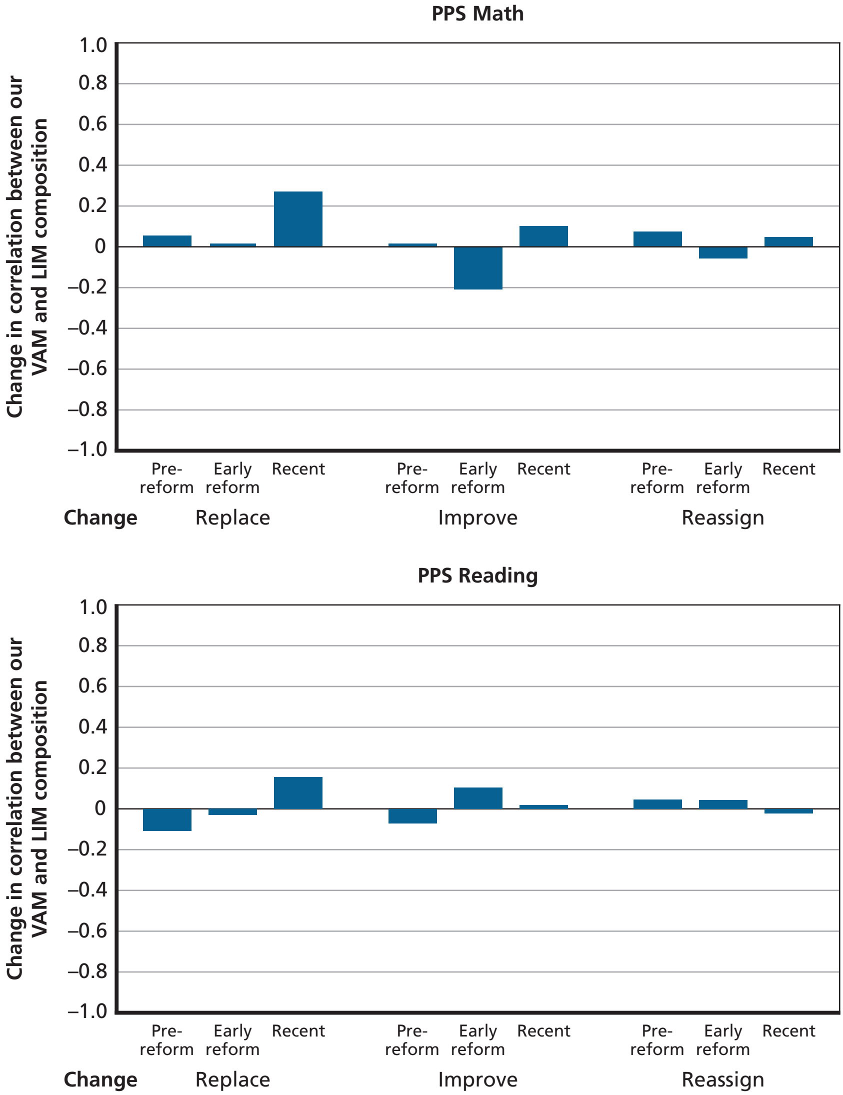
RAND *RR1295/4-A.17*

Figure A.18
Decomposition of Change in Overall Sorting, Memphis City Schools, Composite



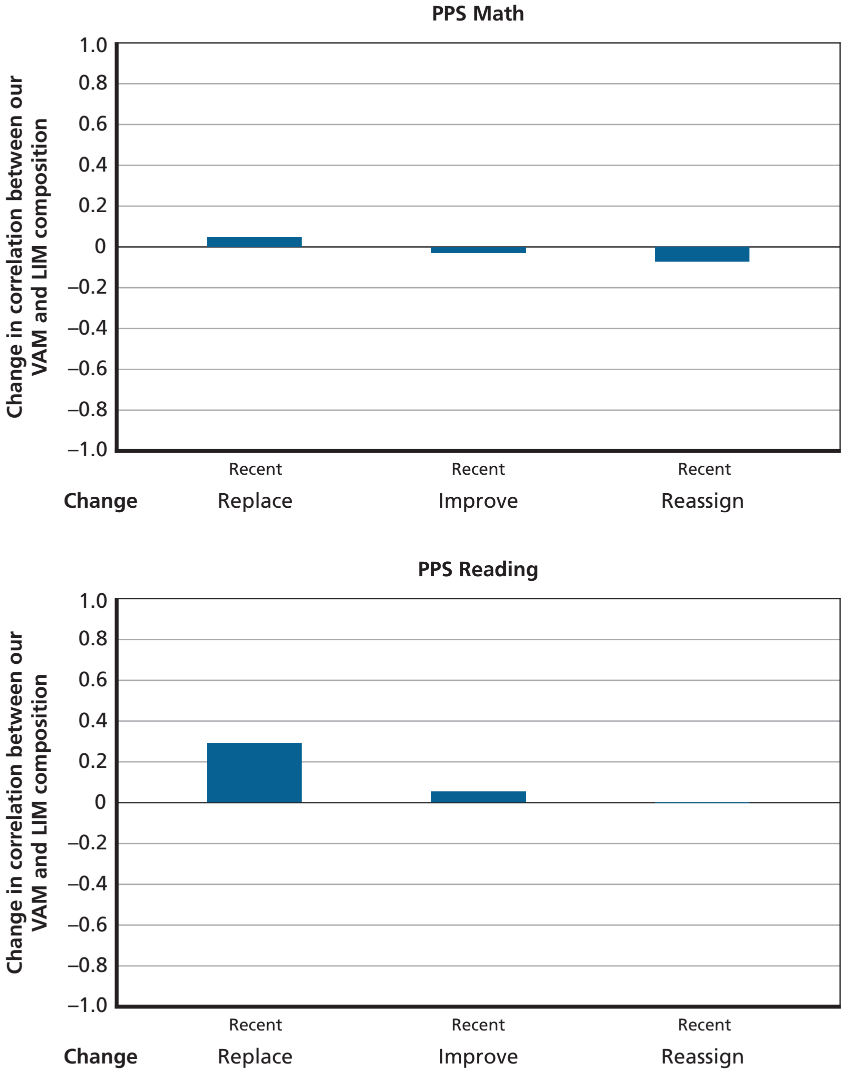
RAND *RR1295/4-A.18*

Figure A.19
Decomposition of Change in Overall Sorting, Pittsburgh Public Schools, Our Value-Added Measure



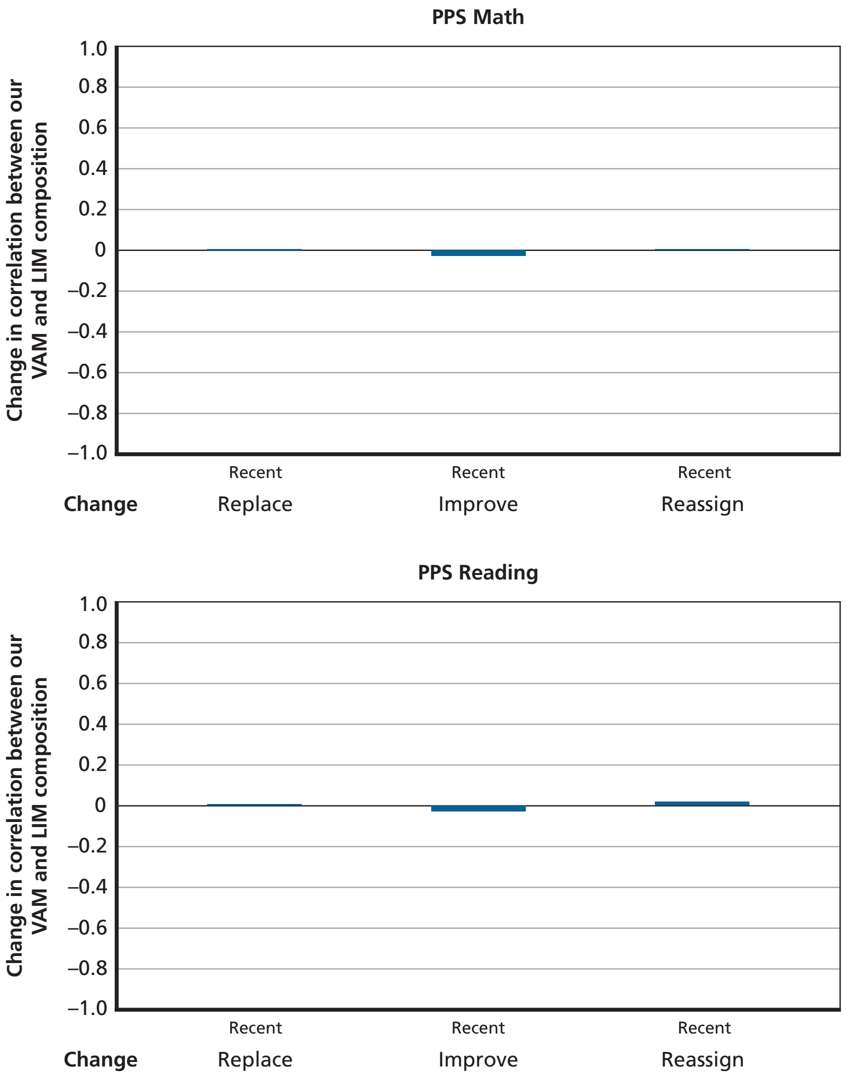
RAND *RR1295/4-A.19*

Figure A.20
Decomposition of Change in Overall Sorting, Pittsburgh Public Schools, Site Value-Added Measure



RAND *RR1295/4-A.20*

Figure A.21
Decomposition of Change in Overall Sorting, Pittsburgh Public Schools, Composite



RAND *RR1295/4-A.21*

measures the change in LIM access to effective teachers caused by the exit and entry of teachers across the two years and what sorting those teachers had. This is weighted by

$$\left(1-p\right)=\frac{\left(p_{new}+p_{exit}\right)}{2},$$

the proportion of teachers that transition on average in those two years. The second element, $\Delta_{improve}=\beta_{V1L0|stay}-\beta_{V0L0|stay}$, measures the change in the sorting coefficients caused by changes in VAM across the two years. It does this by measuring what the change in the sorting coefficients would have been if the fraction of LIM each teacher has does not change across the two years but each teacher's effectiveness is allowed to change as is observed. This is weighted by

$$p=\frac{\left(p_{stay}+p_{exp}\right)}{2},$$

the average fraction of teachers who stay and the fraction who are then returners the next year (in the case that the total number of teachers is the same across years, these two measures are identical). The third element is $\Delta_{reassign}=\hat{\beta}_{V0L1|stay}-\hat{\beta}_{V0L0|stay}$, the portion of the sorting coefficient changed by changes in assignments of teachers and their fractions of LIM students. It does this by measuring how the sorting coefficient would have changed if those teachers' VAMs had stayed the same but their fractions of LIM students changed as was observed in the data. It is weighted by the same *p*. The fourth element, *R*, is the residual difference between the actual difference in sorting coefficients and our decomposition. This is a complicated function of regression coefficients on various samples that largely cancels out. Another difference is that this decomposition does not use the WLS weights that the actual analysis uses. However, the correlation coefficient between the actual (WLS) difference in the sorting coefficients and our decomposition (leaving *R* out) is above 0.96, and a regression of the former on the latter yields an ordinary-least-squares coefficient of 0.903 (*t*-statistic of

19.82) with an intercept of –0.006 (*t*-statistic of –0.36). This demonstrates how close our decomposition is to a complete decomposition (leaving a negligible residual) even while not accounting for the WLS (note, we are not performing any inference on these but using the statistics as guidance), and we use this version, which presents interpretable elements that can be examined.

Tables A.12 through A.14 provide numerical estimates of the effect of the decomposition actions and indicate statistical significance.

Aspire Public Schools

Sorting of Value Added in Mathematics and Reading over Time

Aspire Public Schools is the only CMO in the sample for which we have enough teachers and schools in grades 4 through 8 to estimate teacher sorting patterns over time. Still, the site has far fewer schools and teachers than the large urban districts do, as shown by the sample sizes in Table A.15, so the sorting estimates are substantially noisier than in the sites described above. Although we present all of the results, we urge caution in interpreting these results given the small number of schools and students used in these calculations. We present them in order to show overall patterns and not for direct comparison with the more-precise estimates of other sites.

Bearing in mind that sorting parameters are more noisily estimated in Aspire than in the other sites (we report between-school sorting parameters only when there are at least 30 schools), we observe the patterns described in the rest of this section.

Mathematics

- In 2007–2008, within-school sorting was sharply negative but not statistically significant. In contrast, between-school sorting was strongly positive and significant. The combination of these estimates results in a negative and statistically significant parameter for sorting of effective teachers to LIM students. Overall sorting became positive in the remaining preintervention years because of positive within-school sorting.

Table A.12
Hillsborough County Public Schools: Decomposition of Change in Overall Sorting

Variable	Period	p	$\Delta_{replace}$	$\Delta_{improve}$	$\Delta_{reassign}$	$\hat{\beta}_t - \hat{\beta}_{t-1}$
Mathematics						
Our VAM	1	0.711 (0.009)	−0.030 (0.031)	−0.025 (0.017)	−0.013 (0.018)	−0.008 (0.024)
	2	0.713 (0.009)	−0.014 (0.035)	−0.014 (0.021)	−0.008 (0.019)	−0.009 (0.028)
	3	0.678 (0.009)	−0.122*** (0.037)	−0.011 (0.024)	−0.005 (0.023)	−0.054 (0.031)
Site VAM	2	0.703 (0.013)	0.137*** (0.041)	0.076** (0.030)	0.003 (0.028)	0.080** (0.028)
	3	0.753 (0.009)	−0.053 (0.037)	0.056** (0.026)	−0.011 (0.022)	0.003 (0.029)
Composite	2	0.703 (0.013)	0.099** (0.037)	−0.001 (0.023)	−0.005 (0.021)	0.018 (0.027)
	3	0.753 (0.009)	−0.074* (0.036)	0.035 (0.020)	−0.008 (0.017)	−0.025 (0.028)
Reading						
Our VAM	1	0.721 (0.008)	−0.001 (0.026)	0.031* (0.015)	0.014 (0.014)	0.017 (0.021)
	2	0.718 (0.008)	−0.060** (0.023)	−0.051*** (0.014)	−0.007 (0.015)	−0.048* (0.019)
	3	0.695 (0.008)	0.022 (0.021)	−0.013 (0.014)	0.000 (0.014)	0.000 (0.018)
Site VAM	2	0.719 (0.011)	0.039 (0.025)	0.041** (0.016)	0.005 (0.016)	0.039** (0.017)
	3	0.760 (0.008)	0.013 (0.024)	0.021 (0.016)	−0.007 (0.013)	0.023 (0.018)
Composite	2	0.718 (0.011)	0.039* (0.022)	0.017 (0.013)	0.002 (0.013)	0.015 (0.016)
	3	0.760 (0.008)	−0.016 (0.022)	0.022 (0.011)	−0.002 (0.009)	−0.012 (0.017)

NOTE: *** = $p < 0.01$. ** = $p < 0.05$. * = $p < 0.1$. We used the Benjamini–Hochberg method to adjust significance levels for multiple hypotheses. The numbers in parentheses are standard errors.

Table A.13
Memphis City Schools: Decomposition of Change in Overall Sorting

Variable	Period	p	$\Delta_{replace}$	$\Delta_{improve}$	$\Delta_{reassign}$	$\hat{\beta}_t - \hat{\beta}_{t-1}$
Mathematics						
VAM	1	0.827 (0.010)	0.103 (0.125)	0.200*** (0.053)	−0.019 (0.049)	0.156** (0.067)
	2	0.800 (0.015)	−0.347 (0.178)	−0.030 (0.085)	−0.035 (0.088)	−0.107 (0.081)
	3	0.788 (0.011)	0.022 (0.142)	0.012 (0.061)	−0.012 (0.062)	0.082 (0.086)
Site VAM	3	0.830 (0.011)	0.049 (0.177)	0.108 (0.064)	−0.002 (0.061)	0.055 (0.085)
Composite	3	0.830 (0.011)	0.126 (0.194)	0.072 (0.067)	−0.015 (0.065)	0.060 (0.083)
Reading						
VAM	1	0.825 (0.009)	−0.037 (0.081)	−0.070* (0.037)	−0.008 (0.037)	−0.077 (0.052)
	2	0.787 (0.014)	−0.176 (0.101)	−0.135** (0.053)	0.035 (0.053)	−0.119** (0.049)
	3	0.762 (0.011)	0.193* (0.080)	0.152*** (0.040)	0.040 (0.039)	0.139** (0.047)
Site VAM	3	0.803 (0.010)	−0.007 (0.096)	0.047 (0.039)	0.004 (0.037)	0.051 (0.053)
Composite	3	0.804 (0.010)	0.007 (0.105)	0.047 (0.041)	0.007 (0.039)	0.039 (0.052)

NOTE: *** = $p < 0.01$. ** = $p < 0.05$. * = $p < 0.1$. We used the Benjamini–Hochberg method to adjust significance levels for multiple hypotheses. The numbers in parentheses are standard errors.

- Since the intervention commenced, sorting patterns have been negative but closer to 0 and have been steadily improving.
- The number of schools is too small to provide reliable between-school estimates of sorting except for in 2013. However, a comparison of the within-school and overall sorting parameters shows that, in recent years, within-school sorting has been more pro-

Table A.14
Pittsburgh Public Schools: Decomposition of Change in Overall Sorting

Variable	Period	p	$\Delta_{replace}$	$\Delta_{improve}$	$\Delta_{reassign}$	$\hat{\beta}_t - \hat{\beta}_{t-1}$
Mathematics						
VAM	1	0.772 (0.022)	0.056 (0.132)	0.017 (0.072)	0.075 (0.069)	0.004 (0.109)
	2	0.795 (0.023)	0.017 (0.165)	−0.207* (0.085)	−0.056 (0.088)	−0.183 (0.112)
	3	0.808 (0.024)	0.273 (0.155)	0.103 (0.073)	0.048 (0.077)	0.134 (0.099)
Site VAM	3	0.918 (0.020)	0.047 (0.322)	−0.029 (0.087)	−0.072 (0.093)	0.001 (0.104)
Composite	3	0.854 (0.022)	0.006 (0.178)	−0.028 (0.077)	0.005 (0.076)	0.012 (0.092)
Reading						
VAM	1	0.774 (0.020)	−0.106 (0.120)	−0.069 (0.056)	0.045 (0.057)	−0.057 (0.091)
	2	0.832 (0.019)	−0.030 (0.152)	0.105 (0.068)	0.043 (0.069)	0.003 (0.099)
	3	0.855 (0.019)	0.156 (0.156)	0.017 (0.068)	−0.021 (0.066)	0.111 (0.086)
Site VAM	3	0.942 (0.015)	0.292 (0.402)	0.053 (0.076)	−0.003 (0.078)	0.051 (0.095)
Composite	3	0.900 (0.017)	0.007 (0.203)	−0.029 (0.062)	0.019 (0.058)	−0.008 (0.085)

NOTE: *** = $p < 0.01$. ** = $p < 0.05$. * = $p < 0.1$. We used the Benjamini–Hochberg method to adjust significance levels for multiple hypotheses. The numbers in parentheses are standard errors.

gressive than overall sorting, suggesting that it is also more progressive than between-school sorting.

Reading

- For 2007–2008, sorting within schools was estimated to be negative. Nevertheless, the overall sorting estimate reached 0.3 and was statistically significant.

Table A.15
Aspire Sorting Parameters Using Evaluation Value-Added Measure, by Subject and Year

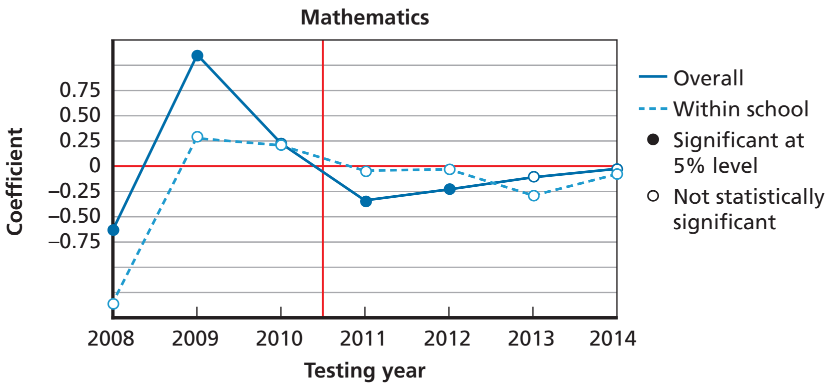
	Beta			Standard Error				
Year	Overall	Between	Within	Overall	Between	Within	Teachers	Schools
Mathematics								
2008	–0.615***	–0.587***	–1.352	0.195	0.200	1.066	52	16
2009	1.096***	1.115***	0.290	0.106	0.106	0.598	75	19
2010	0.228	0.231	0.210	0.130	0.135	0.601	82	22
2011	–0.344***	–0.367***	–0.048	0.087	0.091	0.427	90	24
2012	–0.226***	–0.234***	–0.029	0.081	0.083	0.510	122	27
2013	–0.105	–0.100	–0.291	0.100	0.104	0.543	134	31
2014	–0.024	–0.022	–0.074	0.131	0.132	0.701	118	29
Reading								
2008	0.300***	0.332***	–0.306	0.100	0.101	0.434	55	17
2009	0.067	0.058	0.300	0.121	0.123	0.626	79	20
2010	0.896***	0.898***	0.770	0.102	0.106	0.446	80	23
2011	0.313***	0.275**	0.730	0.107	0.113	0.371	94	24
2012	–0.141*	–0.134*	–0.268	0.071	0.072	0.357	120	26

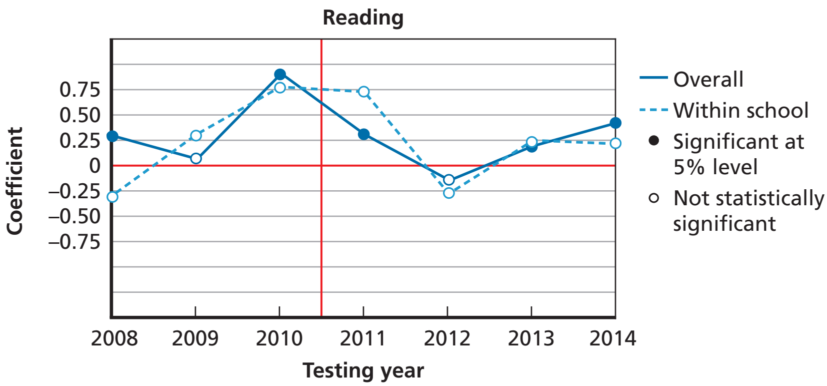
Table A.15—Continued

Year	Beta			Standard Error			Teachers	Schools
	Overall	Between	Within	Overall	Between	Within		
2013	0.189***	0.182**	0.236	0.069	0.070	0.315	132	31
2014	0.420***	0.431***	0.220	0.105	0.108	0.563	28	6

NOTE: The “Beta” columns show three sorting estimates for each subject and year: the overall coefficient, the between-school coefficient, and the within-school coefficient. The “Standard Error” columns show the standard errors associated with the sorting coefficients. The right two columns present the numbers of teachers and schools in each analysis. *** = $p < 0.01$. ** = $p < 0.05$. * = $p < 0.1$. We used the Benjamini–Hochberg method to adjust significance levels for multiple hypotheses. We report between-school estimates only when $n_{schools} \geq 30$.

Figure A.22
Relationship Between Teacher Effects and the Percentage of Students Who Are Low-Income Minority Students in Aspire Public Schools





NOTE: The red vertical line signifies inception of the Intensive Partnerships initiative. The red horizontal line shows a coefficient of 0, which means that a student's LIM status is unrelated to the effectiveness of the teacher to whom the student is assigned. A positive (i.e., above that line) coefficient means that LIM status is associated with *higher* average teacher value-added estimates; a negative (below the red horizontal) coefficient means that LIM students are being taught by teachers with lower value-added estimates than their non-LIM peers are. The light-blue dashed lines represent the estimated relationship of teacher value-added estimates and students' LIM statuses *within* schools and correspond to parameter β'_1 in Equation 3.3 in Chapter Three. The green dash-and-dotted lines represent the estimated relationship of teacher value added in each school and the proportion of students in a school who are classified as LIM, or the *between*-school relationship, and correspond to parameter γ_1 in Equation 3.4 in Chapter Three. The solid dark-blue lines represent overall sorting both between and within schools and correspond to parameter β_1 in Equation 3.2 in Chapter Three.

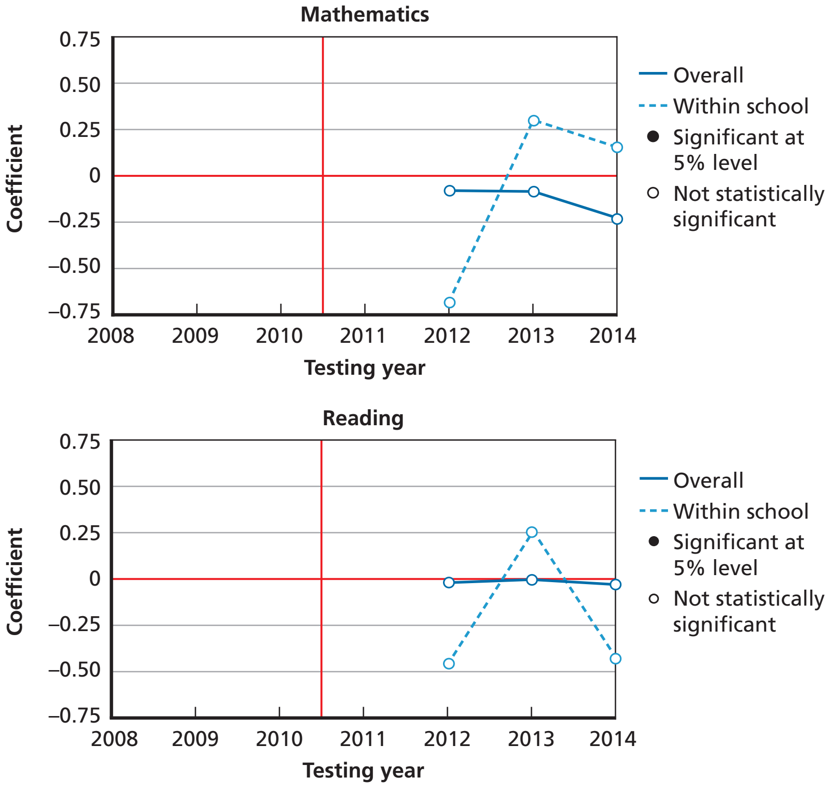
RAND *RR1295/4-A.22*

- Sorting patterns increased until the start of the intervention, at which point all of the sorting patterns became less favorable to LIM students across the next two years.
- However, between 2011–2012 and 2013–2014, the trend reversed itself, and LIM students have gotten increasingly favorable access to effective teachers. The result is that, in 2012–2013 and 2013–2014, LIM students had an overall and statistically significant advantage in access to higher value–added teachers. The 2013–2014 effect estimate of 0.418 corresponded to a difference of 4.2 percent of a student-level standard deviation, on average, between two Aspire reading teachers whose proportions of LIM students differed by 10 percentage points.

For Aspire, we now have site-generated estimates for the 2011–2012 through the 2013–2014 school years. Figure A.23 presents sorting parameters for the sites' teacher-level student-growth estimates, and Figure A.24 shows teacher effectiveness composites. The coefficients, standard errors, and sample sizes corresponding to this figure are shown in the bottom two sections of Table A.16. As before, we urge caution in interpreting these results because of the smallness of the samples.

Although the site estimates for teacher value added are based on an aggregation of student growth to the teacher level, the CMOs in TCRP are actually using an SGP model rather than a value-added model per se. SGPs are not generally classified as value-added models because they do not adjust for student-level or classroom-level attributes other than prior student-level test scores. They are similar to value-added models in that they describe the growth that a teacher's students make compared with other students in a site or district (using that teacher's students' median rather than mean growth), but they are not intended to attribute students' growth causally to a given teacher or set of teachers. For this reason, they do not control for observed student characteristics that might affect the estimate of a teacher's effect (Wright, 2010; Bertelli and Sandoval, 2011). Insofar as teacher effect estimates are confounded with the influence of student background characteristics, such as economic disadvantage, we might expect SGP

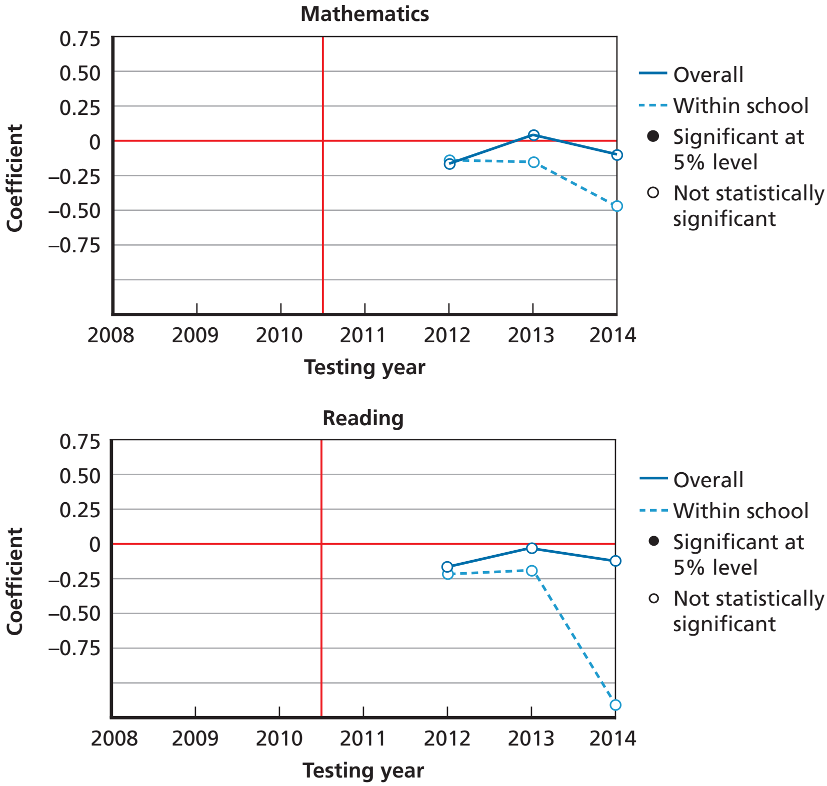
Figure A.23
Sorting of Site-Provided Student Growth Percentile Estimates, by Percentage of Students Who Are Low-Income Minority Students in Aspire Public Schools



NOTE: The red vertical line signifies inception of the Intensive Partnerships initiative. The red horizontal line shows a coefficient of 0, which means that a student's LIM status is unrelated to the effectiveness of the teacher to whom the student is assigned. A positive (i.e., above that line) coefficient means that LIM status is associated with *higher* average teacher value-added estimates; a negative (below the red horizontal) coefficient means that LIM students are being taught by teachers with lower value-added estimates than their non-LIM peers are. The light-blue dashed lines represent the estimated relationship of teacher value-added estimates and students' LIM statuses *within* schools and correspond to parameter β'_1 in Equation 3.3 in Chapter Three. The green dash-and-dotted lines represent the estimated relationship of teacher value added in each school and the proportion of students in a school who are classified as LIM, or the *between*-school relationship, and correspond to parameter γ_1 in Equation 3.4 in Chapter Three. The solid dark-blue lines represent overall sorting both between and within schools and correspond to parameter β_1 in Equation 3.2 in Chapter Three.

RAND *RR1295/4-A.23*

Figure A.24
Sorting of Site-Provided Composite, by Percentage of Students Who Are Low-Income Minority Students in Aspire Public Schools



NOTE: The red vertical line signifies inception of the Intensive Partnerships initiative. The red horizontal line shows a coefficient of 0, which means that a student's LIM status is unrelated to the effectiveness of the teacher to whom the student is assigned. A positive (i.e., above that line) coefficient means that LIM status is associated with *higher* average teacher value-added estimates; a negative (below the red horizontal) coefficient means that LIM students are being taught by teachers with lower value-added estimates than their non-LIM peers are. The light-blue dashed lines represent the estimated relationship of teacher value-added estimates and students' LIM statuses *within* schools and correspond to parameter β'_1 in Equation 3.3 in Chapter Three. The green dash-and-dotted lines represent the estimated relationship of teacher value added in each school and the proportion of students in a school who are classified as LIM, or the *between*-school relationship, and correspond to parameter γ_1 in Equation 3.4 in Chapter Three. The solid dark-blue lines represent overall sorting both between and within schools and correspond to parameter β_1 in Equation 3.2 in Chapter Three.

RAND *RR1295/4-A.24*

Table A.16
Aspire Public Schools Sorting Parameters, by Model and Year

	Beta			Standard Error				
Year	**Overall**	**Between**	**Within**	**Overall**	**Between**	**Within**	**Teachers**	**Schools**
Math, site-based VAM, restricted subsample								
2012	−0.078	−0.059	−0.680	0.093	0.094	0.533	74	25
2013	−0.083	−0.101	0.303	0.116	0.119	0.551	122	31
2014	−0.227	−0.240	0.156	0.135	0.138	0.757	102	29
Reading, site-based VAM, restricted subsample								
2012	−0.018	−0.002	−0.455	0.077	0.079	0.417	69	23
2013	−0.001	−0.013	0.256	0.072	0.075	0.351	120	31
2014	−0.028	−0.015	−0.427	0.136	0.138	0.774	28	6
Math, site-based VAM, all available teachers								
2012	−0.064	−0.079	−0.991	0.108	0.126	0.716	88	25
2013	−0.097	−0.141	0.475	0.139	0.166	0.773	153	31
2014	−0.192	−0.332	0.260	0.144	0.190	1.045	129	29

Table A.16—Continued

Year	Beta			Standard Error			Teachers	Schools
	Overall	Between	Within	Overall	Between	Within		
Reading, site-based VAM, all available teachers								
2012	−0.052	−0.003	−0.681	0.089	0.104	0.555	88	23
2013	−0.061	−0.015	0.349	0.087	0.090	0.426	153	31
2014	−0.108	−0.012	−0.369	0.082	0.111	0.619	129	6
Math, site-based composite, restricted subsample								
2012	−0.168	−0.169	−0.138	0.091	0.093	0.529	74	25
2013	0.042	0.052	−0.153	0.117	0.120	0.553	122	31
2014	−0.101	−0.089	−0.472	0.137	0.140	0.766	102	29
Reading, site-based composite, restricted subsample								
2012	−0.163	−0.163	−0.215	0.074	0.077	0.407	69	23
2013	−0.028	−0.021	−0.192	0.072	0.075	0.351	120	31
2014	−0.123	−0.088	−1.157	0.134	0.137	0.737	28	6
Math, site-based composite, all available teachers								
2012	−0.245**	−0.214	−0.207	0.097	0.117	0.677	88	25
2013	0.010	0.060	−0.131	0.115	0.139	0.645	153	31

Table A.16—Continued

Year	Beta			Standard Error			Teachers	Schools
	Overall	Between	Within	Overall	Between	Within		
2014	−0.121	−0.100	−0.448	0.120	0.158	0.865	129	29
Reading, site-based composite, all available teachers								
2012	−0.201**	−0.196	−0.290	0.079	0.092	0.496	88	23
2013	0.006	−0.023	−0.187	0.073	0.080	0.381	153	31
2014	−0.068	−0.068	−0.976	0.068	0.105	0.557	129	6

NOTE: The “Beta” columns show three sorting estimates for each subject and year: the overall coefficient, the between-school coefficient, and the within-school coefficient. The “Standard Error” columns show the standard errors associated with the sorting coefficients. The right two columns present the numbers of teachers and schools in each analysis. *** = $p < 0.01$. ** = $p < 0.05$. * = $p < 0.1$.

Table A.17
Aspire Public Schools Changes in Low-Income Minority Composition, by Performance Level: Value-Added Measure Tercile

	Prereform				Early Reform				Recent Reform			
Change	**1**	**2**	**3**	**Total**	**1**	**2**	**3**	**Total**	**1**	**2**	**3**	**Total**
Mathematics												
Increase LIM, same school	0.462	0.329	0.280	0.346	0.371	0.408	0.441	0.407	0.154	0.286	0.220	0.246
	(0.144)	(0.094)	(0.170)	(0.072)	(0.134)	(0.076)	(0.128)	(0.059)	(0.128)	(0.068)	(0.125)	(0.054)
Decrease LIM, same school	0.115	0.237	0.440	0.252	0.171	0.233	0.118	0.198	0.288	0.338	0.480	0.355
	(0.184)	(0.100)	(0.150)	(0.077)	(0.154)	(0.086)	(0.161)	(0.068)	(0.117)	(0.066)	(0.102)	(0.050)
Increase LIM, new school	0.077	0.053	0.000	0.047	0.000	0.019	0.029	0.017	0	0	0	0
	(0.188)	(0.112)	(0.000)	(0.087)	(0.000)	(0.098)	(0.169)	(0.076)	—	—	—	—
Decrease LIM, new school	0	0	0	0	0	0	0	0	0.000	0.006	0.020	0.008
	—	—	—	—	—	—	—	—	(0.000)	(0.080)	(0.140)	(0.062)
Out of tested	0.038	0.132	0.160	0.118	0.086	0.136	0.147	0.128	0.212	0.143	0.100	0.148
	(0.192)	(0.107)	(0.183)	(0.083)	(0.162)	(0.092)	(0.158)	(0.071)	(0.123)	(0.075)	(0.134)	(0.058)
Out of district	0.308	0.250	0.120	0.236	0.371	0.204	0.265	0.250	0.346	0.227	0.180	0.242
	(0.163)	(0.099)	(0.188)	(0.078)	(0.134)	(0.088)	(0.147)	(0.066)	(0.112)	(0.071)	(0.128)	(0.054)

Table A.17—Continued

	Prereform				Early Reform				Recent Reform			
Change	**1**	**2**	**3**	**Total**	**1**	**2**	**3**	**Total**	**1**	**2**	**3**	**Total**
Total proportion	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Total number of schools	26	76	25	127	35	103	34	172	52	154	50	256
Reading												
Increase LIM, same school	0.370	0.321	0.154	0.299	0.400	0.419	0.382	0.408	0.118	0.187	0.180	0.171
	(0.153)	(0.092)	(0.180)	(0.072)	(0.131)	(0.074)	(0.135)	(0.058)	(0.132)	(0.074)	(0.128)	(0.057)
Decrease LIM, same school	0.481	0.284	0.308	0.328	0.314	0.229	0.235	0.247	0.235	0.213	0.220	0.219
	(0.139)	(0.094)	(0.163)	(0.071)	(0.140)	(0.086)	(0.150)	(0.066)	(0.122)	(0.072)	(0.125)	(0.056)
Increase LIM, new school	0.000	0.049	0.038	0.037	0.000	0.010	0.000	0.006	0.000	0.007	0.000	0.004
	(0.000)	(0.108)	(0.192)	(0.085)	(0.000)	(0.097)	(0.000)	(0.076)	(0.000)	(0.081)	(0.000)	(0.063)
Decrease LIM, new school	0.000	0.012	0.000	0.007	0.029	0.000	0.000	0.006	0.020	0.000	0.000	0.004
	(0.000)	(0.110)	(0.000)	(0.086)	(0.167)	(0.000)	(0.000)	(0.076)	(0.139)	(0.000)	(0.000)	(0.063)
Out of tested	0.074	0.123	0.231	0.134	0.171	0.086	0.147	0.115	0.157	0.133	0.080	0.127
	(0.185)	(0.104)	(0.172)	(0.080)	(0.154)	(0.093)	(0.158)	(0.071)	(0.129)	(0.076)	(0.136)	(0.059)

Table A.17—Continued

	Prereform				Early Reform				Recent Reform			
Change	**1**	**2**	**3**	**Total**	**1**	**2**	**3**	**Total**	**1**	**2**	**3**	**Total**
Out of district	0.074	0.210	0.269	0.194	0.086	0.257	0.235	0.218	0.471	0.460	0.520	0.474
	(0.185)	(0.099)	(0.168)	(0.078)	(0.162)	(0.084)	(0.150)	(0.067)	(0.102)	(0.060)	(0.098)	(0.046)
Total proportion	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Total number of schools	27	81	26	134	35	105	34	174	51	150	50	251

NOTE: The numbers in parentheses are standard errors

Table A.18
Aspire Public Schools Changes in Low-Income Minority Composition, by Performance Level: Site Value-Added Measure Tercile

	Recent Reform			
Change	**1**	**2**	**3**	**Total**
Mathematics				
Increase LIM, same school	0.100	0.314	0.237	0.255
	(0.150)	(0.076)	(0.142)	(0.062)
Decrease LIM, same school	0.350	0.449	0.368	0.413
	(0.127)	(0.068)	(0.129)	(0.055)
Increase LIM, new school	0	0	0	0
	—	—	—	—
Decrease LIM, new school	0.000	0.008	0.026	0.010
	(0.000)	(0.092)	(0.160)	(0.071)
Out of tested	0.275	0.102	0.158	0.148
	(0.135)	(0.087)	(0.149)	(0.066)
Out of district	0.275	0.127	0.211	0.173
	(0.135)	(0.086)	(0.144)	(0.065)
Total proportion	1.000	1.000	1.000	1.000
Total number of schools	40	118	38	196
Reading				
Increase LIM, same school	0.132	0.172	0.229	0.175
	(0.151)	(0.084)	(0.148)	(0.066)
Decrease LIM, same school	0.263	0.250	0.171	0.238
	(0.139)	(0.080)	(0.154)	(0.063)
Increase LIM, new school	0	0	0	0
	—	—	—	—

Table A.18—Continued

	Recent Reform			
Change	**1**	**2**	**3**	**Total**
Decrease LIM, new school	0.000	0.009	0.000	0.005
	(0.000)	(0.092)	(0.000)	(0.073)
Out of tested	0.237	0.095	0.171	0.138
	(0.142)	(0.088)	(0.154)	(0.068)
Out of district	0.368	0.474	0.429	0.444
	(0.129)	(0.067)	(0.128)	(0.054)
Total proportion	1.000	1.000	1.000	1.000
Total number of schools	38	116	35	189

NOTE: The numbers in parentheses are standard errors.

Table A.19
Aspire Public Schools Changes in Low-Income Minority Composition, by Performance Level: Site Composite Tercile

	Recent Reform			
Change	**1**	**2**	**3**	**Total**
Mathematics				
Increase LIM, same school	0.122	0.316	0.211	0.255
	(0.146)	(0.076)	(0.144)	(0.062)
Decrease LIM, same school	0.439	0.368	0.526	0.413
	(0.117)	(0.074)	(0.112)	(0.055)
Increase LIM, new school	0	0	0	0
	—	—	—	—
Decrease LIM, new school	0.000	0.017	0.000	0.010
	(0.000)	(0.092)	(0.000)	(0.071)
Out of tested	0.220	0.137	0.105	0.148
	(0.138)	(0.086)	(0.153)	(0.066)

Table A.19—Continued

	Recent Reform			
Change	**1**	**2**	**3**	**Total**
Out of district	0.220	0.162	0.158	0.173
	(0.138)	(0.085)	(0.149)	(0.065)
Total proportion	1.000	1.000	1.000	1.000
Total number of schools	41	117	38	196
Reading				
Increase LIM, same school	0.103	0.184	0.222	0.175
	(0.152)	(0.085)	(0.147)	(0.066)
Decrease LIM, same school	0.308	0.254	0.111	0.238
	(0.133)	(0.081)	(0.157)	(0.063)
Increase LIM, new school	0	0	0	0
	—	—	—	—
Decrease LIM, new school	0.026	0.000	0.000	0.005
	(0.158)	(0.000)	(0.000)	(0.073)
Out of tested	0.205	0.123	0.111	0.138
	(0.143)	(0.088)	(0.157)	(0.068)
Out of district	0.359	0.439	0.556	0.444
	(0.128)	(0.070)	(0.111)	(0.054)
Total proportion	1.000	1.000	1.000	1.000
Total number of schools	39	114	36	189

NOTE: The numbers in parentheses are standard errors.

estimates to be more negatively associated with LIM status than a value-added estimate that attempts to remove the effects of student background characteristics.

Bearing this caveat in mind, we make the following observations regarding the sorting of site-generated teacher effectiveness estimates in 2011–2012:

- In the first observed year (2012–2013) for the site SGP, between-school sorting was strongly favorable to LIM students for both math and reading; within-school sorting was unfavorable, leading to a nearly 0 and insignificant overall effect.
- During the three years, despite erratic jumps in the within- and between-sorting parameters, the overall sorting parameters have remained relatively constant (and slightly negative although insignificant) in math and reading for the SGP.
- The composite scores have been even more negative overall than the SGP, consistent with our findings for the other districts. The composite scores show slightly increased access of LIM students to effective teachers, although the coefficients are still negative. Increased between-school sorting drives the improvements.

Mechanisms That Aspire Public Schools Used to Change Distribution

We also estimate the changes in LIM composition for teachers by value-added level as described in Chapter Five for the other sites, and the mechanisms by which access is changed. In Aspire, the estimated effects of the actions are much larger than they are for the districts, but, because of the smallness of the sample, these are imprecise estimates and are never significant in the desired direction. However, Table A.20 and Figure A.28 indicate that the teachers of LIM students had been improving their value added less than the teachers of non-LIM students during the early years of the initiative; that is no longer the case. These and other estimates reflecting the changes in LIM access are presented below.

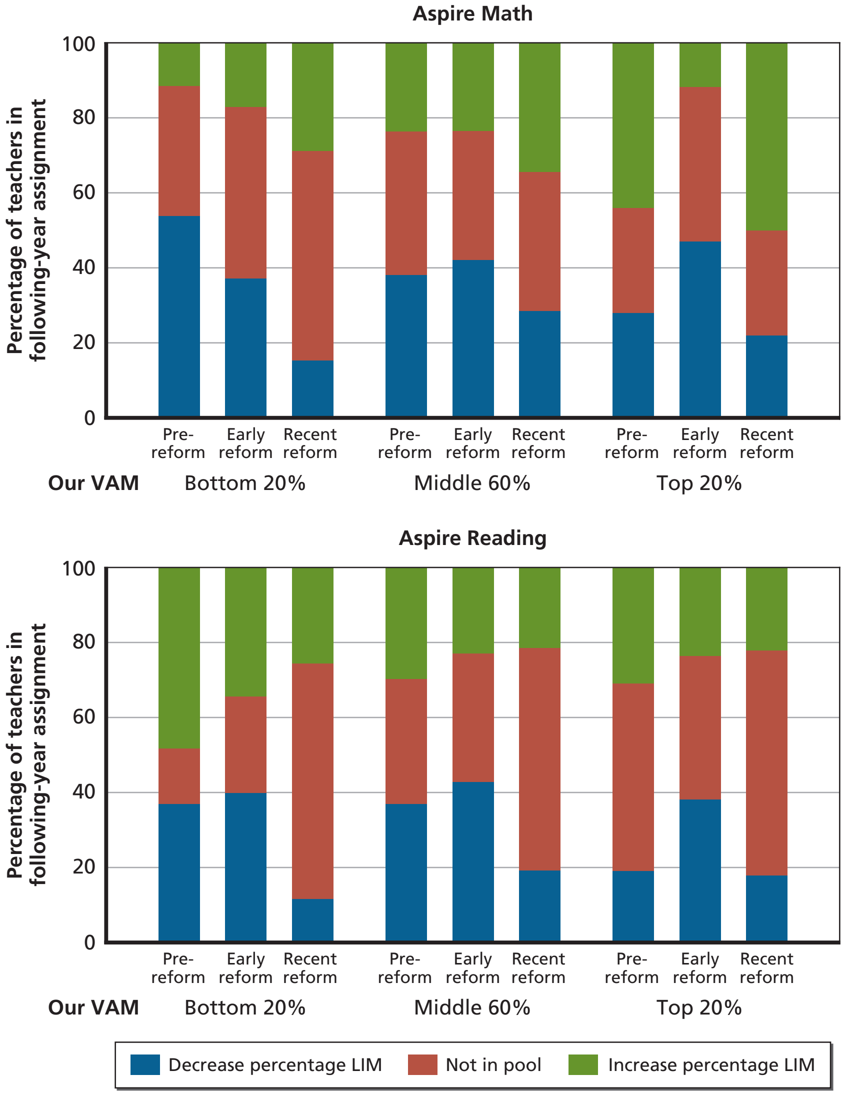
In sum, the sorting estimates presented above, although imprecise, suggest that LIM students' access to effective teaching has increased in Aspire after an initial worsening. LIM students are no longer being taught by less effective math and reading teachers, and LIM students' teachers are no longer improving less than non-LIM students' teachers.

Table A.20
Aspire Public Schools Decomposition of Change in Overall Sorting

Variable	Period	p	$\Delta_{replace}$	$\Delta_{improve}$	$\Delta_{reassign}$	$\hat{\beta}_t - \hat{\beta}_{t-1}$
Mathematics						
Our VAM	1	0.584 (0.047)	0.777 (0.352)	0.021 (0.251)	−0.193 (0.306)	0.422* (0.195)
	2	0.563 (0.041)	−0.307 (0.226)	−0.261 (0.149)	0.020 (0.161)	−0.227 (0.138)
	3	0.633 (0.034)	−0.260 (0.220)	0.226 (0.149)	−0.040 (0.139)	0.101 (0.147)
Site VAM	3	0.709 (0.033)	−0.195 (0.390)	−0.069 (0.187)	0.017 (0.186)	−0.074 (0.164)
Composite	3	0.709 (0.033)	0.401 (0.306)	−0.027 (0.168)	0.045 (0.170)	0.033 (0.164)
Reading						
Our VAM	1	0.619 (0.045)	0.341 (0.242)	0.235 (0.156)	−0.026 (0.165)	0.298* (0.158)
	2	0.604 (0.040)	−0.403 (0.225)	−0.501*** (0.117)	−0.033 (0.126)	−0.518*** (0.138)
	3	0.674 (0.034)	0.200 (0.174)	0.128 (0.088)	0.059 (0.090)	0.280** (0.113)
Site VAM	3	0.742 (0.034)	0.176 (0.238)	−0.023 (0.109)	0.026 (0.113)	−0.005 (0.130)
Composite	3	0.742 (0.034)	0.218 (0.195)	0.000 (0.104)	0.045 (0.106)	0.020 (0.128)

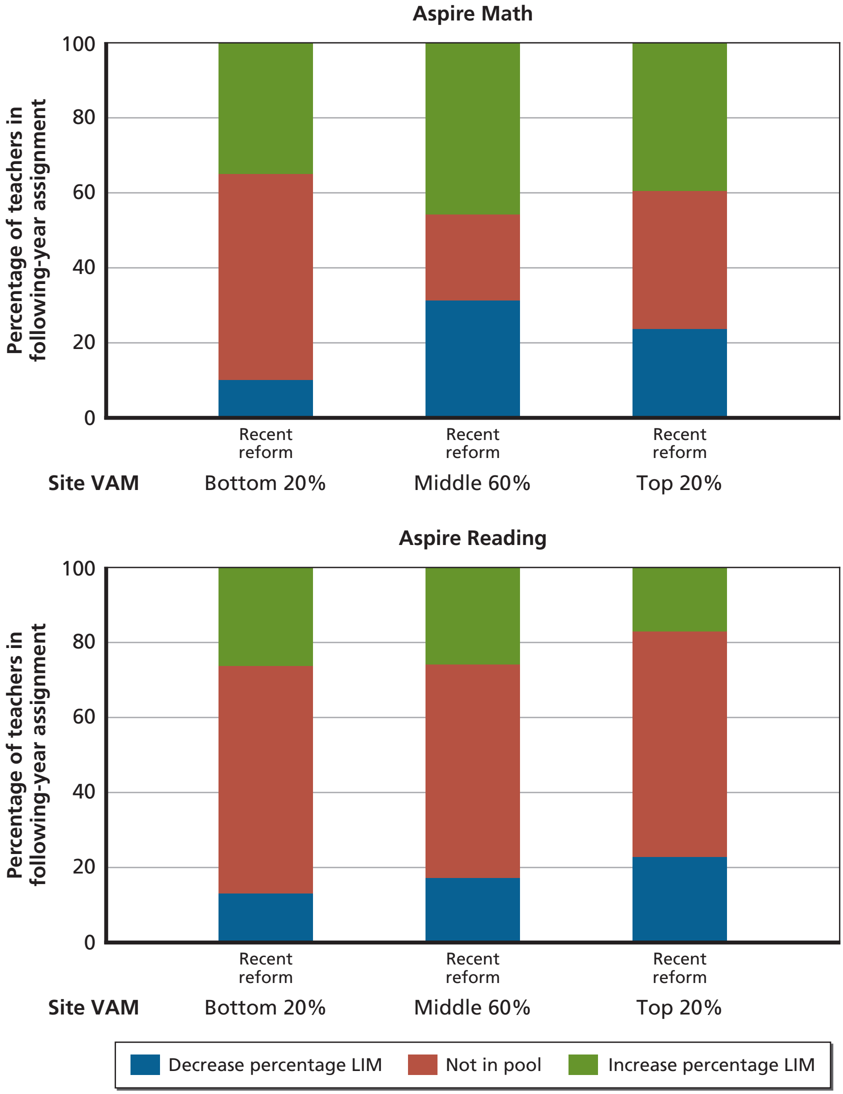
NOTE: *** = $p < 0.01$. ** = $p < 0.05$. * = $p < 0.1$. We used the Benjamini–Hochberg method to adjust significance levels for multiple hypotheses. The numbers in parentheses are standard errors.

Figure A.25
Changes in Low-Income Minority Composition, by Teacher Performance Level, Aspire Public Schools, Our Value-Added Measure



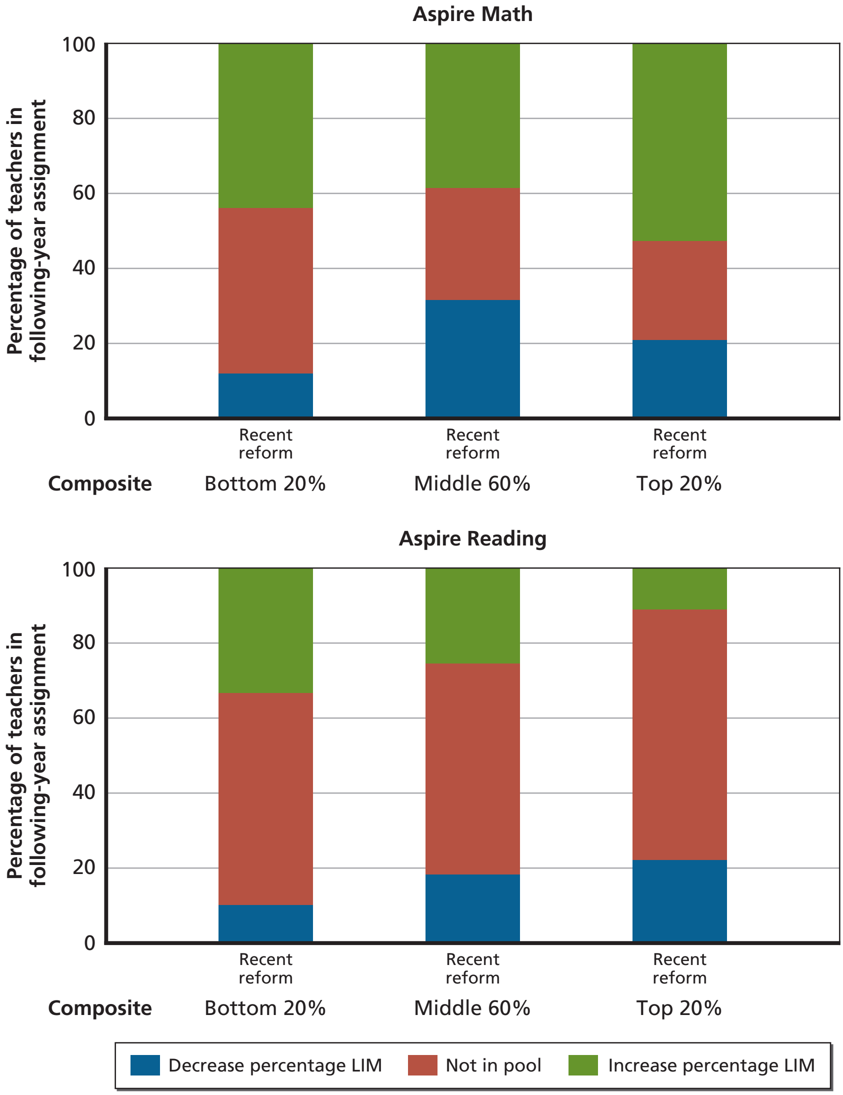
RAND *RR1295/4-A.25*

Figure A.26
Changes in Low-Income Minority Composition, by Teacher Performance Level, Aspire Public Schools, Site Value-Added Measure



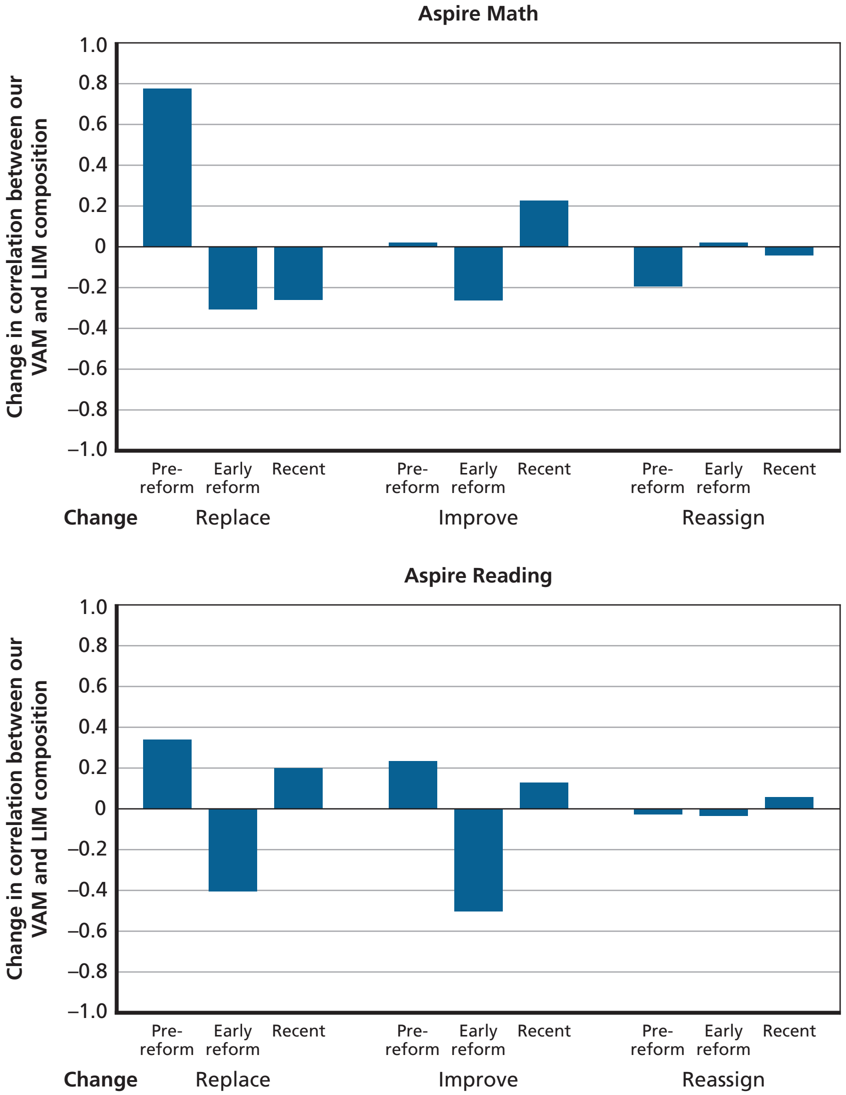
RAND *RR1295/4-A.26*

Figure A.27
Changes in Low-Income Minority Composition, by Teacher Performance Level, Aspire Public Schools Composite



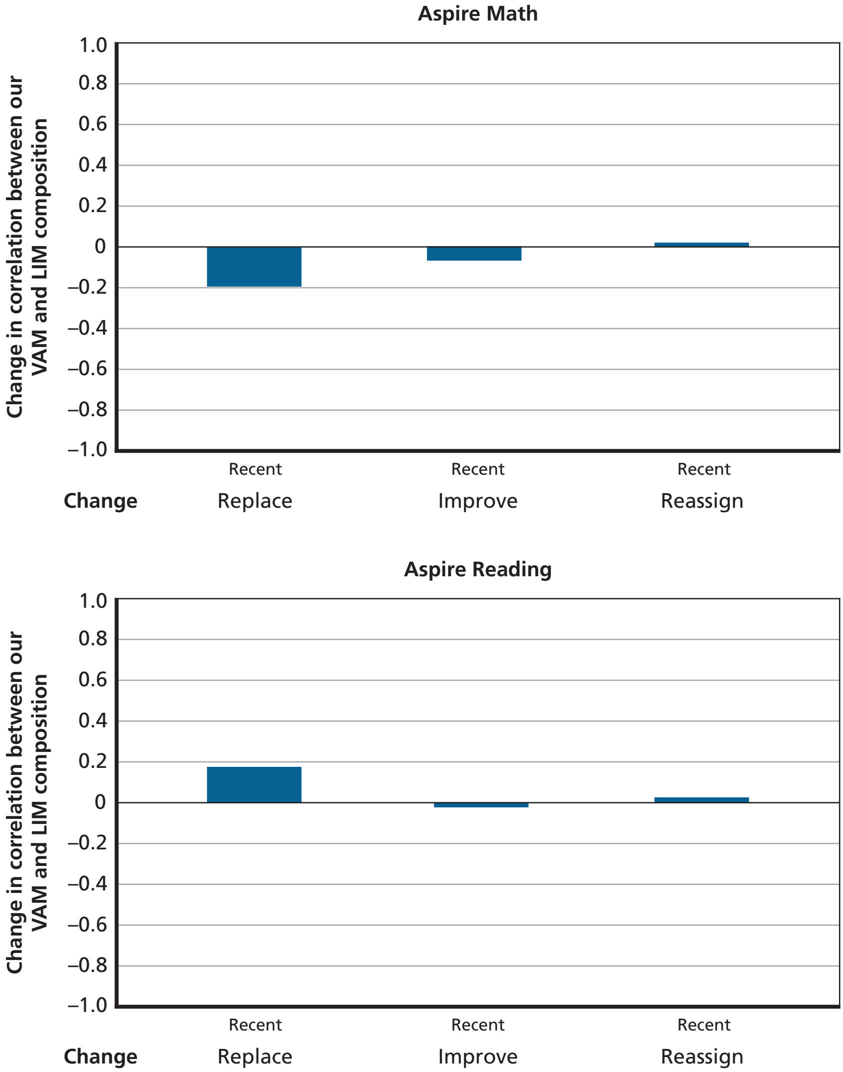
RAND *RR1295/4-A.27*

Figure A.28
Decomposition of Change in Overall Sorting, Aspire Public Schools, Our Value-Added Measure



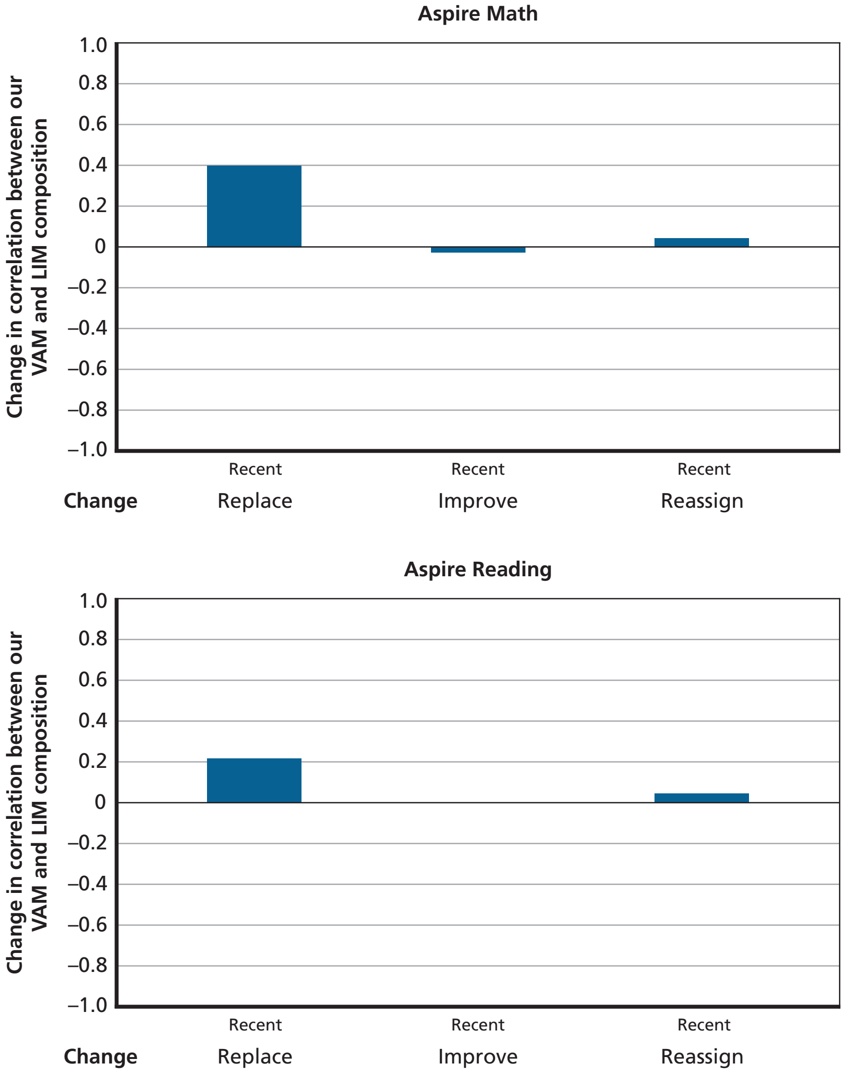
RAND *RR1295/4-A.28*

Figure A.29
Decomposition of Change in Overall Sorting, Aspire Public Schools, Site Value-Added Measure



RAND *RR1295/4-A.29*

Figure A.30
Decomposition of Change in Overall Sorting, Aspire Public Schools, Composite



RAND *RR1295/4-A.30*

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, January 2007, pp. 95–135. As of April 13, 2016: http://eric.ed.gov/?id=ED505650

Benjamini, Yoav, and Yosef Hochberg, "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *Journal of the Royal Statistical Society, Series B (Methodological)*, Vol. 57, No. 1, 1995, pp. 289–300.

Bertelli, Chris, and C. Sandoval, "The College-Ready Promise," presentation for the Education Trust—West conference Innovations in Teacher Evaluation: A Forum on Promoting Teacher Effectiveness, 2011.

Betebenner, Damian W., "Norm- and Criterion-Referenced Student Growth," *Educational Measurement: Issues and Practice*, Vol. 28, No. 4, Winter 2009, pp. 42–51.

Bloom, Howard S., Carolyn J. Hill, Alison Rebeck Black, and Mark W. Lipsey, "Performance Trajectories and Performance Gaps as Achievement Effect-Size Benchmarks for Educational Interventions," *Journal of Research on Educational Effectiveness*, Vol. 1, No. 4, 2008, pp. 289–328. As of April 13, 2016: http://eric.ed.gov/?id=EJ873861

Borjas, George J., and Glenn T. Sueyoshi, "A Two-Stage Estimator for Probit Models with Structural Group Effects," *Journal of Econometrics*, Vol. 64, No. 1–2, September–October 1994, pp. 165–182.

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, Colo.: National Education Policy Center, February 2011. As of April 13, 2016: http://eric.ed.gov/?id=ED516008

Goldhaber, Dan, and Michael Hansen, *Is It Just a Bad Class? Assessing the Stability of Measured Teacher Performance*, Seattle, Wash.: Center for Education Data and Research, Working Paper 2010-3.0, 2010. As of April 13, 2016:
http://eric.ed.gov/?id=ED537151

Gutierrez, Italo, Gabriel Weinberger, and John Engberg, *The Impact on Students of the Intensive Partnership for Effective Teaching Through 2013–2014*, Santa Monica, Calif.: RAND Corporation, in production.

Hock, Heinrich, and Eric Isenberg, *Methods for Accounting for Co-Teaching in Value-Added Models*, Washington, D.C.: Mathematica Policy Research, June 2012. As of April 25, 2016:
http://eric.ed.gov/?id=ED533144

Isenberg, Eric, Jeffrey Max, Philip Gleason, Liz Potamites, Robert Santillano, Heinrich Hock, and Michael Hansen, *Access to Effective Teaching for Disadvantaged Students*, Washington, D.C.: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education, NCEE 2014-4001, November 2013. As of April 13, 2016:
http://files.eric.ed.gov/fulltext/ED544367.pdf

Lockwood, J. R., and Daniel F. McCaffrey, "Correcting for Test Score Measurement Error in ANCOVA Models for Estimating Treatment Effects," *Journal of Educational and Behavioral Statistics*, Vol. 39, No. 1, February 2014, pp. 22–52.

McCaffrey, Daniel F., J. R. Lockwood, Daniel Koretz, Thomas A. Louis, and Laura Hamilton, "Models for Value-Added Modeling of Teacher Effects," *Journal of Educational and Behavioral Statistics*, Vol. 29, No. 1, Spring 2004, pp. 67–101. As of April 13, 2016:
http://www.rand.org/pubs/reprints/RP1165.html

Raudenbush, Stephen W., and Anthony S. Bryk, *Hierarchical Linear Models: Applications and Data Analysis Methods*, Vol. 1, Thousand Oaks, Calif.: Sage Publications, 2002.

Schochet, Peter Z., and Hanley S. Chiang, *Error Rates in Measuring Teacher and School Performance Based on Student Test Score Gains*, Washington, D.C.: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education, NCEE 2010-4004, 2010. As of April 13, 2016:
http://files.eric.ed.gov/fulltext/ED511026.pdf

Stecher, Brian, and Michael S. Garet, "Introduction to the Evaluation of the Intensive Partnerships for Effective Teaching (IP)," paper presented at the 39th annual conference of the Association for Education Finance and Policy, San Antonio, Texas, 2014.

Stecher, Brian M., Michael S. Garet, Laura S. Hamilton, Elizabeth D. Steiner, Abby Robyn, Jeffrey Poirier, Deborah Holtzman, Eleanor S. Fulbeck, Jay Chambers, and Iliana Brodziak de los Reyes, *Improving Teaching Effectiveness: The Intensive Partnerships for Effective Teaching After Five Years*, Santa Monica, Calif.: RAND Corporation, in production.

Stecher, Brian, Mike Garet, Deborah Holtzman, and Laura Hamilton, "Implementing Measures of Teacher Effectiveness," *Phi Delta Kappan*, Vol. 94, No. 3, November 2012, pp. 39–43.

Steele, Jennifer L., Matthew Baird, John Engberg, and Gerald Paul Hunter, *Trends in the Distribution of Teacher Effectiveness in the Intensive Partnerships for Effective Teaching: Progress Report*, Santa Monica, Calif.: RAND Corporation, 2014.

Whitehurst, Grover J., Matthew M. Chingos, and Katharine M. Lindquist, *Evaluating Teachers with Classroom Observations: Lessons Learned in Four Districts*, Washington, D.C.: Brookings Institution, May 13, 2014. As of April 13, 2016: http://www.brookings.edu/research/reports/2014/05/13-teacher-evaluation-whitehurst-chingos

Wright, S. Paul, *An Investigation of Two Nonparametric Regression Models for Value-Added Assessment in Education*, Cary, N.C.: SAS, March 10, 2010.

As part of its effective-teaching initiative, Intensive Partnerships for Effective Teaching, the Bill & Melinda Gates Foundation has partnered with three urban school districts across the United States and a group of four charter management organizations to undertake a set of strategic human-capital reforms. The reforms are intended to improve teachers' overall effectiveness and to ensure that low-income minority (LIM) students have access to highly effective teachers. Lack of access to effective teaching has been identified as a possible contributor to the well-documented achievement gap between LIM students and their more-advantaged peers.

This report attends to the distribution of effective teachers within and across schools in the Intensive Partnership sites. The authors first examine the trends in the distribution of effective teachers between LIM students and other students. They also examine whether any of a variety of mechanisms can explain changes in LIM students' access to effective teaching. These mechanisms include increasing the percentage of LIM students whom effective teachers teach, increasing the effectiveness of teachers with large percentages of LIM students, and replacing less effective teachers of LIM students with more-effective teachers.

RAND EDUCATION

www.rand.org

$34.00

ISBN-10 0-8330-9561-7
ISBN-13 978-0-8330-9561-9

53400

9 780833 095619

RR-1295/4-BMGF