The National Study of Charter Management Organization (CMO) Effectiveness

Charter-School Management Organizations:
Diverse Strategies and Diverse Student Impacts

Joshua Furgeson, Brian Gill,
Joshua Haimson, Alexandra Killewald,
Moira McCullough, Ira Nichols-Barrer,
Bing-ru Teh, Natalya Verbitsky-Savitz
Mathematica Policy Research

Melissa Bowen, Allison Demeritt,
Paul Hill, Robin Lake
Center on Reinventing Public Education

January 2012
Updated Edition

MATHEMATICA Policy Research

center on reinventing public education

Charter- School Management Organizations: Diverse Strategies and Diverse Student Impacts

January 2012

Updated Edition

Mathematica

Joshua Furgeson
Brian Gill
Joshua Haimson
Alexandra Killewald
Moira McCullough
Ira Nichols-Barrer
Natalya Verbitsky-Savitz
Bing Ru Teh

Center on Reinventing Public Education
Melissa Bowen
Allison Demeritt
Paul Hill
Robin Lake

With Assistance From:
Michael Barna
Emily Caffery
Hanley Chiang
John Deke
Melissa Dugger
Emma Ernst
Alena Davidoff-Gore
Eric Grau
Thomas Decker
Mason DeCamillis
Philip Gleason
Amanda Hakanson
Jane Nelson
Antoniya Owens
Julie Redline
Davin Reed
Chris Rodger
Margaret Sullivan
Christina Tuttle
Justin Vigeant
Tiffany Waits
Clare Wolfendale

The National Study of CMO Effectiveness is a longitudinal research effort designed to measure how nonprofit charter school management organizations (CMOs) affect student achievement and to examine the internal structures, practices, and policy contexts that may influence these outcomes. The study began in May 2008 and will conclude in 2012.

This updated edition of the report presents findings on CMO students, resources, and practices as well as CMO impacts on student achievement in middle school, high school graduation, and postsecondary enrollment. It also examines the relationships between CMO practices and impacts. A subsequent report will describe promising CMO practices in more detail.

The study is being conducted by Mathematica Policy Research and the University of Washington's Center on Reinventing Public Education (CRPE). It was commissioned by NewSchools Venture Fund, with the generous support of the Bill & Melinda Gates Foundation and the Walton Family Foundation.

Mathematica Policy Research (www.mathematica-mpr.com) seeks to improve public well-being by conducting studies and assisting clients with program evaluation and policy research, survey design and data collection, research assessment and interpretation, and program performance/data management. Its clients include foundations, federal and state governments, and private-sector and international organizations. The employee-owned company, with offices in Princeton, NJ, Ann Arbor, MI, Cambridge, MA, Chicago, IL, Oakland, CA, and Washington, DC, has conducted some of the most important studies of health care, international, disability, education, family support, employment, nutrition, and early childhood policies and programs.

The Center on Reinventing Public Education at the University of Washington engages in independent research and policy analysis on a range of K-12 public education reform issues, including choice & charters, finance & productivity, teachers, urban district reform, leadership, and state & federal reform. CRPE's work is based on two premises: that public schools should be measured against the goal of educating all children well, and that current institutions too often fail to achieve this goal. Our research uses evidence from the field and lessons learned from other sectors to understand complicated problems and to design innovative and practical solutions for policymakers, elected officials, parents, educators, and community leaders.

CONTENTS

IV *(continued)*

V (*continued*)

TABLES

FIGURES

Technical Working Group Members

- **Diane Burton,** Associate Professor of Sociology, Cornell University
- **Tom Cook**, Professor of Sociology, Psychology, and Education and Social Policy; Faculty Fellow, Institute for Policy Research; and Joan and Sarepta Harrison Chair in Ethics and Justice at Northwestern University
- **Jim Kemple**, Founding Executive Director, Research Alliance for New York City Schools
- **Laura Hamilton**, Senior Behavioral Scientist at The RAND Corporation
- **Jane Hannaway**, Vice President and Director, Analysis of Longitudinal Data in Education Research Program, American Institutes for Research
- **Susanna Loeb**, Associate Professor of Education; Director of the Institute for Research on Education Policy & Practice at Stanford University

Stakeholder Advisory Board Members

- **Michael Casserly**, Executive Director, Council of the Great City Schools
- **Michael Cohen**, President, Achieve
- **Josh Edelman**, Deputy Chief, Office of School Innovation, District of Columbia Public Schools
- **Maria Goodloe-Johnson**, Former Superintendent, Seattle Public Schools
- **Beverly L. Hall**, Former Superintendent, Atlanta Public Schools
- **Jeff Henig**, Professor of Political Science and Education at Teachers College; Professor of Political Science at Columbia University
- **Deb Meyerson**, Associate Professor of Organizational Behavior at Stanford University's School of Education; Co-director of Stanford's Center for Research in Philanthropy and Civil Society
- **Abelardo Saavedra**, Former Superintendent, Houston Independent School District
- **Tony Smith**, Superintendent, Oakland Unified School District
- **Dacia Toll**, President, Achievement First
- **Jonathan Williams**, Founder/Co-director, The Accelerated Schools

ACKNOWLEDGMENTS

This report, The National Study of Charter Management Effectiveness, reflects the contributions of a large number of educators, researchers, advisors, and funder staff without whom the study would not have been possible.

First, we would like to thank the staff from charter-school management organizations (CMOs), districts, and states who generously shared their thoughts and data with the study team. Our findings are based on this invaluable information.

Our Technical Working Group and Stakeholder Advisory Board (listed above) were generous with their time and provided indispensable advice and feedback on the study design, analysis plan, and preliminary findings.

The study was commissioned by NewSchools Venture Fund, with support from the Bill & Melinda Gates Foundation and Walton Family Foundation. At NewSchools, Jim Peyser and Joanne Weiss provided very thoughtful direction, support, and guidance. Todd Kern, a consultant to NewSchools, provided ongoing project management assistance and very helpful feedback on nearly all of our deliverables including this report. Kerri Kerr, another consultant, provided incisive feedback on study design issues and successfully coordinated all of our Technical Working Group meetings. At Gates, David Silver, Lance Potter, and Steve Cantrell provided constructive guidance and direction, as did Marc Holley and Sheree Speakman at Walton.

The study team includes many individuals at Mathematica and CRPE beyond the authors of this report. At Mathematica, several staff provided expert analytic guidance, including John Deke, Hanley Chiang, and Phil Gleason. Margaret Sullivan led the recruitment of CMO schools. Tiffany Waits led the successful surveys of principals and teachers and Eric Grau developed the weights for the survey data. The data cleaning effort was led by Chris Rodger. The expert data cleaning and analysis team included Mike Barna, Thomas Decker, Emma Ernst, Alena Davidoff-Gore, Mason DeCamillis, Amanda Hakanson, Antoniya Owens, Davin Reed, Justin Vigeant, and Clare Wolfendale. Julie Redline carefully implemented the experimental weights and conducted experimental impact estimates. We received extremely helpful comments on the draft report from Christina Tuttle and Phil Gleason, expert editing assistance from Cindy George, and great word processing and graphics help from Jane Nelson, Cindy McClure, and Marjorie Mitchell. Jennifer de Vallance, Joanne Pfleiderer, Adam Coyne, and Amy Berridge provided substantial assistance with communications and dissemination. At CRPE, Michael DeArmond, Betheny Gross, and Katherine Martin provided thoughtful guidance and research assistance; Debra Britt helped with communications.

EXECUTIVE SUMMARY

Charter schools—public schools of choice that are operated autonomously, outside the direct control of local school districts—have become more prevalent over the past two decades. There is no consensus about whether, on average, charter schools are doing better or worse than conventional public schools at promoting the achievement of their students. Nonetheless, one research finding is clear: Effects vary widely among different charter schools. Many educators, policymakers, and funders are interested in ways to identify and replicate successful charter schools and help other public schools adopt effective charter school practices.

Charter-school management organizations (CMOs), which establish and operate multiple charter schools, represent one prominent attempt to bring high performance to scale. Many CMOs were created in order to replicate educational approaches that appeared to be effective, particularly among disadvantaged students. Attracting substantial philanthropic support, CMO schools have grown rapidly in the past decade. Some of these organizations have received laudatory attention through anecdotal reports of dramatic achievement results.

The National Study of CMO Effectiveness aims to fill the gap in systematic evidence about CMOs, providing the first rigorous nationwide examination of CMOs' effects on students' achievement and attainment. The study includes an examination of the relationships between the practices of individual CMOs and their effects on student achievement, with the aim of providing useful guidance to the field. Mathematica Policy Research and the Center on Reinventing Public Education (CRPE) are conducting the study with funding from the Bill & Melinda Gates Foundation and the Walton Family Foundation and project management assistance from the NewSchools Venture Fund. This updated edition of the report provides key findings from the study on CMO practices, impacts, and the relationships between them. A forthcoming report will explore promising practices in greater depth.

A. Research Questions, CMOs in Study, and Data Sources

This study uses multiple data sources to describe CMOs, assess their impacts on students, and identify practices associated with positive impacts in order to address the following research questions:

1. **Characteristics and Context.** How quickly are CMOs growing? Which students and areas do they serve and what resources do they use? What are the practices and structures of CMOs? What state policies and other factors appear to influence the location and growth of CMOs?
2. **Impacts.** What are the impacts of CMOs on student outcomes and to what extent do these impacts vary across CMOs?
3. **Promising Practices.** Which CMO practices and structures are positively associated with impacts?

Previous studies have defined CMOs in various ways. Our study focuses primarily on nonprofit CMOs that directly control four or more schools. To be eligible for the study, an organization had to (1) have four schools open by fall 2007, (2) be nonprofit since inception, (3) not primarily serve dropouts or similar special populations, and (4) directly manage schools (that is, be able to hire and fire school principals). Across the United States, we identified 40 CMOs with 292 schools that satisfied the four criteria. These CMOs schools are located in 14 states, with the largest concentrations in Texas, California, Arizona, Ohio, Illinois, New York, and the District of Columbia.

To examine eligible CMOs and address the research questions, we conducted a survey of CMO central office staff, surveys of CMO principals and principals in nearby conventional public schools, a survey of CMO teachers, and site visits to 10 CMOs and 20 schools. In addition, we collected and analyzed school records with data on student characteristics and outcomes (including test scores), and we examined CMO financial records and business plans.

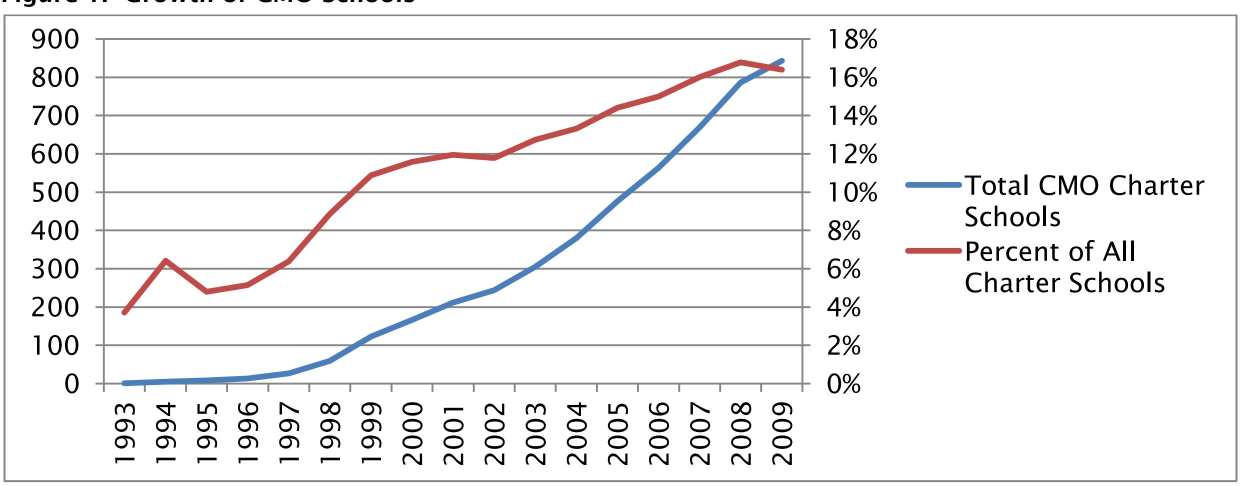
B. Areas and Students Served and Resources Used by CMOs

1. CMO-run schools now represent a substantial share of charter schools and are concentrated in certain states and urban areas

Broadly defined, there are approximately 130 CMOs in the United States, serving nearly 250,000 students.[1] CMOs represent approximately 16 percent of all charter schools operating nationally, up from 12 percent in 2000 (Figure 1). Between 1999 and 2009, the number of CMO schools increased by approximately 20 percent per year, substantially more than the rate for independent charter schools. In 2009, growth in CMO schools appeared to slow to a rate comparable to that of independent charter schools. The growth trajectory of individual CMOs varies according to mission, capacity and constraints imposed by states and charter authorizers. On average, CMOs in our study had opened about one new school per year for the first six years. After seven years of operation, the average pace picks up to approximately two new schools per year.

Nationally, the growth of CMO schools is concentrated in a handful of states. About 80 percent of all CMO-run schools operate in Texas, California, Arizona, and Ohio. Most CMOs have located in states where the charter law offers moderate to high levels of autonomy to charter operators.

Figure 1. Growth of CMO Schools



Source: National Alliance for Public Charter Schools.

[1] These counts employ a broad definition of CMO—namely any nonprofit organization managing two or more schools. By contrast, our study focuses on the 40 CMOs that were managing four or more schools in 2007.

CMO schools are also concentrated in urban areas. About 74 percent of all CMO schools eligible for our study are located in cities. In some cities, such as New Orleans, Newark, Los Angeles, and Chicago, CMOs now represent a substantial fraction of all charter schools. Some CMOs gravitate to urban centers because they are large enough to allow for the creation of a concentrated network of schools, making it easier and more economical for CMO staff to support them. And many CMOs have sought to focus on assisting low-income students and have opened schools in neighborhoods where they hope to attract these students.

2. Relative to their host districts, CMOs serve a disproportionately large number of black, Hispanic, and low-income students but fewer special needs students

Compared to their host districts, the middle school student population served by the average CMO in our study includes a greater percentage of minority and low-income students (Table 1). In the average CMO, approximately 91 percent of middle school students are black or Hispanic compared to 76 percent of their host districts' middle school students. For the average CMO in our sample, approximately 71 percent of middle school students qualify for free or reduced-price lunch, compared to 64 percent of those served by their host districts. On the other hand, CMOs in this study serve a smaller share of students with disabilities and limited English proficiency.

Most CMOs attract students whose previous average test scores are similar to local averages but higher than local averages for black and Hispanic. Most CMOs—18 of 22 in our middle-school sample—enroll students whose prior achievement levels are within a quarter of a standard deviation of the overall average for their districts. But the black and Hispanic students entering most CMOs tend to have higher average test scores than their black and Hispanic peers in other public schools. In 13 of the 22 CMOs black students have significantly larger average math baseline test scores than the average for black students in the host district; only 1 of the 22 CMOs served black students with significantly lower average math test scores than their district peers. The patterns are similar for Hispanics and also similar for reading test scores.

Table 1 . CM Os se rve la rgely lo w- income, m inority stu dents, b ut E nglish- language l earners an d special education students are somewhat under- represented

	Percentage Black or Hispanic	Percentage Free or Reduced-Price Lunch	Percentage Limited English Proficient	Percentage with IEP
CMO Average	91%	71%	14%	9%
Host District Average	76%	64%	19%	13%

Source: State and district school records.

IEP = individualized education plan.

3. Per-pupil expenditures in CMOs vary widely, along with public charter funding

The CMOs in our sample spent between $5,000 and $20,000 per pupil each year. This variation appeared to be partly driven by the per-pupil funding amounts determined by state public charter school funding formulas: the correlation between per-pupil expenditure and per-pupil state funding is 0.61. In addition, philanthropy probably contributed to the variation across CMOs in per-pupil expenditures.

CMOs use their resources to support organizational structures and functions similar to those of school districts, but they vary widely in how they allocate their staff across various functions as well as in the ratio of central office staff to number of students. Some CMOs invest heavily in large central offices, while others maintain a minimal level of administrative staff. Many CMOs also allocate staff to support the creation of new schools.

4. On average, CMOs schools tend to be much smaller than schools in their host districts, with marginally lower student-teacher ratios

Because their organizational mission is often focused on personalized learning and a strong, intimate school culture, one might expect CMOs to have smaller school and class sizes. Indeed, CMO schools are much smaller than nearby schools in their districts. The CMOs in our study average 389 students per school compared to 982 students for nearby district schools. Class sizes and pupil-to-instructor ratios are also smaller in CMO schools than in their host districts. The average pupil-to-instructor ratios in math and reading are about 20.9 students per instructor; by contrast, in comparison schools the ratios are 23.5 in math and 23.2 in reading. (The difference between these CMO and district ratios is statistically significant in math but not in reading).

C. CMO Practices

1. CMOs are less likely than districts to prescribe a particular curriculum or student behavior policy, but CMO principals report more often than district principals that they implement a school-wide behavior strategy

In theory, CMOs could either provide principals with substantial discretion in selecting curricula and instructional materials or require all schools to adopt the same curriculum. We found that, relative to districts, CMOs are less likely to mandate a specific curriculum.

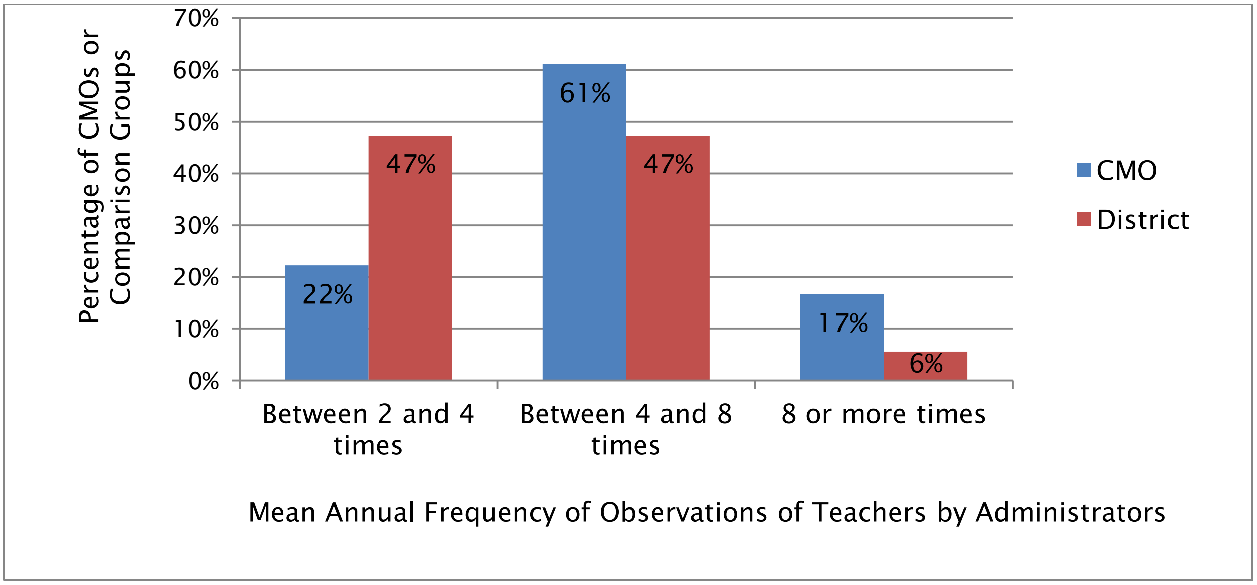
We also compared CMO schools and district schools on student behavior policy. CMO principals are significantly more likely than district principals to report that their schools have school-wide behavior standards (95 percent vs. 76 percent) and require students or parents to sign responsibility agreements (74 percent vs. 54 percent). But they also report that their schools have more flexibility than district principals in defining the details of behavior policies.

2. Relative to districts, CMO principals report that their teachers receive more coaching and that their teachers' pay is more likely to be based on performance

Some CMOs appear to pay considerable attention to the support and evaluation of teachers. Principal surveys suggest that CMO schools engage in more intense monitoring and coaching of teachers relative to district schools. On average, CMO principals report more frequent observations of teachers by administrators (see Figure 2), more frequent feedback to teachers from individuals observing them, and more frequent submission of lesson plans by teachers for review.

CMO schools are also more likely than district schools to use performance-based compensation. On average, 69 percent of CMO principals report using student test scores to evaluate teachers, compared to 46 percent of principals in nearby district schools. In addition, CMO principals report placing a higher priority on student test scores and teacher observations than on tenure and education in determining teachers' pay.

Figure 2. CMO Administrators Observe Teachers More Often

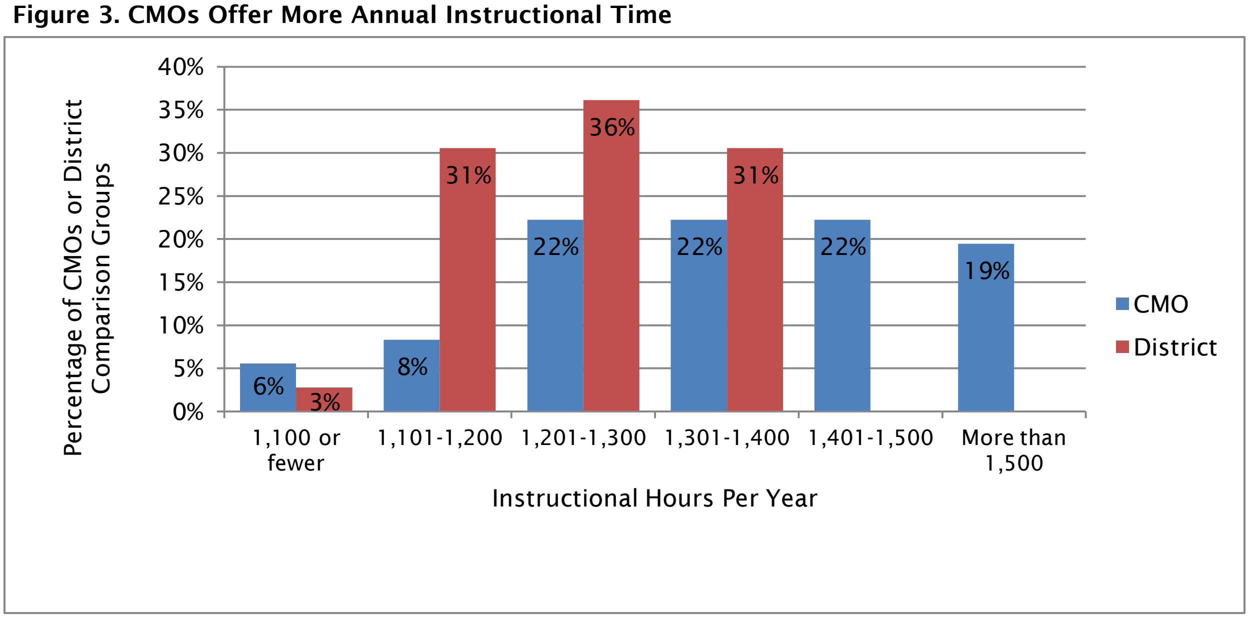


Source: Principal survey.

3. Relative to district principals, CMO principals report that their schools provide more instructional time

Some CMOs view increased instructional time as a key strategy for promoting achievement. Relative to district schools, CMOs tend to require significantly more time in school. The principal survey indicates that the average CMO provides 1,373 hours of instruction time per year compared to 1,239 hours in the nearby district schools. About 40 percent of CMOs report more than 1,400 instruction hours annually, and none of the district schools have this much (see Figure 3). The difference is driven largely by the length of the school day (rather than days in the school year), averaging 7.5 hours in the CMOs and 6.9 hours in district schools.

Figure 3. CMOs Offer More Annual Instructional Time



Source: Principal Survey.

4. CMOs can be categorized into four clusters that are differentiated largely by their emphasis on schoolwide behavior policies, teacher coaching, instructional time, formative assessment use, and performance based compensation

We used cluster analysis to categorize CMOs into four groups defined by their practices. One group ("Data Driven") is distinguished by its emphasis on performance-based compensation and use of formative assessment data. A second group ("Time on Task") places greatest emphasis on school-wide behavior policies and requires the most instructional hours for its students. Both of the first two groups also make extensive use of teacher coaching. A third set of CMOs ("Incremental Innovation") deviates the least from practices typical of conventional, district-operated schools. The fourth group ("Alternative Approach") is made up of a pair of CMOs that place the least emphasis on all of these practices.

D. CMO Impacts on Student Achievement and Attainment

To estimate the effects of CMOs on student achievement and attainment, we compared the outcomes of students enrolling in CMOs with those of matched comparison groups of students who resembled the CMO students in terms of baseline test scores and other key characteristics. Students who transferred out of CMO schools after the first year were kept in the CMO "treatment" group for the analysis. This ensures that impact estimates are not artificially inflated by the departure of low-achieving students. It also means that our impact estimates are conservative, in the sense that students who remain enrolled in CMOs for more than a year are likely to experience larger impacts than the ones we report here.

We were able to estimate middle school achievement impacts for 22 CMOs, high school graduation impacts for 6 CMOs, and college enrollment impacts for 4 CMOs. These were the only CMOs where the relevant school records data were available. CMO elementary schools could not be included in the impact analysis, because pre-kindergarten students do not typically have test scores, which are needed to match the CMO students to a valid comparison group.

We tested the validity of our propensity-score matching method in a subset of CMO schools where it was possible to conduct a randomized experiment—the "gold standard" of evaluation methodology. The matching approach successfully replicated experimental estimates of achievement impacts, thereby providing some confidence that it can produce valid impact estimates in the much larger number of CMO schools where it can be applied but where the experimental analysis is not possible. We report impacts in standard deviation units (also known as z-scores) to allow comparisons across grades and states that have different test scales.

1. Achievement impacts for individual CMOs are more often positive than negative, but vary substantially in both directions

In examining the distribution of CMO middle school achievement impacts, we focus on math and reading impacts two years after students enroll in a CMO schools. Two years is the longest period for which we can examine all 22 CMOs with available data.

The number of CMOs with significant positive middle school impacts is larger than the number with significant negative impacts. Two years after enrolling in a CMO school, students experience significantly positive math impacts in half of the CMOs (11 of 22) covered by the impact analysis, while students in about one-third of the CMOs (7 of 22) do significantly worse in math. Similarly, students in nearly half of the CMOs (10 of 22) experience significantly positive impacts in reading, while students in about a quarter of CMOs (6 of 22) experience reading impacts that are significantly negative. Table 2 shows that half of the CMOs (11 of 22) have significantly positive impacts in math or reading and nine have significantly negative impacts in one or both subjects. Moreover, 10 of the 22 CMOs have significantly positive impacts in both subjects while only four have significantly negative impacts in both subjects.

Table 2. Number of CMOs with Significant Positive and Negative Impacts in Math and Reading

	Math Impacts		
Reading Impacts	Significant Positive	Insignificant	Significant Negative
Significant Positive	10	0	0
Insignificant	1	2	3
Significant Negative	0	2	4

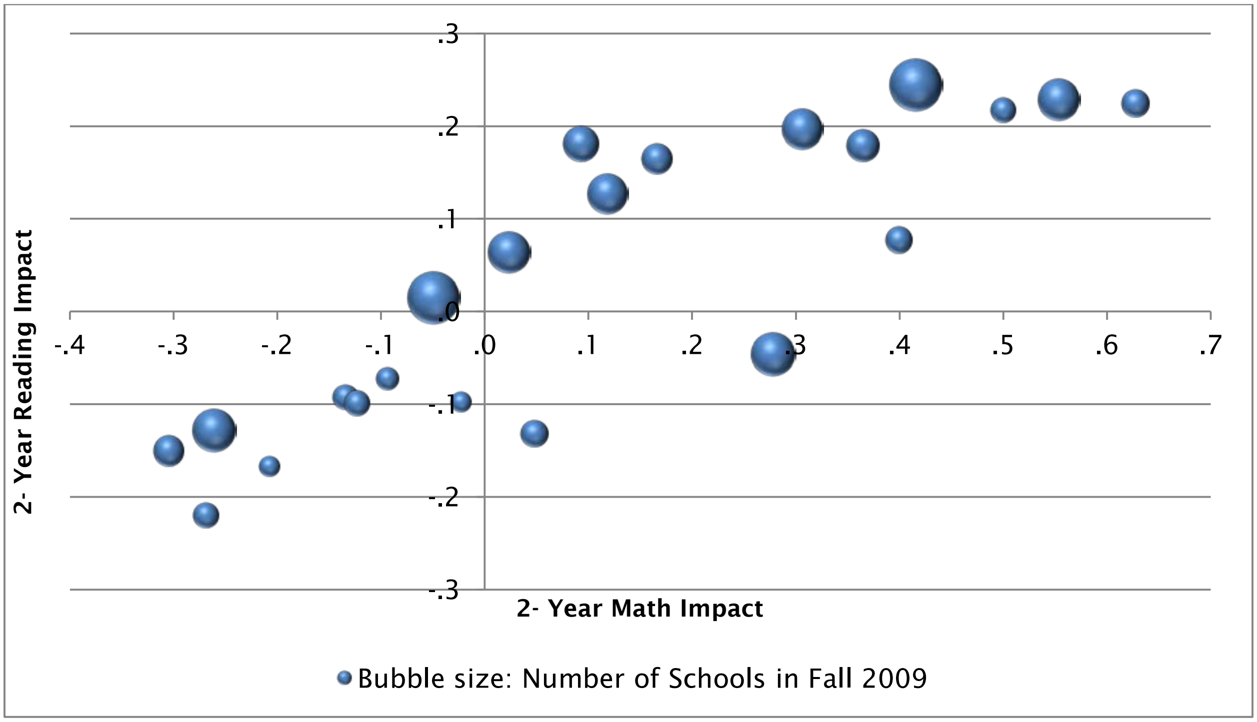
Source: State, district, and CMO school records.

There is also substantial variation in the magnitude of impacts. Figure 4 shows the distribution of estimated two-year math and reading impacts respectively on the x and y-axes. The size of the bubbles in Figure 4 represents the number of schools each CMO operated in fall 2009. Two years after CMO enrollment, math impacts range between -0.30 and 0.63 while reading impacts range between -0.22 and 0.24. The majority of two-year impacts are statistically significant (18 out of 22 for math; and 16 out of 22 for reading). At the extremes, the impacts are substantial: A few CMOs produce impacts that might generate three years of learning gains within two years of enrollment (Bloom et al. 2008). These numbers suggest that the CMOs at the high end of the scale have the potential to measurably reduce achievement gaps, especially in math. Meanwhile, the lowest performing CMOs are producing negative achievement effects that are nearly as large as the effect of a year of schooling—that is, their students achieve not much more than one year of learning after two years in the classroom.

As indicated in Figure 4, most of the larger CMOs in our sample have positive impacts, which might indicate that funders have had some success in supporting the expansion of CMOs that are more effective. Among the CMOs covered by the impact analysis, 8 of the 12 larger ones (those operating more than 8 elementary, middle, or high schools in 2009-10) have significant positive impacts in math or reading; by contrast, this is the case for only 3 of the 10 small CMOs (those operating 8 or fewer schools in 2009-10). As a result, the schools operated by these 22 CMOs in fall 2009 are concentrated in the CMOs with positive impacts. Approximately 57 percent of these schools are managed by CMOs with significant positive impacts in either math or reading, while only 26 percent are managed by CMOs with significant negative impacts in either subject; the remaining 17 percent are affiliated with CMOs with impacts in both subjects that are indistinguishable from those of nearby schools.[2] Despite this pattern, effectiveness is not related to size in a linear way: Correlations between size and impacts (and between growth rate and impacts) are not statistically significant.

[2] Ignoring statistical significance, approximately two thirds of these CMO schools are managed by CMOs with positive math impacts and the same fraction are managed by CMOs with positive reading impacts. Among the subset of CMO middle schools included in the impact analysis, the proportions of schools with positive and negative impacts are very similar to these percentages.

Figure 4. Test- Score Impacts in Math and Reading Vary Considerably Across CMOs



Source: State, district, and CMO school records.

Although the larger CMOs often have positive impacts, this does not mean that CMOs increase their performance as they grow. Within individual CMOs, some show declining impacts as they add schools, while others do not. In math, there is no clear pattern of changes in impacts as CMOs grow, but in reading, the impacts of most CMOs declined as they grew.

2. **Test score impact estimates for the average CMO are positive in all four subjects, but they are not statistically significant**

We estimated impacts of the average CMO in reading, math, science, and social studies, one to three years after a student's initial enrollment in the CMO school (though science and social studies test scores were available only for a subset of CMOs and years). Average CMO impacts are positive in all cases but one (one-year reading impact), but they are not statistically significant (at the .05 level), despite reaching a non-trivial magnitude in math by the third year after enrollment (Table 3). Our statistical power to detect an average impact across CMOs is limited by the fact that only 22 CMOs are included in the analysis (fewer for science and social studies). If we were to estimate the effect of these CMOs on the average student (rather than the average CMO impact), thereby giving the larger CMOs more weight, the math impact would be significantly positive.

Table 3. Average CMO Test Score Impacts, by Year After CMO Enrollment

	1-Year Impact	2-Year Impact	3-Year Impact
Math Effect Size	0.06	0.11^	0.15
(Standard Error)	(0.05)	(0.06)	(0.09)
Number of CMOs	22	22	14
Reading Effect Size	-0.01	0.03	0.05
(Standard Error)	(0.02)	(0.03)	(0.04)
Number of CMOs	22	22	20
Science Effect Size	N.A.	N.A.	0.06
(Standard Error)			(0.09)
Number of CMOs			11
Social Studies Effect Size	N.A.	N.A.	0.09
(Standard Error)			(0.09)
Number of CMOs			9

Source: State, district, and CMO school records.

^Significantly different from zero at the .10 level, two-tailed test.

3. The variation in school-level impacts is mostly due to differences between rather than within CMOs, indicating that some CMOs are systematically outperforming others

One of the primary aims of CMOs is to promote consistent results across multiple schools. Our analysis of middle school achievement impacts sheds light on the extent to which the CMOs are achieving this objective. Specifically, for each middle school in a given CMO, we estimated impacts separately on students' reading and math achievement two years after enrollment. We then calculated the between-CMO and within-CMO variation in school-level impacts.

Most of the variation in school-level impacts occurs between rather than within CMOs. We found that CMO-level impacts account for 87 percent of the variation in school-level impacts in math and 73 percent in reading.

Several CMOs produce consistently positive school-by-school results. Among the 18 CMOs with impact estimates for two or more schools, in math 7 have uniformly positive school-level impacts, 5 have uniformly negative impacts, and 6 have mixed effects. In reading, 7 CMOs have uniformly positive reading effects, 1 has uniformly negative impacts, and 10 have mixed results.

4. CMOs that produce positive impacts in one middle school subject tend to produce positive impacts in other subjects, including science, social studies, reading, and math

The upward-sloping diagonal pattern of the results in Figure 4 shows that CMOs' math and reading impacts are positively correlated. Similar patterns are evident for science and social studies. Within CMOs, impacts are highly correlated across the four subjects. Science and social studies impacts, like reading and math impacts, vary substantially across CMOs. But we found no evidence that CMOs are focusing on some subjects at the expense of others:

5. In several CMOs, middle school math and reading test score impacts are larger for Hispanic students than for other students

Prompted by prior studies of charter schools (Angrist et al. 2011; Gleason et al. 2010) that have found suggestive evidence of greater benefits for low-income minority students in urban areas, we compared two-year math and reading impacts for particular subgroups of students. We found some evidence of larger two-year math and reading impacts for Hispanic students in the nine CMOs for which sufficient data was available. Other subgroups—including African Americans and groups defined by gender, income, and prior academic achievement—do not show clear patterns of differential positive or negative impacts.

6. Among the CMOs where high school graduation and college enrollment data are available, impacts are more often positive than negative, but average impacts are not statistically significant.

While most CMOs place a high priority on improving student achievement, most also seek to increase students' educational attainment. Nine out of 10 CMO principals say that preparing students for college is very important. In addition to examining CMOs' impacts on middle school test scores, we also estimated the longer-term impacts of CMO high schools on graduation and postsecondary enrollment. While data on these attainment outcomes were available only in a small number of CMOs, we still found substantial variation in impacts.

Among the six CMOs with graduation data, three had positive and statistically significant effects, and only one had a negative impact on graduation. In the remaining two cases, impacts were positive but not significant. The range of impacts is also quite substantial: one CMO increased the probability that a student would graduate from high school in four years by 23 percentage points while another reduced this probability by 22 percentage points (Table 4). On average, these six CMOs increased students' probability of graduating by 7 percentage points, but this average impact is statistically insignificant (p=0.35).

Table 4. CMO Impacts on High School Graduation and Postsecondary Enrollment (In Percentage Points)

	High School Graduation	Postsecondary Enrollment
CMO #1	23%**	23%**
CMO #2	17%**	21%**
CMO #3	12%*	3%
CMO #4	8%	4%
CMO #5	3%	N/A
CMO #6	-22%**	N/A
CMO Average	**7%**	**13%**

Source: State and district school records.

*Significantly different from zero at the .05 level, two-tailed test.
**Significantly different from zero at the .01 level, two-tailed test.

The two CMOs with the largest impacts on high school graduation also had significant, positive impacts on postsecondary enrollment (Table 4). These two CMOs appear to increase postsecondary enrollment by 21 percentage points and 23 percentage points, respectively. Two other CMOs had insignificant impacts on postsecondary enrollment; one of these two had a significant positive impact on high school graduation while the other had an insignificant positive impact on graduation.. (No college enrollment data were available for either the CMO that had large negative impacts on high school graduation or the remaining CMO with an insignificant positive effect on graduation.) On average these four CMOs had a positive impact on college entry of 13 percentage points, but this effect is statistically insignificant (p=0.10).

E. Practices Associated with Positive Impacts

Understanding which CMO practices are associated with the largest impacts can help identify potentially promising educational strategies. To be sure, the associations we observed between impacts and specific CMO practices might not indicate a causal effect of the practices. It is possible that a practice that is positively associated with impacts may in fact be correlated with some other practices we do not observe that are the real driver of student outcomes. But examining associations of practices with impacts is the necessary first step toward identifying promising practices.

1. Comprehensive behavior policies in schools are associated with larger CMO impacts

Student impacts in math and reading are larger in CMOs whose schools have comprehensive behavior policies. We found positive associations between student impacts and multiple measures of school behavior policies: consistent behavior standards and disciplinary policies within a school, zero tolerance policies for potentially dangerous behaviors, behavior codes with student rewards and sanctions, and responsibility agreements signed by students or parents.

2. CMOs with intensive coaching of teachers tend to have larger positive impacts on student achievement

Student impacts in math and reading are larger in CMOs with schools that place a greater emphasis on intensive coaching of new teachers. Impacts are associated with a composite measure of teacher coaching that captures the frequency with which teachers are observed and the frequency with which they receive feedback on their performance and their lesson plans. In addition impacts are larger in those CMOs providing substantial professional development support to their schools.

3. Several other notable CMO characteristics do not show significant relationships with impacts

We found no significant relationship between impacts and three other factors that we posited might contribute to student achievement. Specifically, impacts are not correlated with (1) the extent to which CMOs define a consistent educational approach through the selection of curricula and instructional materials, (2) performance-based teacher compensation, or (3) frequent formative student assessments (although impacts are larger when teachers frequently *use* student test results to modify lesson plans). Nor are impacts significantly associated with school or class sizes.

Math impacts are positively correlated with more hours of annual instruction, but this relationship appears to be largely due to the association of instructional time with behavior policies and coaching. We ran multivariate regressions of impacts on key practices that were significantly associated with impacts in bivariate regressions. In the multivariate regressions, the association between impacts and instructional time declined substantially and became statistically insignificant.

F. Questions for Future Research

As is often the case in studies of this kind, some of the interesting findings raise other important questions for future research on CMOs:

- **To what extent do CMOs produce positive effects on longer term student outcomes?** Most of our impact analysis focused on how CMOs affect academic achievement, although this report also includes findings related to high school graduation and postsecondary enrollment impacts in a handful of CMOs. More research is needed on how CMOs affect postsecondary enrollment, degree completion, civic behavior, and earnings.
- **What explains why some CMOs have negative impacts on test scores?** Over 40 percent of the CMOs covered by our impact analysis are falling short of the performance of nearby district schools in math or reading. This raises questions about whether some CMOs might be scaling up the wrong models or are attempting to scale up a promising model but have had difficulty replicating it. It is also possible that these CMOs are focused on student outcomes other than test scores.
- **Which promising strategies should CMOs implement and how should they implement them?** Our forthcoming report on promising practices of CMOs will explore this issue in more depth, drawing on both our surveys and qualitative research.
- **To what extent do CMOs add value compared to independent charter schools?** Whether CMOs can take advantage of scale without losing the flexibility associated with charter status is a key question.
- **Are new CMOs using the same strategies and producing the same impacts as more established CMOs?** Because our study began four years ago, we focused on CMOs operating multiple schools in 2007. It is possible that newer CMOs are either more or less effective than the CMOs we examined.
- **What other factors might contribute to CMO impacts?** Among the other important hypotheses that we could not examine in detail are ways in which impacts might be related to high expectations in the classroom, funding levels, peer effects, grade configuration, and specific approaches to classroom instruction.

I. INTRODUCTION

A. Policy Context and Rationale for CMOs

Charter schools—public schools of choice that are operated autonomously, outside the direct control of local school districts—have become more prevalent over the past two decades. Several thousand across the country are educating students, and federal policy has supported their growth over three successive presidential administrations. Research studies on charter schools have multiplied along with the schools themselves, and an increasing number of studies, using a variety of methods, have attempted to measure charter schools' effects on student achievement. Debates over the achievement impacts of charter schools have been heated, but no clear consensus has emerged about whether, on average, charter schools across the country are doing better or worse than conventional public schools at promoting student achievement. Nonetheless, one finding of the research on charter school impacts is clear: Effects vary widely among different charter schools.

Large variation in the performance of charter schools follows directly from the design of the policy itself. A major purpose of charter school laws is to encourage the creation of varied educational models. The question for current and future funders and operators is how to identify and replicate the more successful charter schools and their effective practices.

Charter school management organizations (CMOs)—which, as the name implies, operate multiple charter schools and create new schools under a common structure and philosophy—represent one prominent attempt to leverage the success of high-performing charter schools. Many CMOs were created to replicate educational approaches that appeared to be effective, particularly for disadvantaged students, in a small number of charter or other schools. Attracting substantial philanthropic support, CMO schools have grown rapidly from encompassing about 6 percent of all charter schools in 2000 to about 17 percent of a much larger number of charter schools by 2009 (Miron 2010). Some of these organizations have received laudatory attention following reports of dramatic achievement results for disadvantaged students. Many of these reports, however, have relied on incomplete evidence.

The National Study of CMO Effectiveness aims to fill the gap in evidence about CMOs, providing the first rigorous nationwide examination of the effects of CMO schools on student achievement. To provide information on the variability of CMO effectiveness, the study estimates achievement effects separately for each CMO for which data were available. In addition, we seek to understand the relationships between CMO practices and their effects on student achievement; the aim is to inform the field by identifying promising practices associated with positive impacts. Mathematica Policy Research and the Center on Reinventing Public Education (CRPE) have conducted the study with funding from the Bill & Melinda Gates Foundation and the Walton Family Foundation and with project management assistance from the NewSchools Venture Fund.

This updated edition of the report provides key findings on CMO practices, impacts, and the relationships between them. A future report will explore promising practices in greater depth. The rest of this introduction describes (1) the research questions, the CMOs in the study, and the data sources, and (2) the organization of this report.

B. Research Questions, CMOs in Study, and Data Sources

The National Study of CMO Effectiveness is designed to describe CMOs in existence in 2007, assess their impacts on students, and identify practices associated with positive impacts. Drawing on multiple data sources, the study focuses on three sets of research questions:

1. **Characteristics and Context.** How quickly are CMOs growing? Which students and areas do they serve, and what resources do they use? What are the practices and structure of CMOs? What district policies and other contextual factors appear to influence the location and growth of CMOs?
2. **Impacts.** What are the impacts of CMOs on student achievement and attainment and to what extent do these impacts vary across CMOs?
3. **Promising Practices.** Which CMO practices and structures are positively associated with impacts?

1. CMOs eligible for this study

Previous studies have defined CMOs in various ways, and these differing definitions have influenced which organizations the studies covered. Some researchers include organizations that do not directly manage charter schools but that provide a brand name, support network, and other academic services; other researchers include organizations whose services are limited to administrative functions such as payroll. Although some studies focus only on not-for-profit organizations, others have also examined for-profit organizations. Studies also vary in terms of the number of schools that must be managed by a CMO (Miron et al, 2010).

Our study focuses on nonprofit CMOs that directly control four or more schools. More specifically, to be eligible for the study, an organization had to satisfy four criteria: it had to (1) have four schools open by fall 2007, (2) be nonprofit since inception, (3) not primarily serve dropouts or similar special populations, and (4) directly manage schools (that is, hire and fire school principals). This definition excludes for-profit organizations, charter networks that lack direct operational authority, and organizations that provide only back-office services.

When the study began, we identified 40 CMOs containing 292 schools that satisfy the four criteria. As discussed in Chapter II, these CMOs schools are located in 14 states, with the largest concentrations in Texas, California, Arizona, Ohio, Illinois, New York, and the District of Columbia.

2. Study data sources

To examine eligible CMOs and address the research questions, the study makes use of the following data sources:

- **Survey of CMO Central Office Staff**. Conducted in 2009, this web-based survey was completed by managers of 37 CMOs. The survey included questions on the role of CMO staff in managing schools, the number of schools operated by the CMO, CMO growth goals, the composition of CMO and school staff, educational priorities, teacher compensation and evaluation, and staff development, student behavior, and evaluation policies.

- **Survey of CMO and District Principals.** The sample of this web-based survey, conducted in 2010, comprised all the principals of 292 CMO schools open by fall 2007 and an equal number of comparison principals in traditional public schools. The comparison schools were the closest district schools with a similar mix of students. Approximately 70 percent of the principals responded to this survey. The survey included questions on the school's educational priorities, curriculum, behavioral policies, staff development and evaluation, and student assessments as well as on the role of the CMO or district central office in managing and supporting the school.

- **Survey of CMO Teachers.** The sample for this web-based survey, conducted in 2010, comprised randomly selected teachers in each of the CMO schools. We selected two teachers per grade for each elementary school and one English and math teacher per grade for each middle and high school. Approximately 76 percent of the teachers responded to the survey. The survey included questions on participation in staff development and curriculum planning activities, observation and evaluation of the teacher, job satisfaction and perceptions of the work environment, and the teacher's training and background.

- **Student School Records.** For the impact analysis and description of student characteristics, we collected administrative records from states, districts, and CMOs. The key outcome variables collected were individual student test scores and, where available, high school graduation and postsecondary enrollment. We also requested various background and demographic characteristics including race, gender, and baseline Individual Education Program (IEP) and English language learner (ELL) status.

- **Site Visits.** During 2009 the study team visited 10 CMOs and 20 CMO schools. We selected the CMOs for diversity in terms of their theory of action, size, and region of the country. The visits included semi-structured interviews with the senior leadership team of the CMOs, two principals, and selected teachers as well as brief classroom observations. The research team collected information on the CMO theory of action, structural choices, growth strategies and challenges, alignment of CMO and school priorities, interactions with authorizers and districts, and teachers' classroom management approach.

- **CMO Financial Records and Business Plans.** The team collected the Form 990 tax statements for each of the 40 eligible CMOs to estimate per student expenditures. For 17 of the CMOs, researchers also examined business plans to secure more detail on CMO strategies and financial projections.

C. This Report and Other Study Reports

This report addresses each of the three sets of study questions. The description of CMO characteristics and practices draws primarily from the surveys and school records but also is informed by the site visits. The impact analysis focuses on the impact of CMO middle schools on academic achievement and the impact of CMO high schools on high school graduation and college enrollment.[1] The middle school achievement analysis covers 22 CMOs, the high school graduation

[1]A previous version of this report released in November 2011 contained the CMO-level middle school impacts; this version adds findings on CMO high school impacts as well as the school-level middle school impacts.

analysis covers 6 CMOs, and the college enrollment analysis covers 4 CMOs. In addition to providing information on the distribution of impacts for these CMOs, we report the extent to which the middle school impacts are correlated with specific CMO practices and structures. A subsequent report will describe CMO promising practices in greater detail.

This report also draws upon the study's interim report. In 2010 CRPE completed an interim study report which describes the practices, challenges, and strategies of CMOs (Lake et. al, 2010). Drawing on the survey of CMO headquarters staff and the site visits to 10 CMOs, the report described the diverse educational strategies employed by CMOs, the challenges they encounter as they expand, and ways they differ from, and are perceived by, school districts.

The rest of this report is organized as follows:

- **Chapter II: CMO Growth, Students, and Resources.** This chapter reviews recent growth of CMOs, the geographic concentration of CMOs in specific states and districts, the mix of students served by CMOs, CMO central office staffing information, CMO student-instructor ratios, and per pupil expenditures of CMOs.
- **Chapter III: Practices of CMOs and Their Schools**. Drawing on the three surveys of CMO headquarter staff, principals, and teachers, this chapter describes the practices and policies of CMOs and their schools and the extent to which they differ from district practices and policies, and ways in which CMOs bundle multiple practices. It also examines the relationship between CMO practices and school-level instructional coherence and organizational health measures.
- **Chapter IV: Impacts on Student Achievement and Attainment**. This chapter presents estimates of the impacts of individual CMOs on middle school student achievement, high school graduation, and college enrollment,. We examine the range of positive and negative impacts as well as the overall average impact. In addition, the chapter examines the middle school impacts for specific subgroups defined by race, ethnicity, and other background characteristics.
- **Chapter V: Promising Practices**. Combining data from surveys and impact results, this chapter examines which CMO practices are associated with middle school student impacts.
- **Chapter VI: Questions for Future Research.** This chapter identifies some policy-relevant questions that could be examined in the future.

II. CMO GROWTH, STUDENTS, AND RESOURCES

Key Findings

- CMOs account for approximately 16 percent of all charter schools, with a higher percentage in urban areas. The CMO share of the charter sector has grown steadily over the past 15 years, although growth appears to have slowed recently.
- Relative to their host districts, CMOs serve disproportionately large numbers of students who are black, Hispanic, and low-income, but somewhat fewer special education students and English-language learners.
- Students entering CMO middle schools typically have prior achievement levels that are similar to the local average and somewhat higher than the local average for black and Hispanic students.
- CMOs vary widely in per pupil spending, which reflects, in part, variation in state and district funding.
- CMO schools tend to be smaller than schools in their host districts, with marginally lower student-teacher ratios.

A. Introduction

CMOs operate in the context of state and local laws, local economies, and funders' priorities. CMOs can choose where they locate based in part on whether state and charter laws accommodate nonprofit organizations seeking to create networks of effective charter schools. In making location decisions, CMOs also need to consider whether they can reach their target population of students. Thus the location of CMOs can influence and be influenced by the mix of students they serve. In addition, CMOs are funded by federal, state, and local education agencies as well as philanthropies. The amount of funding they attract affects the resources they can deploy in their central offices, schools, and classrooms. Examining variations among CMOs along these and other dimensions can shed light on the role and characteristics of CMOs as well as the findings presented in other chapters of this report.

B. Number and Geographic Distribution of CMOs

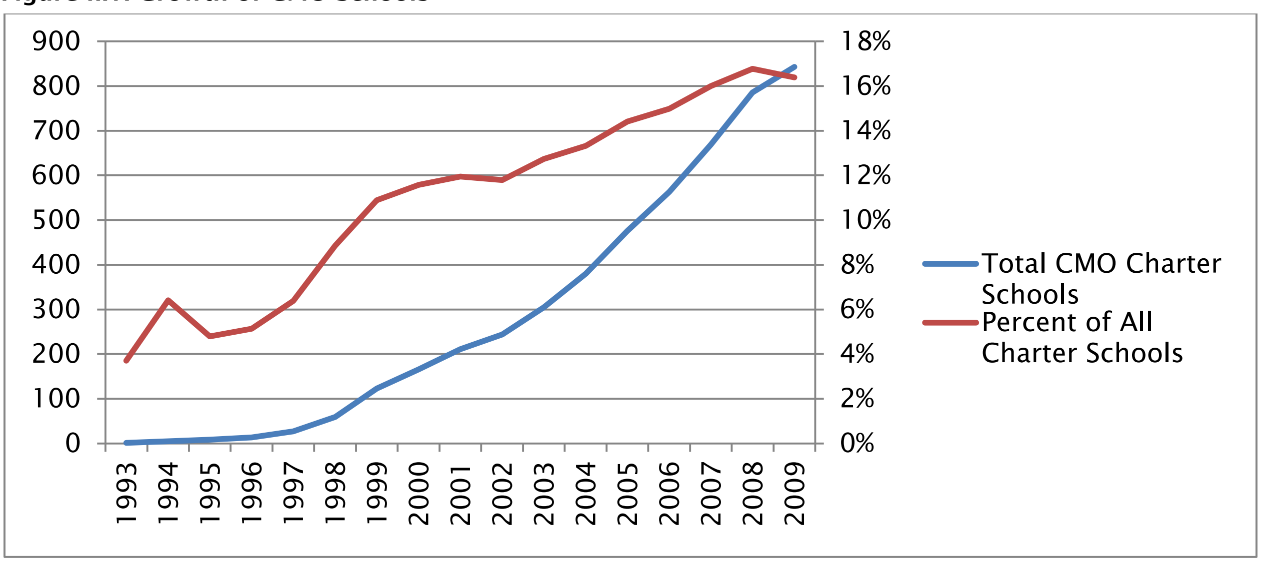
This section presents nationwide data on the overall CMO landscape, including the number of CMOs nationally, where they are located, how fast they are growing, and how their growth trajectory compares to charter schools that are not operated by CMOs (as a sector). For those questions, we focus on the broadest population of CMOs, employing the definition of CMOs identified by the National Alliance for Public Charter Schools (NAPCS) and by the study team's additional investigation (Miron et al. 2010). This all-inclusive identification of CMOs includes approximately 130 nonprofit organizations that manage more than one charter school. Note that other sections of this chapter and the rest of this report employ a narrower definition of CMOs that includes only organizations managing four charter schools as of fall 2007 and excluding organizations that had once operated as for-profit organizations, serve only dropouts, or that do not have direct operational control over their schools. We contrast the broader set of 130 CMOs with the 40 CMOs that satisfy the narrower definition of CMO used in our study. For some purposes we rely on data for students entering CMO middle schools, for which we are able to systematically compare the baseline (i.e., pre-CMO entry) characteristics of students with those of students in the same grades and districts. This is useful, for example, in assessing the prior achievement levels of students entering CMOs.

1. CMOs represent a growing presence in the charter landscape

Using the broadest definition of CMOs, there were more than 130 CMOs operating 843 schools as of the 2009–2010 school year.[1] Nationwide, these CMOs served nearly 250,000 students, representing approximately 15 percent of all charter school students and 0.5 percent of total public school enrollment. CMO-run schools constitute approximately 16 percent of the charter schools operating nationally in 2009–2010, compared to just 12 percent in 2000 (see Figure II.1).

The number of new CMO schools has grown rapidly alongside the number of CMOs. In 1993, the broad universe of CMOs accounted for just 5 percent of all new charter schools (see Figure II.2). At their peak growth in 2007, CMO-run schools represented 25 percent of all new charter schools. The rapid growth in CMOs and affiliated schools partly reflected the substantial philanthropic investments in CMOs. These donations totaled more than $500 million since 2000 (Lake et al., 2010).

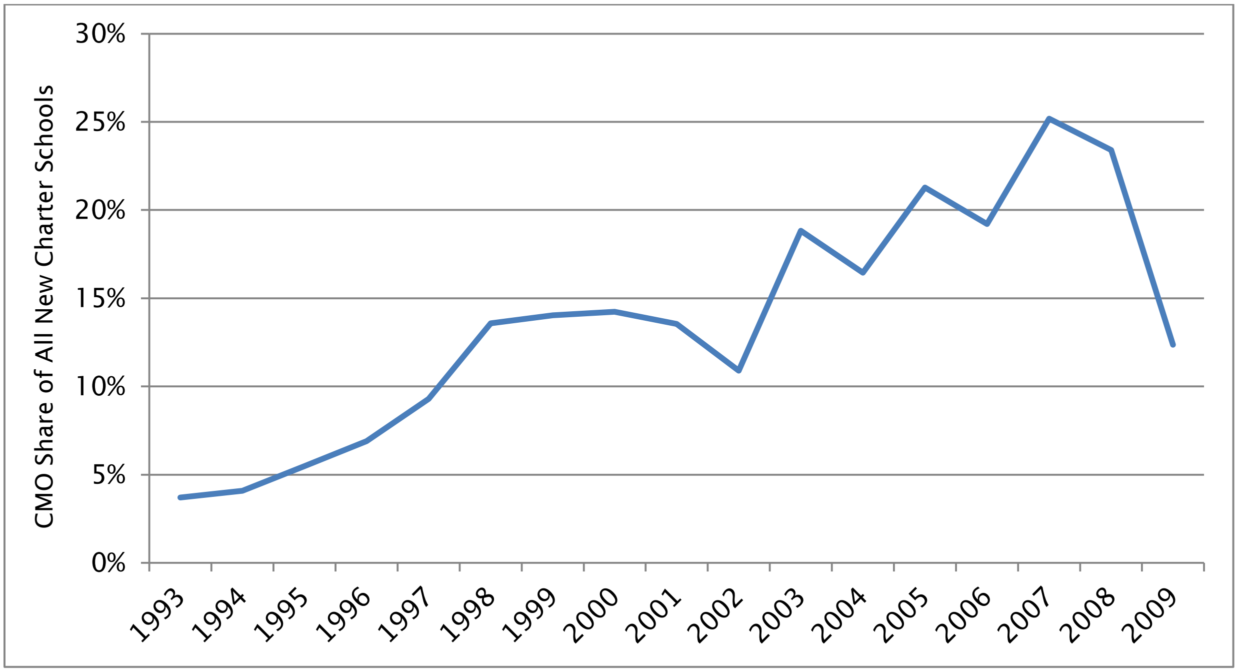
Figure II.1. Growth of CMO Schools



Source: Analysis of data from the National Alliance for Public Charter Schools.

[1]Again, of the 130 organizations satisfying this broad definition of CMO (non-profits operating more than one charter school), 40 satisfied the specific criteria for inclusion in this study.

Figure II.2. Share of New Charter Schools That Were CMO- Operated, by Year, 1993- 2009



Source: Analysis of data from the National Alliance for Public Charter Schools.

By 2009, the rate of growth in CMO schools appears to have slowed. In that year, CMOs were responsible for only 10 percent of new charter schools. It is too early to say whether this represents a trend. Assuming CMO growth has slowed, there are numerous possible explanations including (but not limited to) reduced philanthropic investments, effort to slow growth to enhance quality, reluctance by authorizers to issue charters to CMOs, and lack of facilities.

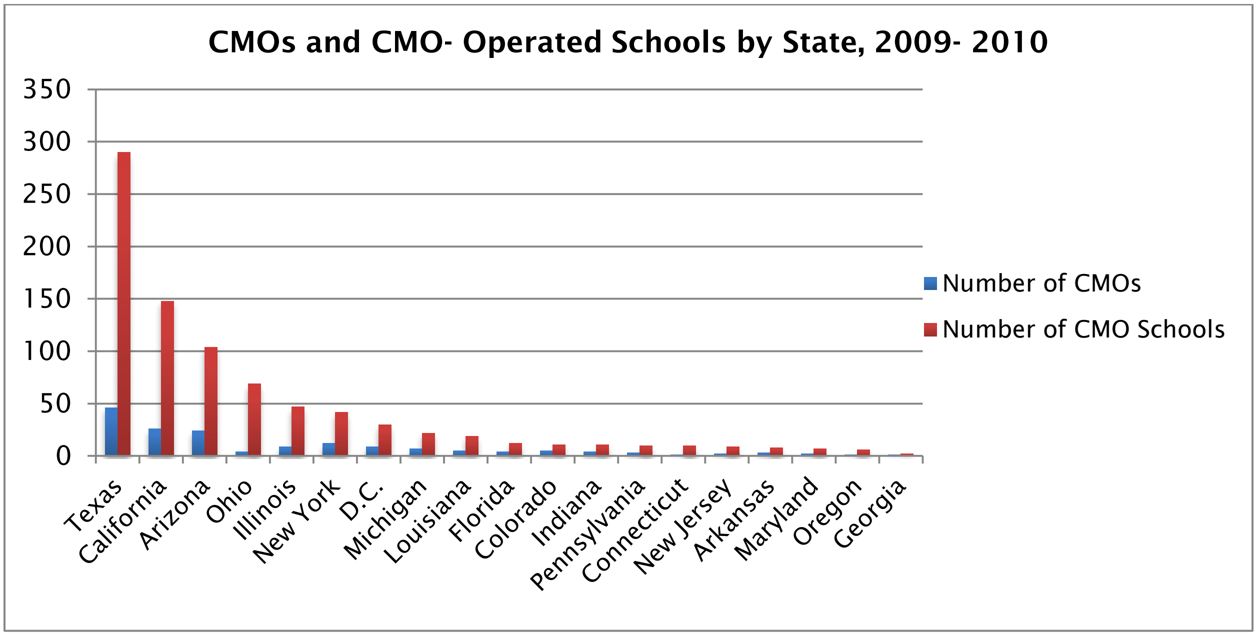
2. CMO presence is concentrated in states with growth-friendly charter laws and in several urban areas

Nationally, charter school growth is concentrated in some states, such as California and Arizona. The growth of CMO schools is even more concentrated in many of the same states. As Figure II.3 shows, 80 percent of all CMO-run schools (again using the broadest definition of CMO) operate in Texas, California, Arizona, and Ohio compared to 40 percent of charter schools overall. More than half the states that have charter schools do not have any CMO-operated schools.

An important aspect of the policy context that may influence where CMOs choose to open and operate schools is the amount of autonomy granted by state law. Greater autonomy gives CMOs more freedom to operate their schools in the manner they see fit, and makes it easier to expand the number of schools they operate.

Indeed, most of the states attracting substantial numbers of CMOs provide at least a moderate amount of autonomy to charter schools (Table II.1). Using indicators developed by the NAPCS, we developed an overall measure of charter autonomy for each state with a charter school law. The indicators measure the degree to which state law provides for: (1) fiscally and legally autonomous schools with independent charter school boards, (2) automatic exemptions from many state and district laws and regulations, (3) automatic collective bargaining exemptions, and (4) the ability to operate more than one campus under one charter contract and/or board. These are factors that the

Figure II.3. Distribution of CMO Home Offices and CMO- Operated Schools by State, 2009



Source: Analysis of data from the National Alliance for Public Charter Schools.

Table II.1. CMO Location by Amount of Autonomy Offered in State Law

Autonomy Category	'Autonomy Index' Score	Number of States in Category*	States in Each Category*	States with CMOs	Number of CMOs Operating in All States in Each Category
High	12-16	10	AZ, CA, DC, DE, FL, ID, NH, OK, OR, UT	AZ, CA, DC, FL, OR	64 (in 5 states)
Medium	8-11	18	AR, CO, CT, GA, IL, LA, MA, MI, MN, MO, NC, NJ, NM, NY, PA, SC, TN, TX	CO, CT, GA, IL, LA, NJ, NY, PA, TX	83 (in 9 states)
Low	0-7	12	AL, HI, IA, IN, KS, MD, NV, OH, RI, VA, WI, WY	IN, OH, MD	11 (in 3 states)
None/no charter law	n.a.	11	AL, KY, ME, MI, MT, NE, ND, SD, VT, WA, WV	0	0

Source: Analysis of data from National Alliance for Public Charter Schools

n.a. = not applicable.

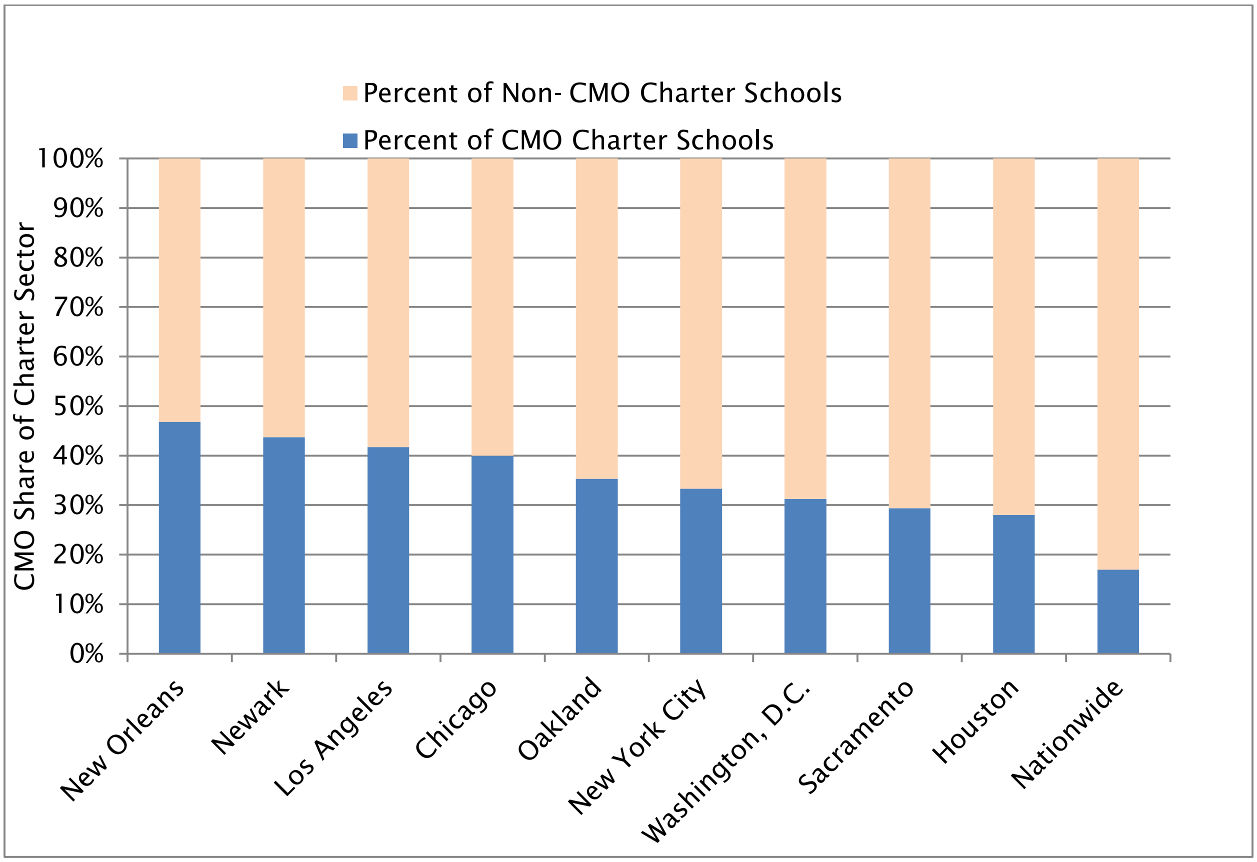
CMO executives we interviewed said influence their location decisions.[2] Ohio is one of the few states with low levels of autonomy attracting many CMOs.

[2] States were scored on a scale of 0 (low) to 4 (high) on each of these indicators, and we weighted the indicators equally to create an overall 'autonomy index' with possible scores ranging from 0 to 16. We then created four autonomy

CMO schools are also concentrated in urban areas in general and certain cities in particular. Seventy-four percent of the schools of the 40 CMOs eligible for our study are located in cities (defined as small, mid-sized, or large cities by U.S. Census). CMOs represent a large fraction of the charter schools in New Orleans, Newark, Los Angeles, Chicago, Oakland, New York City, the District of Columbia, Sacramento, and Houston (see Figure II.4.). In the cities most saturated with charter schools, CMOs have significant presence, typically representing 30 to 50 percent of all charter schools.

This pattern may partly reflect CMOs' effort to develop dense networks of schools that can be easily supported by CMO staff. Several CMO leaders reported that they have sought to restrict their growth to certain metropolitan areas in order to capture local economies of scale. These staff said they seek to support growth and maintain quality across their network through close monitoring and active support of their schools (Lake et al. 2010).

Figure II.4. CMOs' Market Share in Larger Charter Markets



Source: Charter statistics provided by National Alliance for Public Charter Schools; Non- CMO charter data collected by Center on Reinventing Public Education from state and charter agencies.

(continued)

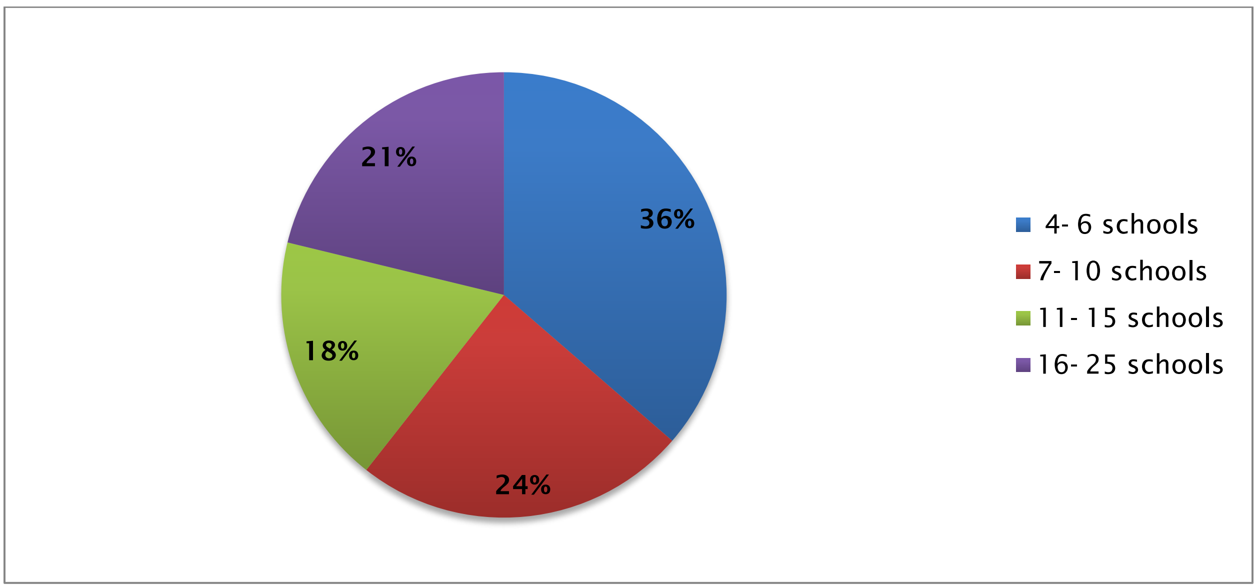
categories corresponding to four equal ranges of the index scores. We collapsed the two lowest autonomy categories into one as states in the lowest two 'low autonomy' categories have managed to attract fewer than a dozen CMOs.

3. Relative to the urban districts in which most CMOs are situated and their own growth goals, most CMOs are small organizations

Some CMOs have encountered challenges when they sought to expand. These challenges include recruiting staff with the skills needed to implement or adapt the CMO's educational approach, maintaining consistent quality across many schools, and avoiding rigid bureaucracy. These issues are discussed in the study's interim report (Lake et al. 2010), which noted that the types of challenges CMOs encounter can change as it grows.

Most CMOs oversee a modest number of schools relative to the districts in which they operate. Among the CMOs eligible for our study (all of which manage at least 4 schools), about 57 percent manage 10 or fewer schools (Figure II.5); in the broader universe of CMOs (those managing two or more schools), about 87 percent of CMOs manage 10 or fewer schools (Figure II.6). On one hand, the average CMO in our study is larger than 87 percent of all school districts in the country in terms of number of schools. However, compared to the urban school districts where they are typically located, they are much smaller, with the average urban district overseeing 32 schools.[3] Moreover, staff at most of the CMOs in our study say they would like to grow, hoping to add up to 20 schools each by 2025 (Lake et al. 2010). By those standards, most CMOs should still be considered in their early development stage, with the average CMO just entering the size range (beyond 8–10 schools) that can cause significant expansion challenges (Lake et al. 2010).

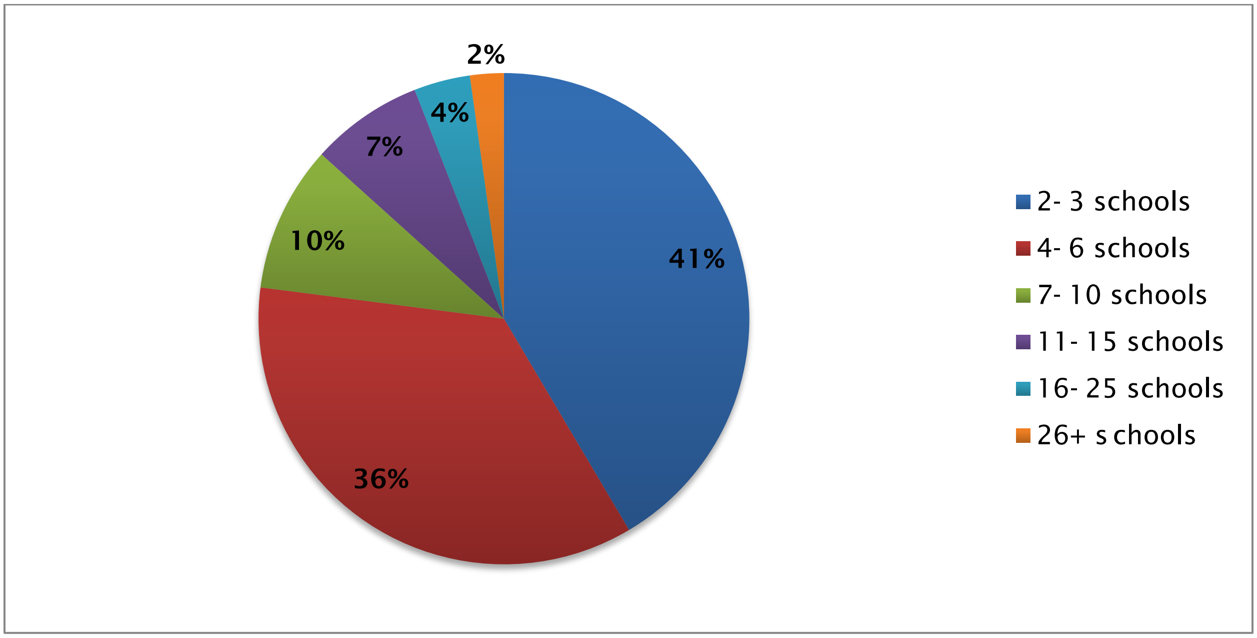
Figure II.5. Number of Schools Managed by CMOs Eligible for This Study, 2009



Source: CMO Central Office Staff Survey.

[3] We define an urban district as any district that is classified as either a small, mid-sized, or large city by the U.S. Census and serves four or more schools.

Figure II.6. Number of Schools Managed by CMOs in Broader Universe, 2009



Source: Analysis of data from National Alliance for Public Charter Schools.

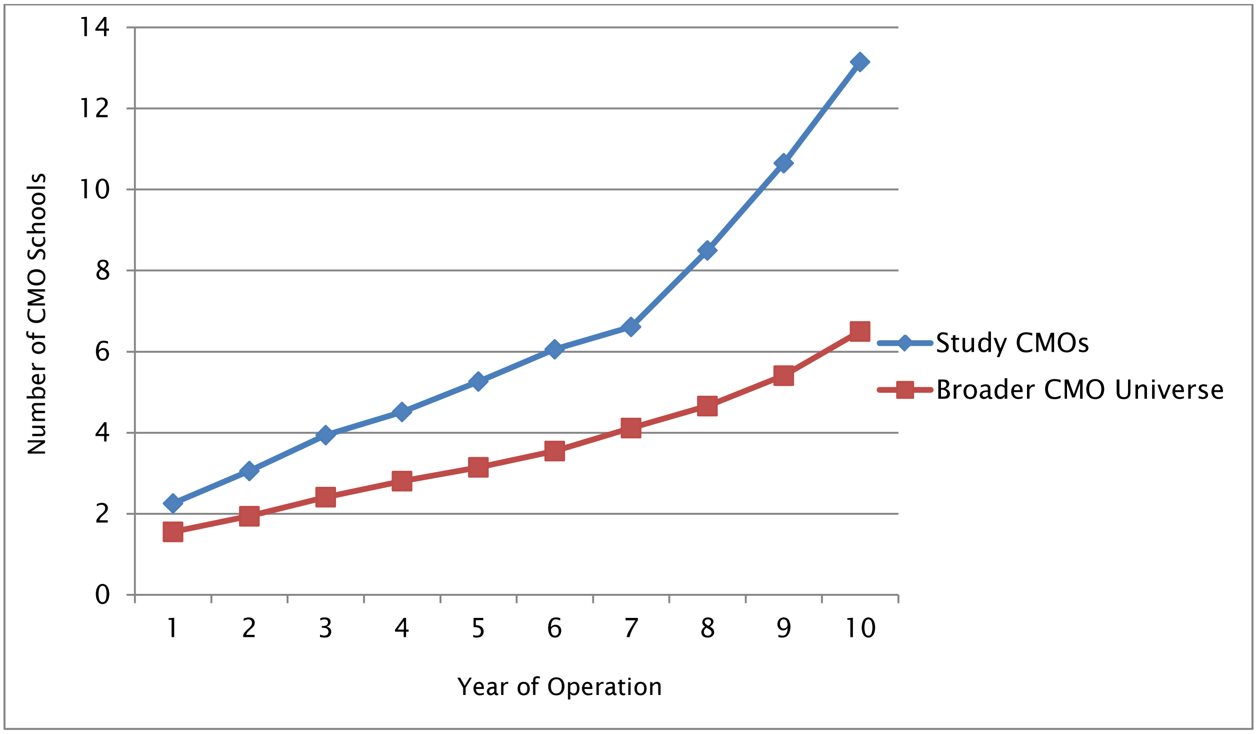
The growth trajectory of individual CMOs varies dramatically according to their mission and capacity. Some pursue aggressive growth, aiming to open 3–5 new schools per year. Others aim for much slower expansion. On average, CMOs in our study are opening no more than one new school per year for the first six years (see Figure II.7). After seven years of operation, the average pace picks up to approximately two new schools per year. This pace of expansion is faster than that of the broader universe. By Year 10, the average CMO in our study had approximately 13 schools, while the CMOs in the broader universe had approximately 6.[4]

C. Students Served

In this section we rely on data for students entering middle schools in 23 CMOs for which we were able to acquire student-level data (including all of the CMOs included in the middle-school achievement analyses described in Chapter IV). Using information on students prior to CMO entry allows us to examine baseline achievement levels of CMO entrants (which would be impossible for students entering CMO schools in kindergarten), and it provides more confidence that data on classifications related to poverty and special needs were consistently collected (rather than potentially being biased by differences in data collection processes of conventional public schools and CMO schools).

[4] Approximately 54 percent of CMOs in our study and 58 percent of the broader universe have been operating for 10 or more years.

Figure II.7. Number of Schools Per CMO, by Year of Operation, for Study CMOs and Broader Universe



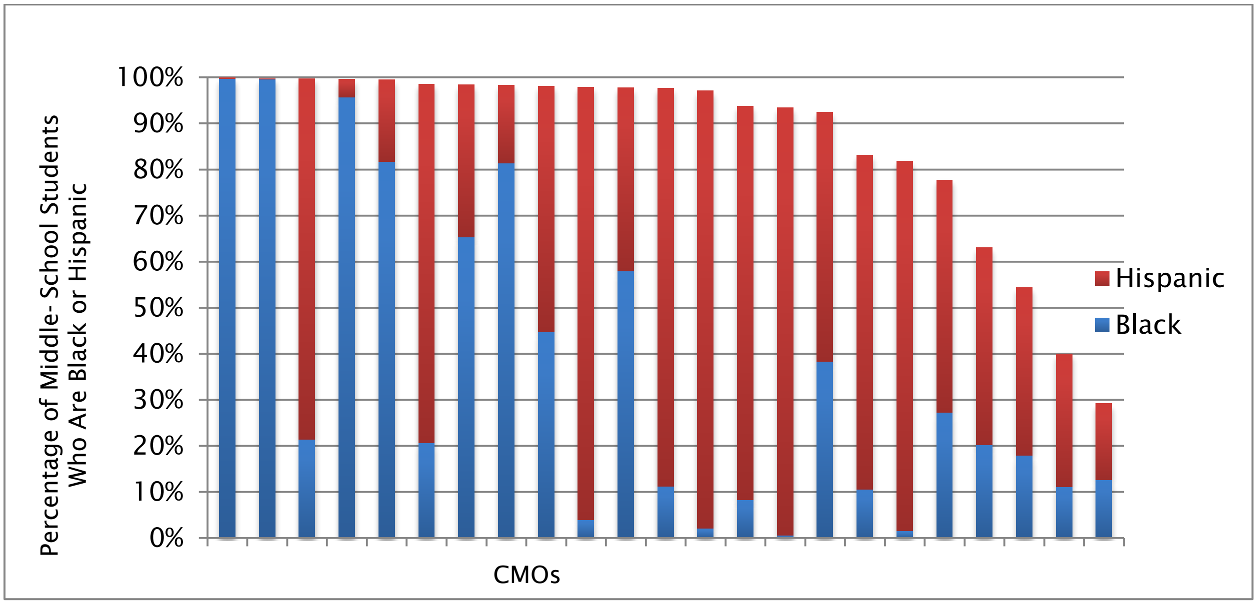
Source: CMO Central Office Staff Survey; broader CMO "universe" data from National Alliance for Public Charter Schools.

1. CMO schools serve a greater share of minority and low-income students than do their districts of residence, but fewer students with special needs or limited English proficiency

The decision by several CMOs to operate in large urban areas reflects not only an interest in achieving local economies of scale but also their mission to serve disadvantaged students. All but two CMOs in our sample serve mostly Black and Hispanic ("minority") students (as Figure II. 8 shows). As indicated in the figure, over half of sample CMOs at the middle school level serve nearly all Black or Hispanic students, while others serve a mix of both groups, and in some cases other students, such as Caucasian and Asian students. (High school-level results are similar.)

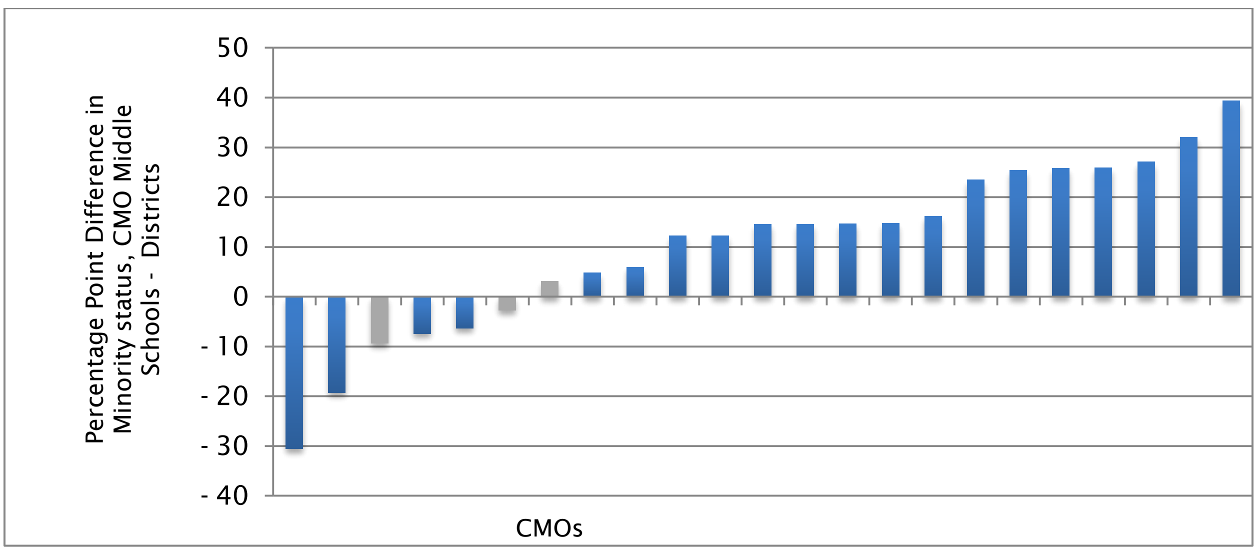
Even when compared to their host districts (which tend to be urban and high-minority), the student population served by the average CMO middle school in our study includes a greater percentage of minority students (see Figure II.9). This may reflect CMOs' efforts to target minority families and communities within their host districts. In the average CMO, approximately 91 percent of (middle-school) students are minorities compared to 76 percent of their host districts' students in the same grades (Table II.2). In Figure II.9 the CMOs with bars above the x-axis (horizontal line) have a greater proportion of minority students than their host districts.

Figure II.8. CMO Middle Schools Primarily Serve Black and Hispanic Students



Source: State and district school records.

Figure II.9: CMO Middle Schools Serve a Greater Percentage of Minority Students Than Their Districts of Residence



Source: State and district school records

Note: Blue bars represent statistically significant differences at the 5% level

Table II.2. Student Demographics in CMO and Host District Middle Schools

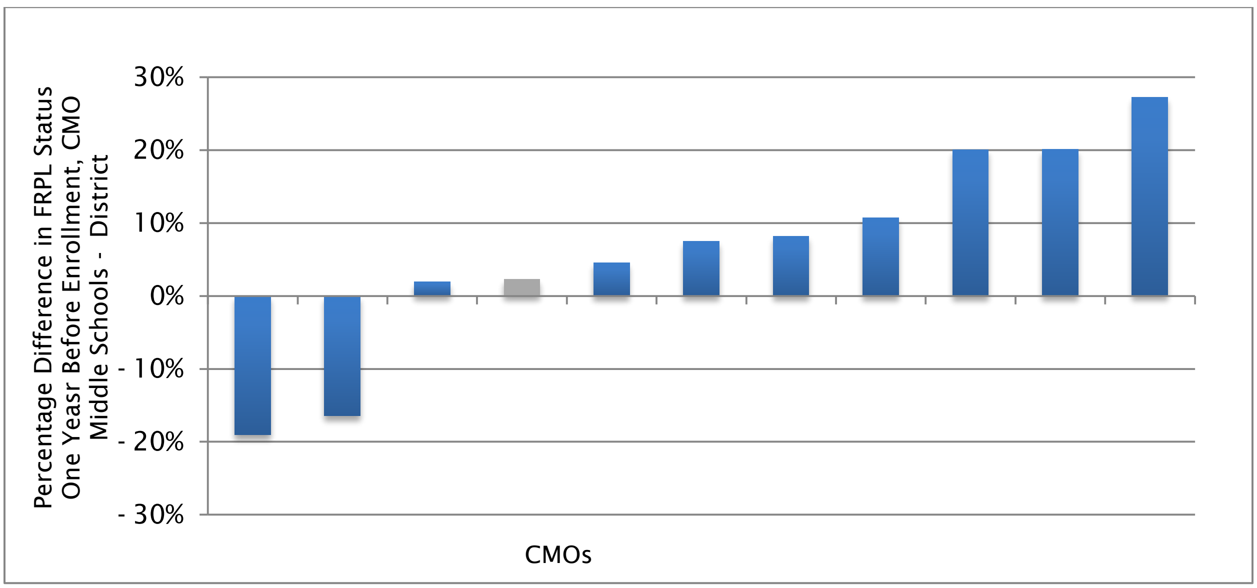
	Percentage Minority	Percentage Free and Reduced-Price Lunch	Percentage Limited English Proficient	Percentage with IEP
CMO Average	91%	71%	14%	9%
CMO Host District Average	76%	64%	19%	13%

Source: CMO and district numbers are from state and district school records; these are unweighted simple averages (rather than student-weighted averages).

NA = not available.

CMOs also appear to focus on serving low-income students, although relevant data are only available for 11 of these CMOs. For the average CMO, approximately 71 percent of students entering middle school qualify for free and reduced-price lunch, compared to 64 percent of students served by their host districts in the same grades (see Figure II.10). Eight of the CMOs tend to serve significantly more low-income students than their corresponding districts, while only two serve fewer low-income students.

Figure II.10. CMO Middle Schools Serve Greater Percentage of Students Who Qualify for Free and Reduced-Price Lunch than Host Districts



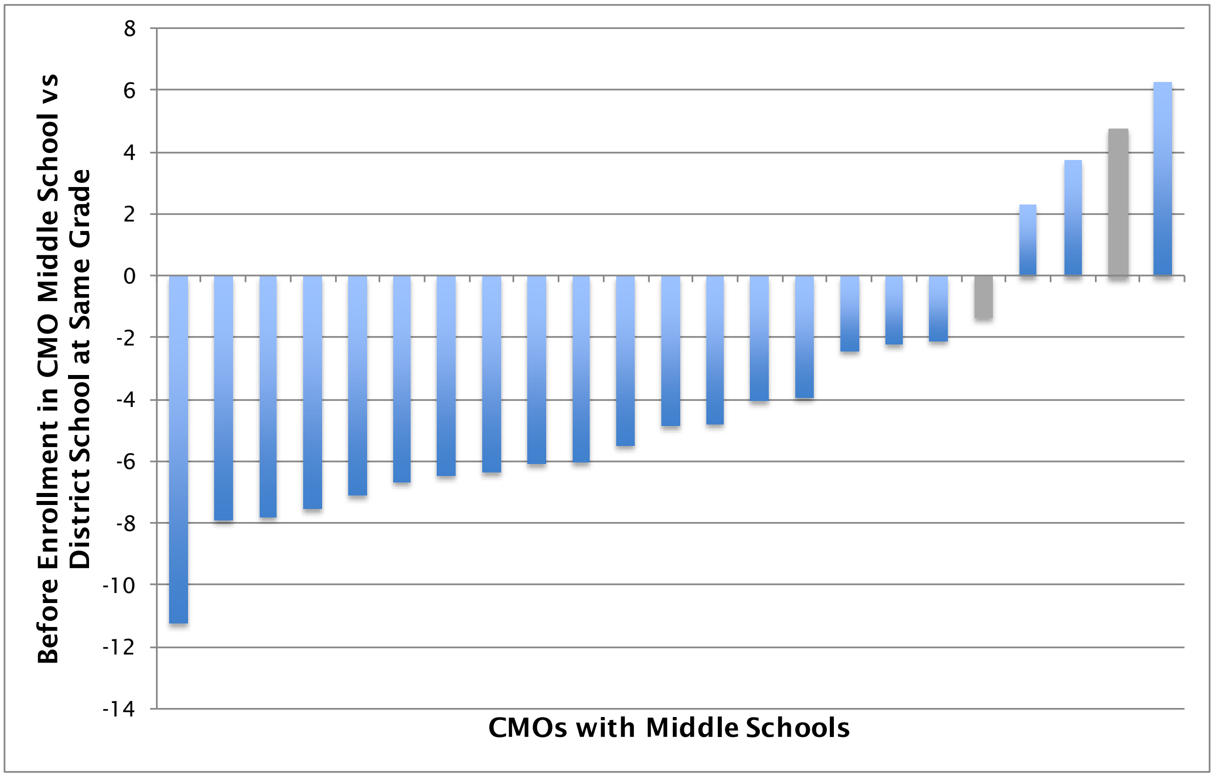
Source: State and district school records.

Note: Blue bars represent statistically significant differences at the 5 percent level.

FRPL = free or reduced-price lunch.

On the other hand, special education students and English-language learners appear to be less likely to enroll in CMO schools (see Figures II.11, II.12). On average, nine percent of the students entering these CMOs are students with disabilities that qualify them for Individualized Education Plans (IEPs), compared to 13 percent of host district students (Figure II.11). Similarly, students with limited English proficiency (LEP) appear to be less likely to enroll in CMOs (Figure II.12). It is possible, however, that these differences are over-stated because of differences in the way schools categorize special needs and LEP status.[5]

Figure II.11. Most CMOs Serve Fewer Special Education Students than Host Districts

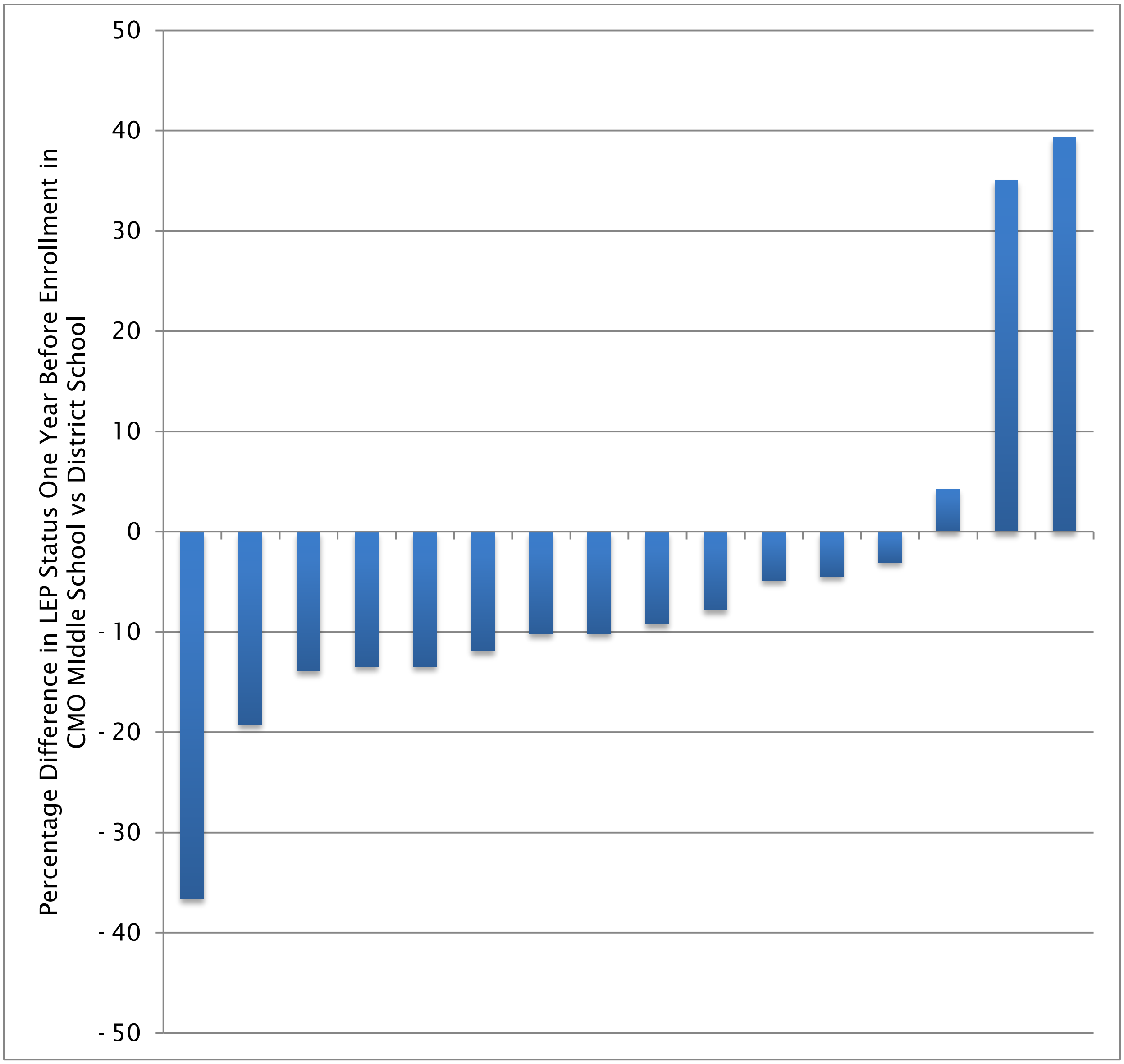


Source: State and district school records.

Note: Blue bars represent statistically significant differences at the 5 percent level.

[5] The data on IEP and LEP status are drawn from records of students prior to entering the CMO school and for all students in the same grades in district schools. Most of the CMOs draw somewhat more students from charter feeder schools than do regular district schools in their host districts. If charter schools systematically under-identify special needs and limited English abilities, then the rates of special needs and LEP status may not be fully comparable for the CMOs and districts.

Figure II.12. Most CMOs Serve Fewer Limited English Proficient Students than Host Districts



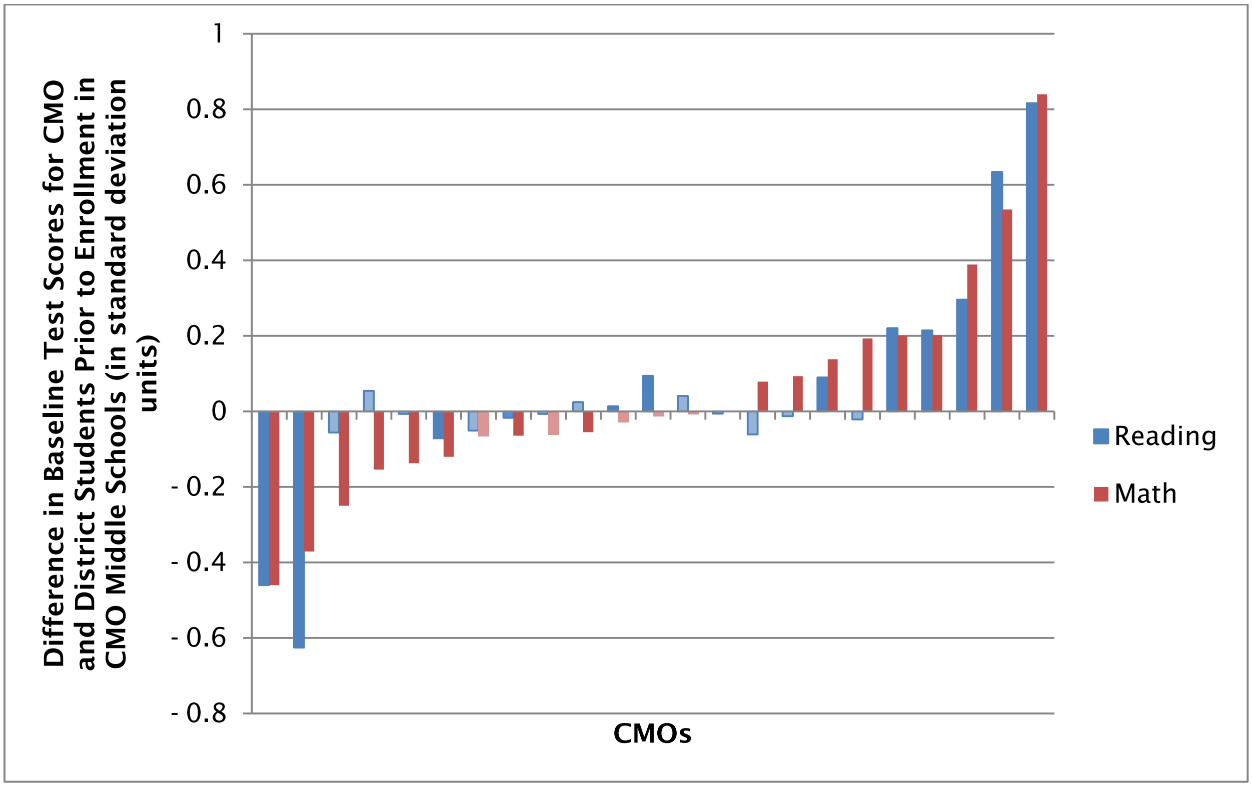
Source: State and district school records.

Note: Blue bars represent statistically significant differences at the 5 percent level.

2. Students entering CMO middle schools typically have prior achievement levels that are similar to the local average and somewhat higher than the local average for black and Hispanic students

We also examined students' academic performance before enrolling in a CMO middle school. The majority of CMOs—18 of 23—enroll students whose prior achievement levels are reasonably similar to (within a quarter of a standard deviation) those of other students in their districts. Two CMOs enroll students with baseline test scores that are more than 0.25 below those of their districts, while three CMOs enroll students with baseline test scores that are more than 0.25 above those of their districts (see Figure II.13).

Figure II.13. Incoming Reading and Math Scores for CMO Students and Their District Peers



Source: State and district school records.

Dark blue and red bars represent statistically significant differences between CMOs and their host districts at the .05 levels, two- tailed tests.

However, most CMOs attract somewhat higher achieving students of color relative to those served by their host districts. Thirteen of 22 CMOs in our sample serve black students who had significantly higher average pre-entry reading test scores than the averages for their black peers in the host district; only two CMOs served black students with scores significantly lower than those of black students locally. Likewise, the pre-entry reading scores of Hispanic students in 13 of 23 CMOs were significantly higher than Hispanic averages locally, and only three CMOs served Hispanic students with significantly lower baseline reading achievement than that of other Hispanic students in their districts.[6] The percentages are similar for reading test scores. Thus, while CMOs attract a disproportionate number of black and Hispanic students, these students tend to have higher test scores on average when they enter the CMO than their black and Hispanic peers in the host districts.

3. Like many independent charter schools, many CMO schools offer grade configurations, that differ from those of district schools

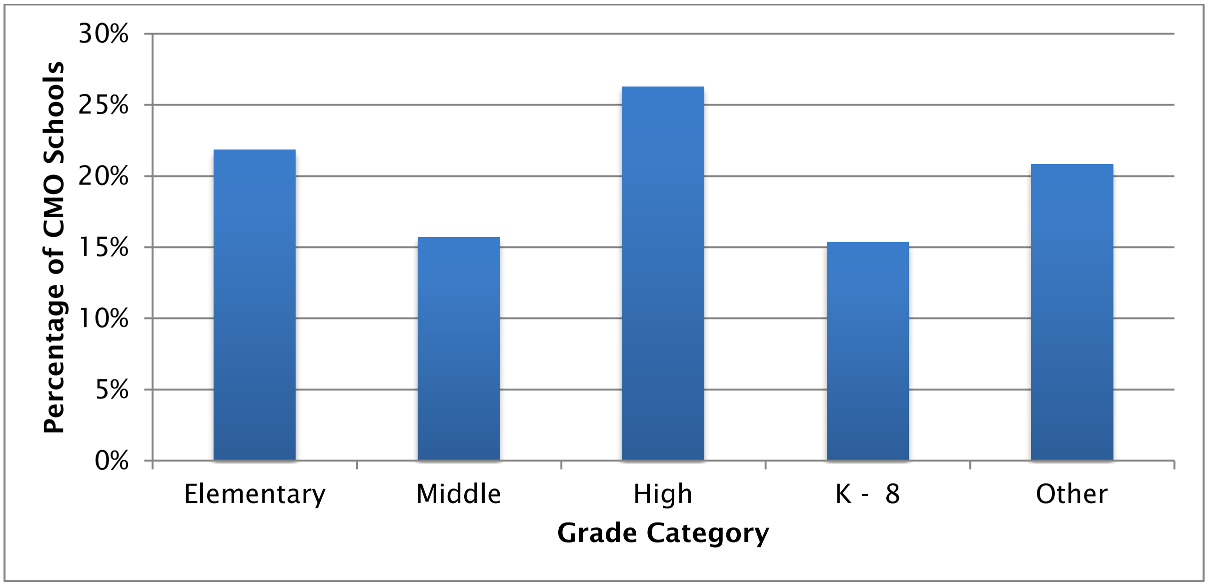
CMOs have the ability to choose which grade levels they wish to serve overall and in a single school. Many use that flexibility to develop schools that have a broader range of grades than is the

[6] The number of CMOs examined differs for black students and Hispanic students because one CMO has no black students.

case among most district schools. As Figure II.14 shows, K-8 schools are fairly common in CMOs, as are other configurations that span traditional middle and high-school grades or even elementary, middle, and high-school grades. Not captured in the figure is the fact that CMO middle schools are also often configured unconventionally, serving grades 5-8 instead of 6-8 or 7-8.

Approximately one-third of CMOs appear to specialize in schools with a single grade configuration (e.g., offering only middle schools or only K-8 schools). The rest combine more than one category, often in feeder patterns (e.g., an elementary CMO school near a middle school).

Figure II.14. Percentage of CMO Schools Serving Various Grade Level Categories



Source: Data assembled by study team from CMOs

Note: Elementary includes all schools that serve only students in grades K-6, middle includes all schools that serve only students in grades 5-8, high includes all schools that serve only students in grades 9-12, K-8 includes all schools that span grades K-8 but do not include high-school grades, and other includes all schools that serve any other mix of grades, spanning middle and high or elementary, middle, and high.

D. Resource Use

1. Per-student spending reflects state charter funding patterns

As with other charter schools and school districts, CMOs are funded through a combination of public funding and charitable donations or grants. Expenditures for the 39 study CMOs for which data was available ranged from $5,000 to $20,000 per pupil (see Figure II.15). This variation appears to be driven in part by the per-pupil funding amounts determined by public funding formulas in each state's charter school law (Figure II.16). The correlation between per pupil CMO expenditures and per pupil state funding of charter schools is 0.61.

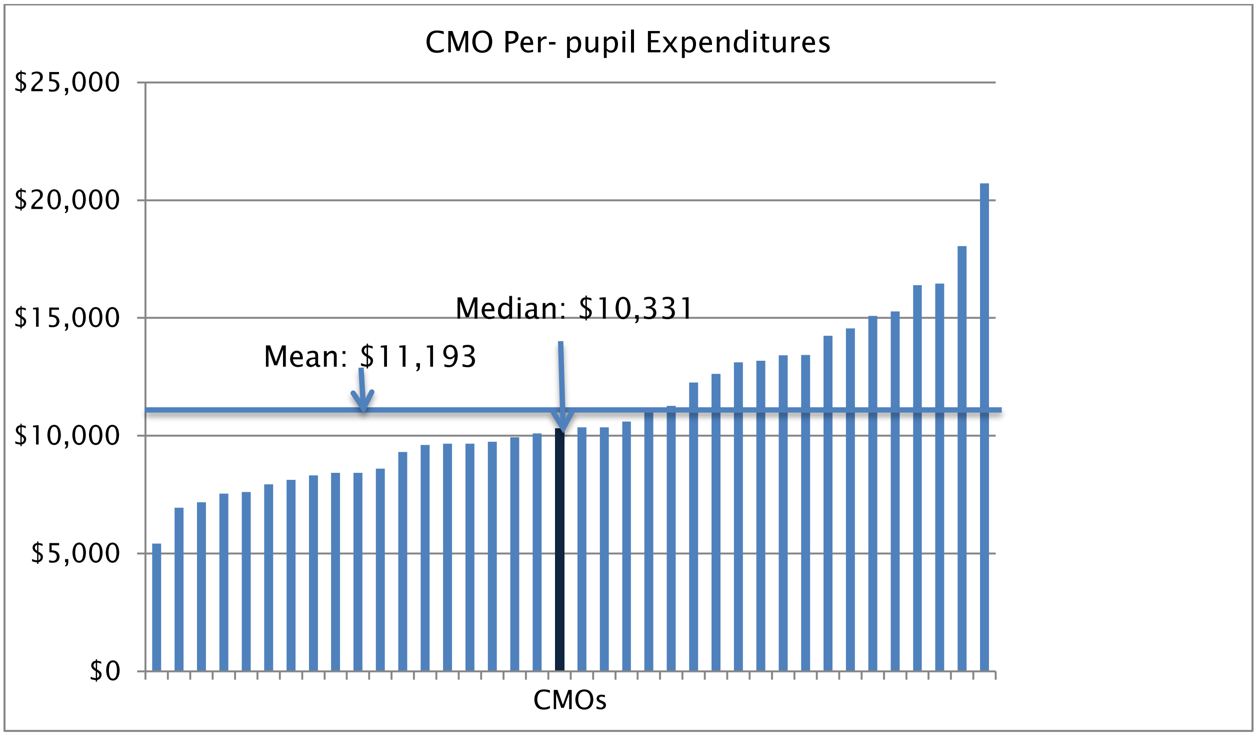
In addition to state funding, some CMOs receive philanthropy and other funding. At least 9 of these CMOs spend more than $1,000 per pupil beyond the amounts allocated from public sources and four CMOs spend more than $4,000 per pupil more. Because public per-pupil revenue allocations to charter schools are often lower than those granted to traditional public schools, it remains unclear whether the CMOs' per-pupil spending is larger or smaller than that of nearby district schools.

2. Most central office staff devoted to educational support, operations, and finance

CMOs look very much like school districts, both in organizational structure and functions served. Central office staff provide supports, services, and oversight for the schools they manage. Among the 37 CMOs responding to our survey of central office staff, the majority of CMO positions are directed at educational supports (such as professional development, coaching, assessment, and data analysis), operations (such as payroll and facilities management), and finance. Figure II.17 shows how these CMOs allocate their central office staff by functional area, as reported by each CMO. Some CMOs invest heavily in large central offices, while others maintain a fairly minimal level of administrative staff. Decisions about how to allocate central staff appear to be more a function of CMO preference than a function of size, and CMOs vary widely in how they allocate their staff across categories.

The overall size of the central office in relation to number of students served also varies widely. Although one might expect this ratio to drop as CMOs grow because of economies of scale, there is no significant relationship between size and staff-to-student ratio. This may be because some large CMOs attempt to provide more coaching, guidance, or other support to their schools.

Figure II.15. Variation in Per- Pupil Expenditures by CMO



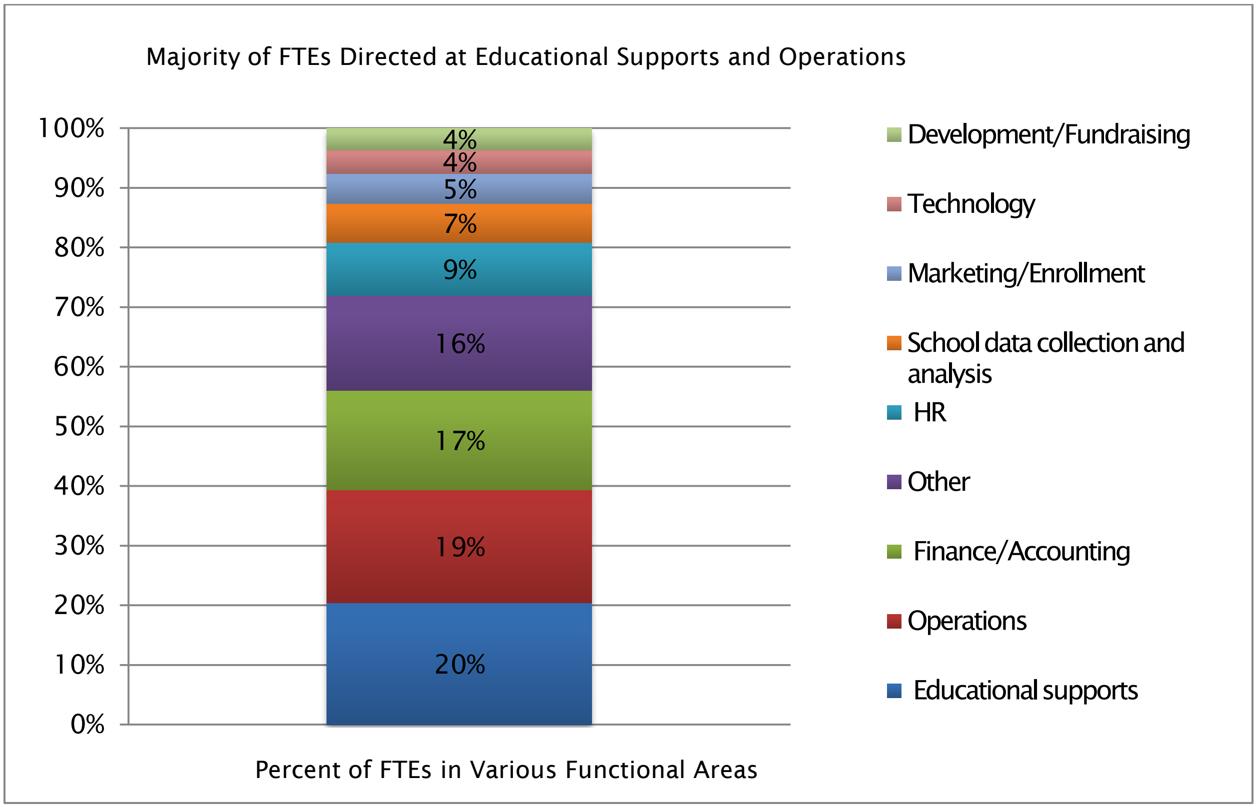
Source: IRS 990 forms for tax year 2009.

Figure II.16. CMO Expenditures Compared to Charter Per- Pupil Public Funding

Source: CMO expenditure data from IRS 990 forms for tax year 2009, per- pupil revenue data from Ball State University, *Charter School Funding: Inequity Persists.*

PPE = per- pupil expenditures; PPR = per- pupil revenue.

Figure II.17. Average Distribution of CMO Staff, by Functional Category



Source: CMO Central Office Staff Survey

FTE = fu ll- time equivalent.

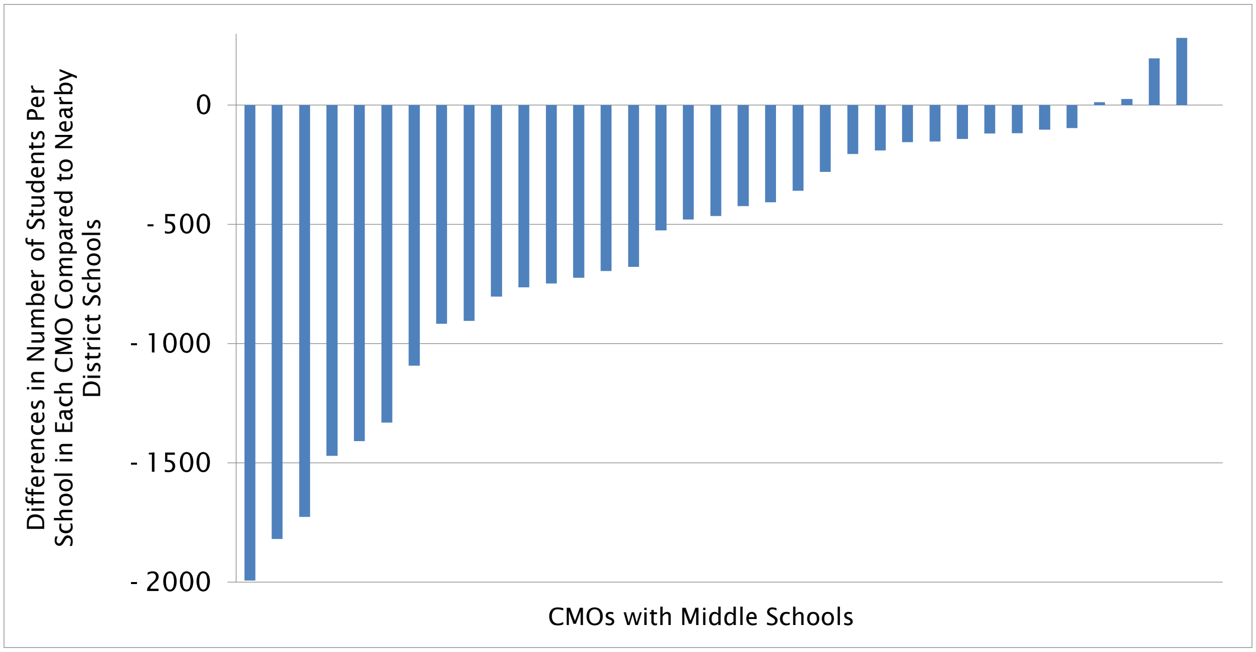
3. CMO schools tend to be smaller than schools in their host districts, with marginally lower student-teacher ratios.

CMOs are generally free to restrict their school and classroom sizes as they choose. Because their organizational mission often focuses on personalized learning and a strong, intimate school culture, one might expect CMOs to have smaller school and class sizes.

Indeed, CMO schools are much smaller than nearby schools in their districts (Figure II.18). The CMOs in our study average 389 students per school compared to 982 students per school in nearby district schools. CMO schools have approximately the same number of grades per school as their district counterparts so the difference in total enrollment is largely due to a smaller number of students per grade.

Class sizes and student-to-teacher ratios are also somewhat smaller in CMO-run schools than in their host districts, although the average differences are not as great as the school size differences. The average pupil-to-instructor ratio in math and reading are about 20.9 students; by contrast in comparison schools this ratio is 23.5 in math and 23.2 in reading. These CMO-district differences are statistically significant in math but not in reading.

Figure II.18. Enrollment Per School in CMOs Compared to Nearby District Schools



Source: Principal Survey

III. CMO PRACTICES AND SCHOOL OUTCOMES

Key Findings

- Relative to district principals, CMO principals report that their schools provide more instructional time, are more likely to have a comprehensive school-wide behavior policy, practice more frequent teacher coaching and monitoring, and place more emphasis on performance-based compensation.
- According to principals, CMOs are less likely than districts to prescribe a particular curriculum, behavior policy, or staff evaluation approach; however CMO principals are more likely to report implementing a school-wide behavioral strategy than their district counterparts.
- Different groups of CMOs tend to adopt distinct sets of practices: those emphasizing behavior policies also tend to have longer instructional time while those emphasizing performance-based compensation tend to make frequent use of formative assessments; both of these groups provide intensive teacher coaching.

A. Introduction

In this chapter, we describe the policies and practices of CMOs and CMO schools. Any effects that CMOs have on student achievement presumably occur through their influence on their schools and on students' educational experiences. Our study is configured to identify some of the ways that CMO school practices and policies differ from those of nearby district schools. We also explore the differences among CMOs in their central office-level and school-level practices. Drawing upon CMO central office, principal, and teacher surveys, this chapter addresses the following questions:

1. How do CMO school policies and practices differ from district schools? What are the key elements that distinguish CMO instructional systems, human capital strategies, and their approach to school culture and parental involvement?
2. To what extent do CMOs centralize decisions about school-level practices? Are CMOs more or less prescriptive than districts? How does prescriptiveness vary across CMOs? To what extent are schools within CMOs consistent in their implementation of specific practices?
3. How much variation is there in CMO practices? Is it possible to identify types of CMOs that place more or less emphasis on specific sets of practices?
4. To what extent do CMO practices appear to influence the instructional coherence and organizational health of their schools?

Any differences in practices among CMOs (relative to nearby district schools) could potentially contribute to differences in their impacts on student achievement. Before estimating CMOs' achievement impacts, we identified a number of practices and structures that might be associated with impacts. These practices were selected using the following three criteria. First, we attempted to identify characteristics that CMOs or researchers believe is a promising practice or structure that is important to improving achievement. Second, using our surveys, we looked for practices where there is substantial variation across CMOs in the characteristic and, most importantly, in the differences between each CMO and its comparison schools. Third, using the principal survey, we sought to identify practices that distinguish CMOs schools from nearby district-operated schools.

The practices included among our seven "primary hypotheses" are the focus of our descriptive analyses in this chapter. We also describe some additional practices that distinguish CMO schools from district schools and could impact student outcomes. The seven primary hypotheses focus on whether impacts are associated with the following seven CMO characteristics:

1. Amount of instructional time
2. Consistent educational approach, including curriculum and instructional materials
3. Student behavior policies that include specific rewards, sanctions, and commitments
4. Intensive teacher coaching and monitoring
5. Performance-based teacher evaluation and compensation
6. Frequent review and analysis of student formative assessment data
7. Number of CMO schools

The first six characteristics are practices that are discussed in this chapter; Chapter V examines how these practices relate to impacts on student achievement. Chapter II contains information on the number of schools in each CMO and Chapter IV examines how this variable relates to impacts.

We also hypothesized that CMO practices may have an effect on student achievement indirectly via two mediating factors: (1) instructional coherence—the degree to which different aspects of the school's instructional program reinforce one another, and (2) organizational health—the degree to which the school is efficient and effective in maintaining a stable environment with few administrative problems. Instructional coherence might be impacted by whether a CMO has a consistent educational approach, for example. Organizational health may be influenced by factors such as centralization of decisions at the CMO level or the CMO's size.

Methods

Several of our measures of CMO practices are composite variables. These composites were created by combining closely related survey items into a single measure, reducing measurement error and capturing the breadth of a construct. See Appendix A for details on how these composites were created.

To measure the CMO characteristics described in this chapter, we chose to rely primarily on the principal survey, rather than on our central office staff survey or teacher survey, for several reasons. First, relative to the central office survey, the principal survey reflects the perceptions of respondents who are closer to the implementation of CMO practices in schools. Second, the principal survey provided a unique opportunity to contrast practices in CMO schools with those in nearby district schools. Each CMO principal eligible for the survey was matched with the principal of the district school with the same grade range in closest geographic proximity, and we attempted to survey all matched comparison principals in addition to all CMO principals.

Our unit of analysis is the CMO.[1] Some of our measures incorporate survey questions that ask about policies and activities of the CMO central office. However, we also rely on principal survey responses about school activities to infer overall CMO-level characteristics, averaging principal responses within each CMO and within the group of comparison schools associated with that CMO using weights to adjust for nonresponse.[2] There is substantial consistency within CMOs in these practices relative to the variation between CMOs (see Appendix A.).

We acknowledge that our measures of CMO practices rely on principal reports of CMO practices and not on direct observation of these practices in CMO schools. Given that the surveys were conducted for a study of CMOs, it is possible, for example, that CMO principals tended to report school-level practices in the most positive light. We examined correlations between similar teacher and principal survey responses and confirmed that all correlations were positive and, on most individual items, substantial. In addition, when measuring school-level intermediate outcomes in the domains of instructional coherence and organizational health, we used the teacher survey exclusively for all measures of instructional coherence and for a measure of teacher satisfaction in the organizational health domain.

In this chapter, we begin by describing average differences in CMO and district school practices in the following areas, in turn: instructional time, educational approach, student behavior strategies, and teacher effectiveness. We also describe the variation across CMOs in the emphasis on these practices in their schools. Next, we use cluster analysis to explore whether CMOs can be classified into groups based on their management strategies and bundling of key practices. Finally, we describe associations between the intermediate school outcomes in instructional coherence and organizational health and the school practices identified in the first part of the chapter.

B. CMO and School Practices: Significant Differences from District Schools

CMO and district schools appear to differ in many potentially important respects. Based on the principal surveys, it appears that CMO and district schools are appreciably different in terms of their instructional time, educational approach, student behavior policies, compensation and evaluation policies, and strategies for teacher coaching and monitoring. Although we focus on the noteworthy differences here, there were also areas of similarity. For example, the frequency with which CMO and district school and central office staff review and analyze student assessment data—one of our primary hypotheses—was not significantly different (Table III.1).

---

[1] The unit of analysis for the comparison sample corresponds to the group of district schools matched with the schools within a CMO. Therefore, for CMOs with schools located in more than one district, this comparison unit aggregates responses from multiple districts, weighted by the number of CMO schools located in each district.

[2] Of 292 CMO principals eligible for the study, 76 percent responded to the survey. Among the 292 matched comparison principals, the response rate was 59 percent. Five CMOs eligible for the study declined to participate in the survey. We developed weights to adjust for principal nonresponse.

Table III.1. CMO Schools and District Schools Diverge on Most Primary Hypotheses Practices

Primary Practice	Significant CMO- District Difference?
Consistent educational approach, including curriculum and instructional materials	√
Student behavior policies that include specific rewards, sanctions, and commitments	√
Frequent review and analysis of student formative assessment data	
Intensive teacher coaching and monitoring	√
Performance- based teacher evaluation and compensation	√
Amount of instructional time	√

1. On average CMOs offer more instructional time

Recent literature suggests that extended learning time may be a key component distinguishing more successful schools.[3] Expanding learning time is a strategy implemented by some CMOs. We explored the variation in instructional time in CMOs relative to districts and examined differences after decomposing instructional time into hours per day and days per year.

Relative to districts, CMO schools tend to require significantly more time in school both during and outside of the academic year. On average, CMOs provide 1,373 hours of instruction per year compared to 1,239 for districts, a difference of 134 hours (Table III.2). The difference between CMO and district schools is driven more by extending the length of the school day than by extending the length of the school year. CMO schools have significantly longer school days. The average length of a CMO school day, as reported by CMO principals, is 7.5 hours, while the mean district comparison school reported a school day of 6.9 hours.[4] CMO schools and district schools report similar numbers of instructional days per year, with a mean of 182 at the CMO level and a mean of 180 at the district comparison level.

Table III.2. CMO and District School Instructional Time Practices[5]

Practice	n	CMO Mean	Comparison Mean	p- value
Yearly Instructional Hours	36	1,373 (185)	1,239 (98)	$p<.001$
Required Summer Session Prior to Enrollment, Proportion	36	0.17 (0.27)	0.05 (0.11)	$p=.02$

Source: Principal Survey.

[3] Angrist et al. 2011.

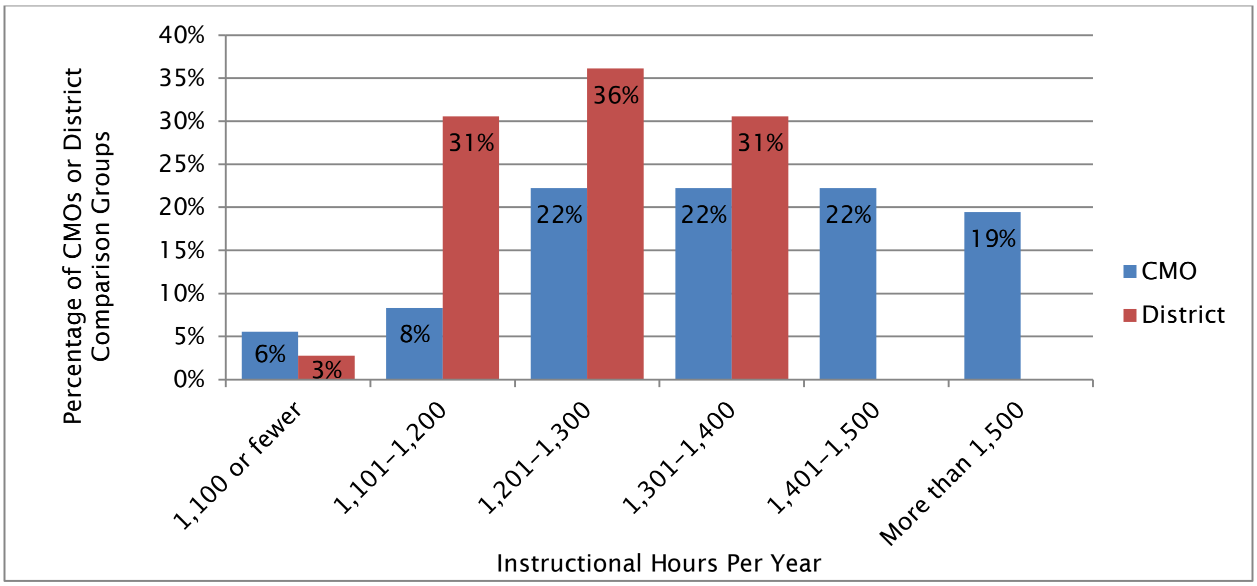
[4] $p=0.0001$

[5] In this table and throughout this chapter, the significance of the CMO-comparison difference was tested using a two-tailed t-test with unequal variance assumption. P-values smaller than 0.05 were considered significant.

Although the number of days per year in CMO and district schools is similar, CMO principals were more likely to report the use of a mandatory summer session for new students prior to enrollment. An average of 17 percent of principals in a CMO report this requirement, relative to 5 percent among district schools (p=0.02).

Although CMO principals report more instructional time than district principals on average, the mean masks the substantial variation across CMOs in the amount of time students are required to be at school. Approximately one-fifth of the CMOs require an average of more than 1,500 hours annually, while 36 percent require 1,300 hours or fewer. As Figure III.1 indicates, there is substantially more diversity in instructional time requirements among CMO schools than among district schools. Despite the significant variation across CMOs in instructional time,[6] standards appear to be very consistent within CMOs.[7]

Figure III.1. CMO Students Spend More Time in School



Source: Principal Survey.

2. CMO principals report more autonomy choosing their curriculum

Independent charter schools are often established in order to provide more flexibility in defining an educational approach. As collections of charter schools, CMOs could in theory either provide principals with substantial discretion in selecting curricula and instructional materials or require all schools to adopt the same curriculum. We investigated the extent to which CMO principals perceive flexibility in their curricular choices, whether they report more or less autonomy relative to district principals, and how these practices vary within and across CMOs.

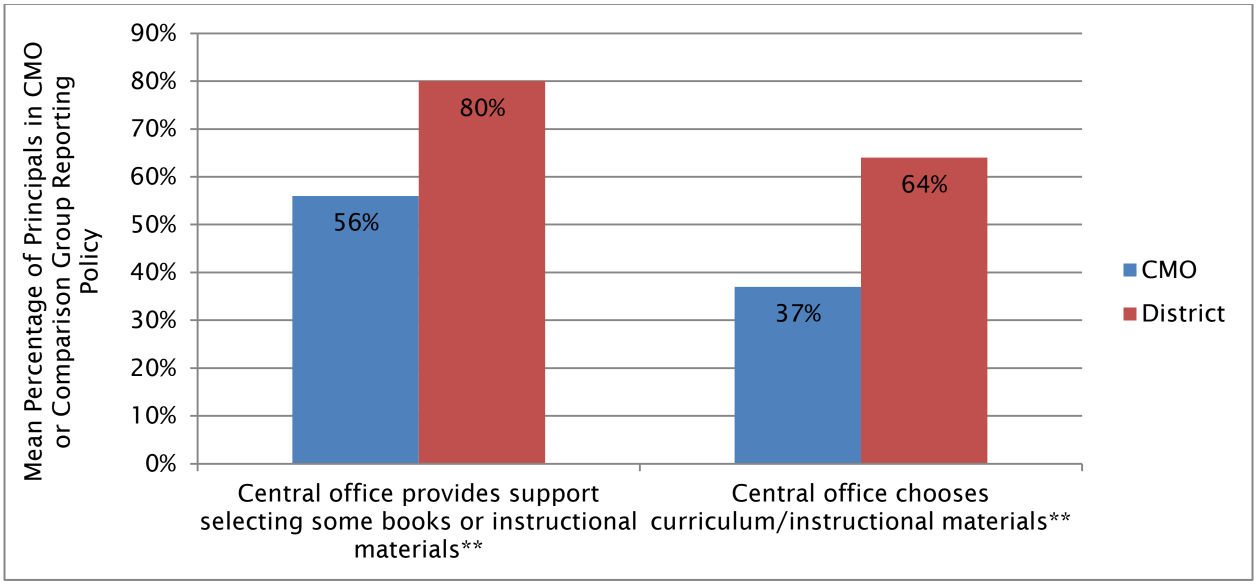
---

[6] A one-way analysis of variance (ANOVA) confirms that the CMO means are not identical, and the average of the CMO means in the top half is significantly larger than that of CMOs in the bottom half (p<.0001).

[7] The intraclass correlation coefficient for yearly instructional time is 0.64, indicating that substantial variance among CMO schools is explained by clustering at the CMO level rather than by school-level practice.

Relative to districts, CMOs are less likely to mandate a specific curriculum or instructional materials and more frequently allow their schools to make such choices independently. On a composite measure of central office-level centralization of educational approach, CMO schools reported significantly less centralization than district schools, on average. This difference was driven primarily by of the extent of central office prescriptiveness in selecting instructional materials. As indicated in Figure III.2, significantly fewer CMO principals report that their central office plays a role in selecting some books or instructional materials than comparison principals. CMO schools are also less likely than district schools to report that their central office is responsible for actually choosing their curricula or instructional materials.

Figure III.2. CMO Schools Report More Autonomy in Choosing Instructional Materials



Source: Principal Survey.

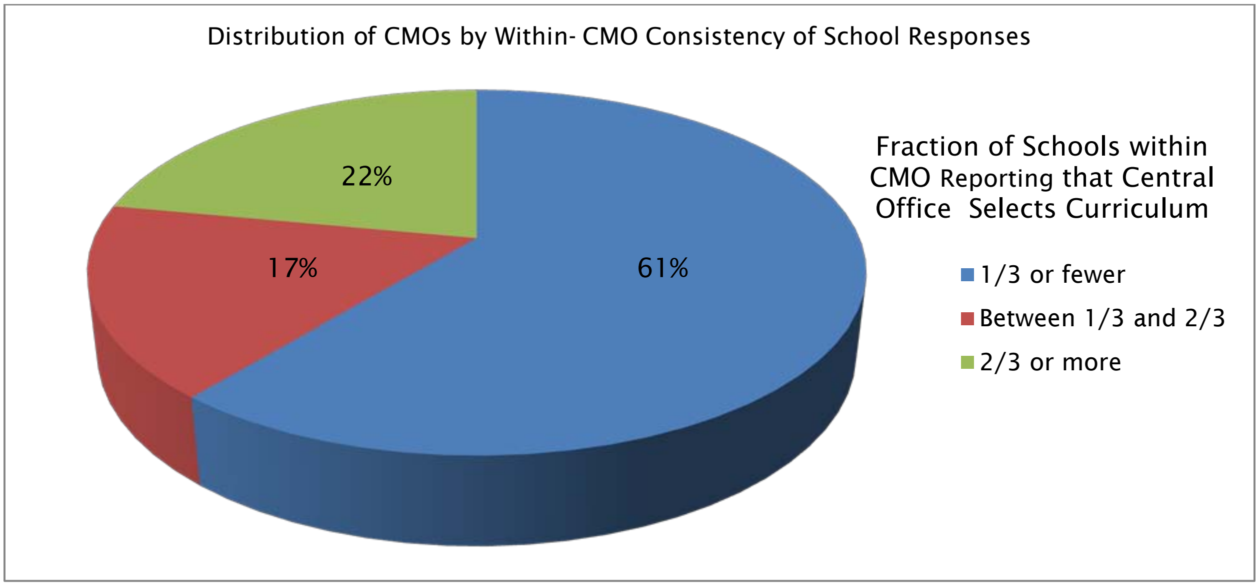
**Statistically significant difference; $p<.01$.

The overall extent to which CMOs centralize decisions on curriculum and materials varies significantly across CMOs but appears to be relatively consistent within CMOs.[8] CMO practice varies most substantially with regard to whether the CMO central office plays any role in helping select some books or instructional materials: in one-third of CMOs, most[9] principals report no central office involvement, while just over one-fifth of CMOs have very mixed principal responses, and the remaining CMOs are reportedly more consistently involved with their schools on this dimension. On the other hand, most CMOs do not explicitly select curricula for their schools, and principal responses indicate consistency in the policy within CMOs. Figure III.3 illustrates that, in a majority of CMOs, most principals report that they have flexibility in choosing curricula or instructional materials; that is, the central office does not mandate a specific curriculum.

[8] The intraclass correlation coefficient for the composite measure of centralization of educational approach is 0.51.

[9] Two-thirds or more of principals.

Figure III.3. Most CMO Principals Report Consistency in Decentralization of Curriculum



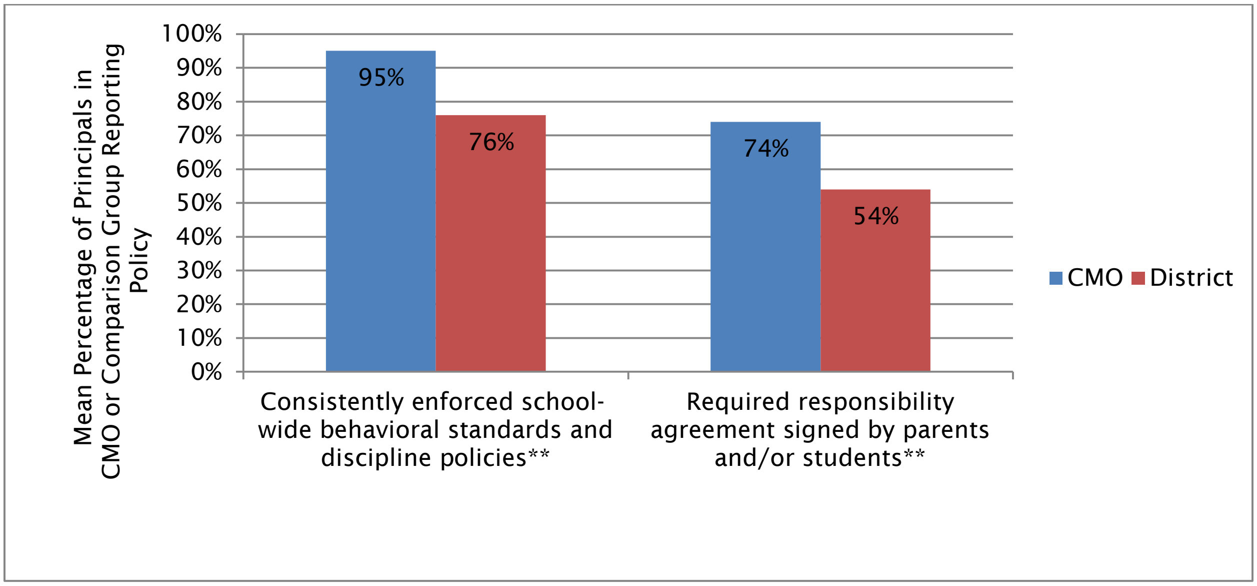
Source: Principal Survey.

3. CMO schools report comprehensive behavior policies, but flexibility in defining the details

Independent charter schools often have a fair amount of autonomy in defining behavior policies relative to district schools, but there has been relatively little information on how CMOs influence their schools' behavior policies. CMOs can potentially influence the behavior policies of their schools in two ways. The first is the extent to which they encourage their principals to develop and enforce some type of comprehensive behavior code. The second is in whether they give principals flexibility in defining the details of that behavior policy, allowing different schools within the CMO to adopt different policies. We first examine school-level practices in developing a comprehensive behavior policy and then turn to CMO-level prescriptiveness, as reported by principals, in mandating specific disciplinary and behavior policies across their schools.

CMO principals were more likely than district principals to report that their schools define and enforce a comprehensive set of behavioral standards and require responsibility agreements. We found significant differences between CMO school practices and district school practices, on average, on a composite measure of comprehensive behavior policy that includes, among other items, the use of rewards and sanctions and the inclusion of a zero tolerance policy ($p<.01$). The difference between CMO and district school practice appears to be driven by two components (Figure III.4). First, significantly more CMO schools implement consistent school-wide behavior codes: on average, 95 percent of principals within a CMO report that behavioral standards and discipline policies are established and enforced consistently across the entire school. Second, CMO schools are significantly more likely to require parents or students to sign an agreement describing their responsibilities.

Figure III.4. CMO Schools Emphasize Behavior Standards and Responsibility Agreements

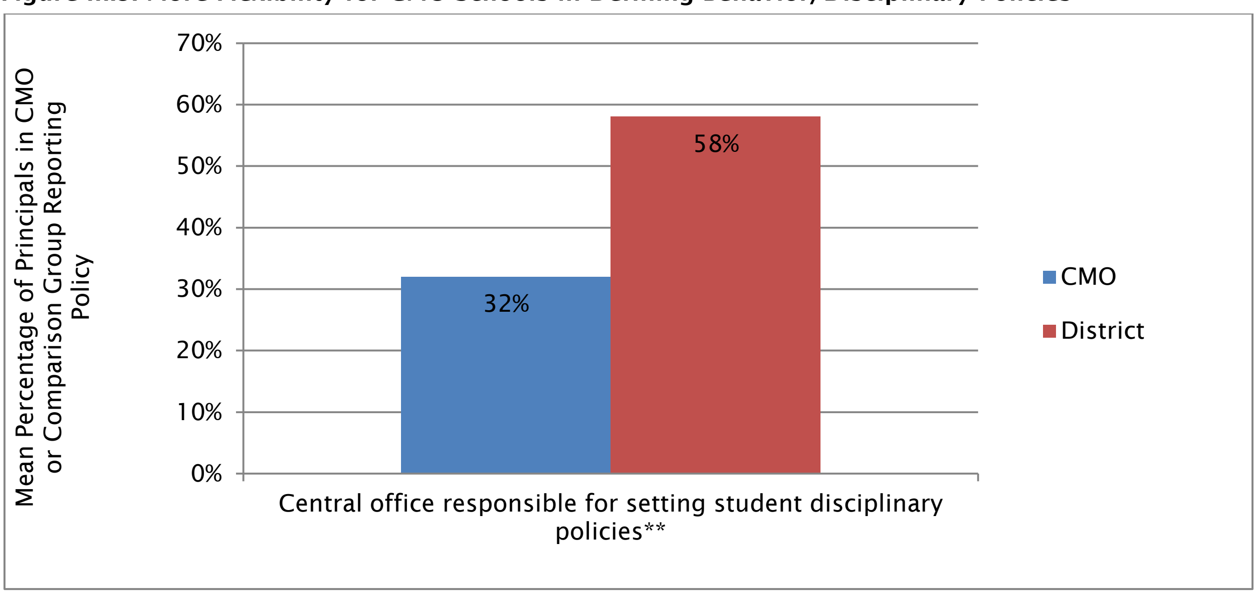


Source: Principal Survey.

**Statistically significant difference; p<.01.

That said, CMOs appear to be less likely than districts to require all their principals to adopt a uniform behavior strategy. Rather, CMO principals report more flexibility in defining the details of behavior policies than do district principals. On a composite measure of the extent to which central office staff provide or mandate disciplinary policies and behavior codes, CMOs averaged significantly lower levels of centralization (p=.007), as illustrated with a specific survey item response in Figure III.5.

Figure III.5. More Flexibility for CMO Schools in Defining Behavior/Disciplinary Policies

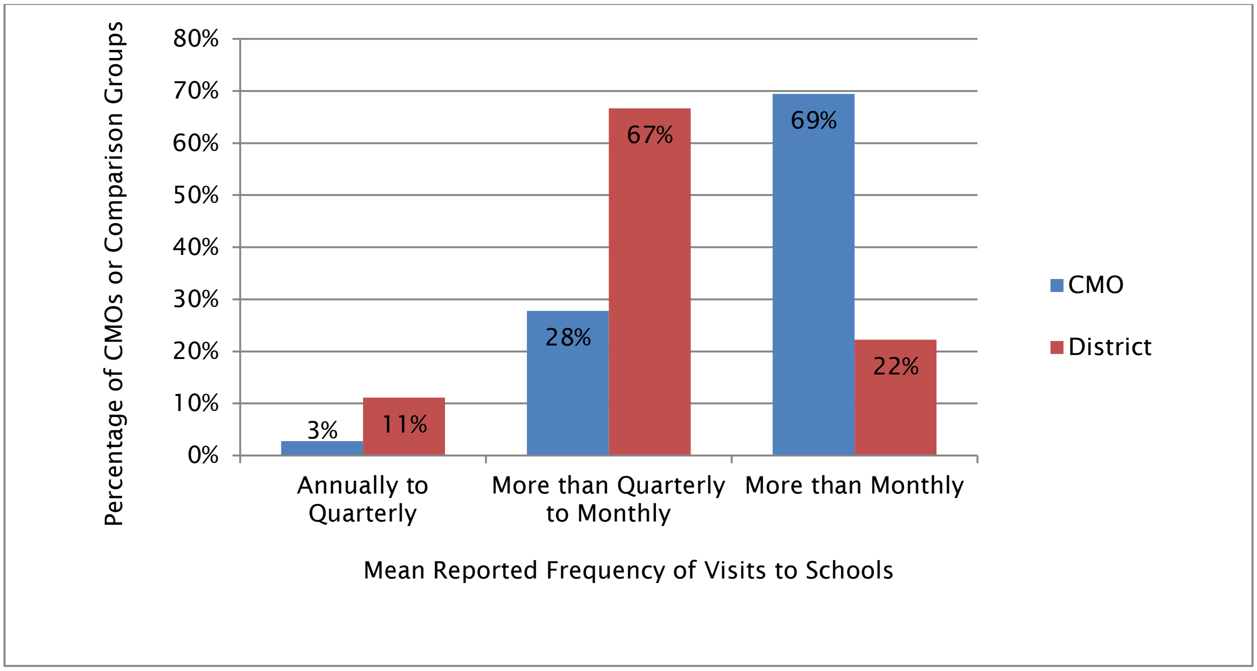


Source: Principal Survey.

**Statistically significant difference; p<0.01.

Although CMOs are less likely than districts to mandate a specific behavior policy, CMO central office staff visit schools more frequently than do district office staff, as reported by principals (Figure III.6.). More frequent visitation, an average of 1 to 4 times per month, might permit them to influence or monitor school climate in ways other than prescribing specific practices.

Figure III.6. CMO Schools Report More Frequent Central Office Staff Visits



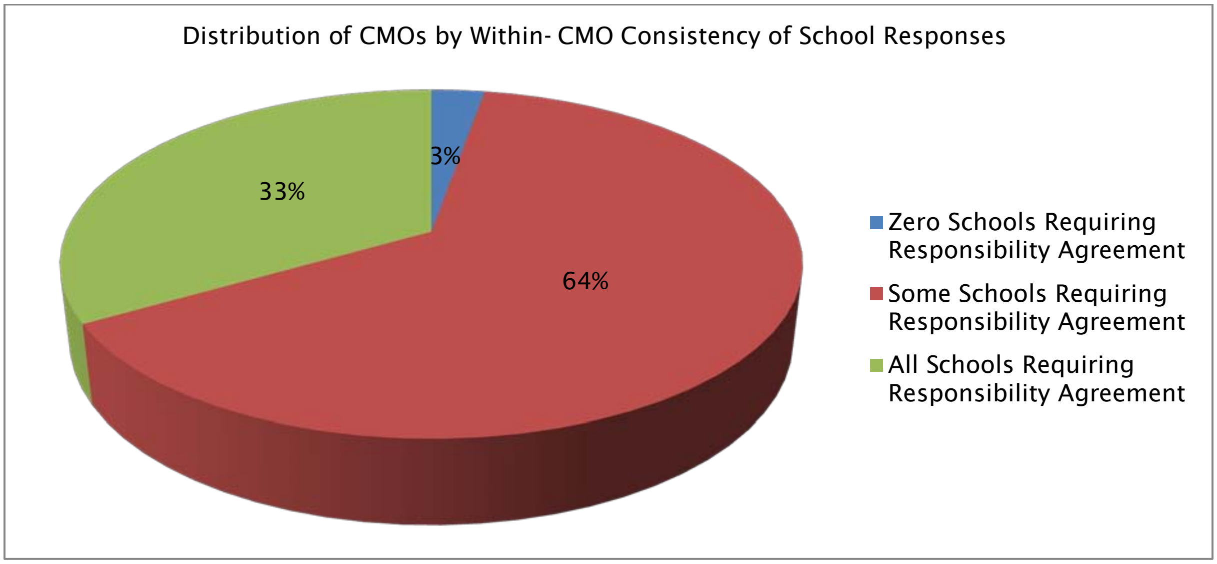
Source: Principal Survey.

Although CMOs vary significantly on the composite measure of school-level behavior policy, there is also substantial variation in the behavior policies across schools within the same CMO.[10] This variation may be expected given the tendency of CMOs not to mandate specific components of behavior and disciplinary strategies. An example of this lack of consistency within CMOs is illustrated in Figure III.7. Among more than 60 percent of CMOs in the sample, at least one but not all surveyed principals reported requiring students and/or parents to sign responsibility agreements.

The survey data suggests that in CMOs with the strongest average emphasis on schoolwide behavior policies, central office staff are either encouraging or requiring schools to adopt specific practices. Using the CMO central office staff survey, we examined CMO reports of behavior policies among the subgroup of CMOs scoring at or above the median CMO score on the schoolwide comprehensive behavior policy index, which was based on principal reports. These CMOs were most likely to report that they require student uniforms (94 percent) and that they require students or parents to sign a contract or letter of commitment (71 percent).

[10] The intraclass correlation coefficient for our composite measure of comprehensive behavior policy is 0.30.

Figure III.7. Substantial Within- CMO Variation in Behavior Requirements



Source: Principal Survey.

CMOs whose schools emphasize enforcing behavioral expectations also tend to require longer instructional time. Of the 18 CMOs with values at or above the CMO median on a composite of behavior policy consistency, 78 percent (14 CMOs) also place in the top half of CMOs when ranked by instructional hours.[11] The CMOs may seek to emphasize discipline both by defining and enforcing behavior policies and by requiring students to spend a long time at school.

4. CMO schools emphasize targeted recruitment and performance-based compensation

One of our core hypotheses focuses on the extent to which CMOs reward teachers based on their performance. More broadly, we were interested in examining how CMOs hire, evaluate, and compensate classroom teachers; whether these practices differ from those employed by districts; and whether these staffing strategies vary within and across CMOs.

CMOs appear to strive to hire and reward school staff who seem to be committed to the CMO mission and effective in the classroom. Relative to district schools, CMO schools place a significantly higher priority on screening teacher applicants based on their performance leading a sample class ($p=0.03$) and their expressed commitment to the school mission ($p=0.01$). CMO schools also appear to draw from a larger applicant pool than do district schools, on average: CMO principals report that the number of applicants per vacancy is approximately 60 compared to 23 for district principals ($p<.01$).[12]

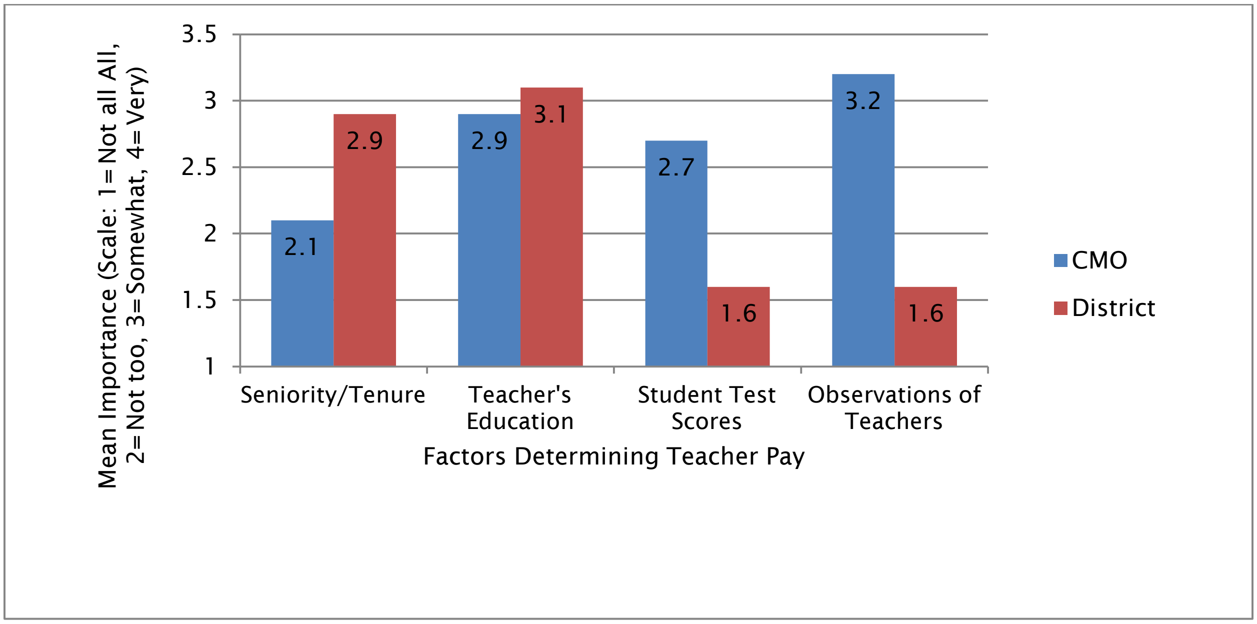
---

[11] A Fisher's exact test confirms that the distribution of CMOs on our behavior policy measure is significantly related to the distribution of CMOs on our instructional time measure ($p=0.0002$).

[12] These differences between CMO schools and district schools in hiring practices and applicant pool could be driven in part by differences in their certification requirements. While we have no data for districts, the extent to which CMOs hire teachers from alternative certification programs, as measured by our CMO central office staff survey, varies substantially: the mean proportion of teachers from Teach For America or Teaching Fellows programs is 17 percent but ranges from 0 to 55 percent.

CMO schools are significantly more likely than district schools to employ a system of performance-based compensation for both teachers and principals, on average.[13] In particular, based on the principal survey responses, we found CMO-district differences in four out of four practices. First, while an average of 69 percent of principals in a CMO report using student test scores to evaluate teachers, only 46 percent of principals in a comparable group of district schools report doing so ($p<.01$). Second, compared to district principals, CMO principals report placing more importance on student test scores and teacher observations and less importance on tenure and education in determining teacher pay ($p<.01$) (Figure III.8). Third, the emphasis on performance-based compensation extends to principals as well: CMO principals are significantly more likely than district principals to have an opportunity to earn a bonus for student achievement ($p<.01$). Finally, CMO teachers are less likely to have an opportunity to earn tenure. Among CMO schools, an average of only 7 percent of principals report offering tenure, while 67 percent of district principals do ($p<.001$).

Figure III.8. Student Test Scores and Observations More Important to Teacher Pay in CMO Schools



Source: Principal Survey.

Although CMO school-level practices indicate a consistently strong emphasis on performance-based compensation and evaluation, it appears that their schools have some latitude in shaping these policies. Nearly all (99 percent) district comparison principals report that the district is responsible for setting teacher salaries, but significantly fewer CMO principals (75 percent, on average) report that the CMO central office does so ($p<.01$). Similarly, nearly all district principals (97 percent) indicated that the central office requires a specific protocol for evaluating teachers, but only 81 percent of CMO principals reported that their central office had a comparable policy ($p<.01$). This flexibility appears to result in some variation within CMOs, but CMOs still account for a substantial

[13] On a composite measure of emphasis on performance-based compensation and evaluation, the average CMO value was significantly higher relative to the district comparison ($p<.001$).

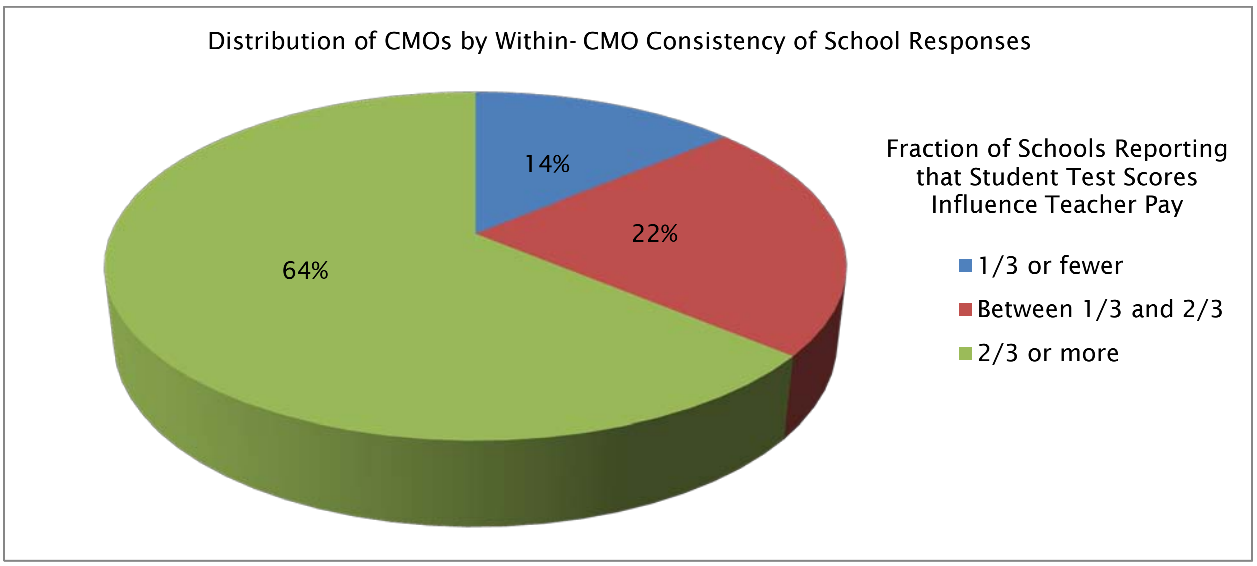
portion of the variation across CMO schools.[14] Further, despite some variation within CMOs on specific approaches, there is substantial consistency within and across CMOs on one particular component: the use of student test scores to evaluate teachers (Figure III.9).

CMOs emphasizing performance-based compensation and evaluation may be applying a broader strategy focused on the use of student data. Indeed, the CMOs that evaluate and compensate their teachers based on their performance also ask teachers and principals to review and analyze formative assessment data more frequently (correlation of .43, $p<.01$). In addition, these CMOs tend to be more likely to apply a centralized instructional model and materials (correlation of .29, $p=.09$). The size and significance of these correlations stand out relative to the correlations of an emphasis on performance-based compensation with other practices.

5. Frequent teacher coaching and monitoring emphasized over workshops

In addition to hiring, compensation, and evaluation strategies, schools may also aim to increase teacher effectiveness via professional development. Professional development can be offered in a variety of ways, including in-service workshops, graduate-level coursework, and peer coaching. We examined both the centralization of professional development in CMOs and the frequency of teacher coaching and monitoring, exploring CMO-district school differences, variation across CMOs, and whether intense teacher coaching correlates with other practices.

Figure III.9. Consistent Use of Student Test Scores to Evaluate Teachers within CMOs



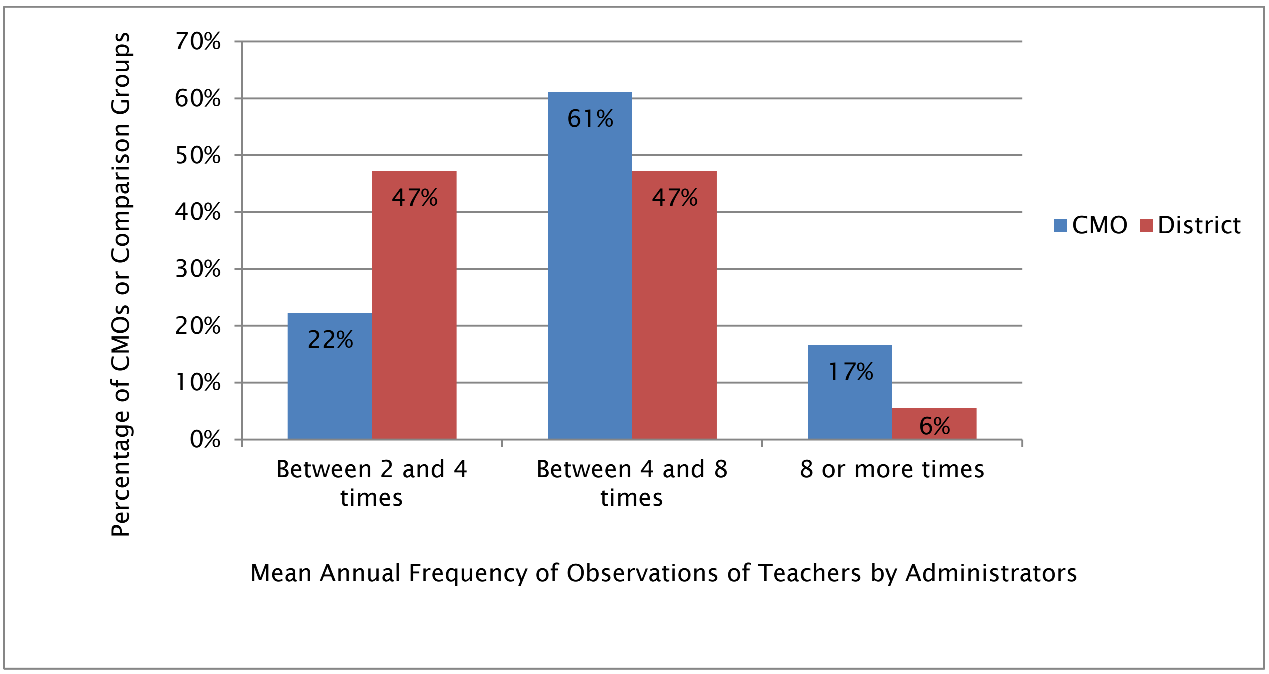
Source: Principal Survey.

[14] The intraclass correlation coefficient for our composite measure of emphasis on performance-based compensation and evaluation is 0.40.

At the school level, CMOs emphasize a higher intensity of teacher coaching and monitoring relative to nearby district schools. On a composite measure of principal reports on the frequency of observations, review of lesson plans, and provision of feedback for teachers, CMOs score significantly higher than the district schools (p=0.01). We find marginally significant (p-values fall between 0.05 and 0.06) differences between CMO and district school practice on three areas in particular. On average, CMO principals reported that the following practices were implemented in their schools with higher frequency relative to comparison district school principals: (1) observation of new teachers by principals or other administrators (Figure III.10), (2) providing feedback from observations to new teachers, and (3) requiring new teachers to submit lesson plans for review.

CMOs are distributed along a spectrum in the extent to which they coach and monitor teachers. A one-way analysis of variance confirms that there is significant variation across CMOs in the degree of intensity of coaching and monitoring. Figure III.11 provides one example of this variation. However, we also found substantial variation in focus on teacher coaching and monitoring at the school level that is not primarily explained by being part of a particular CMO.[15]

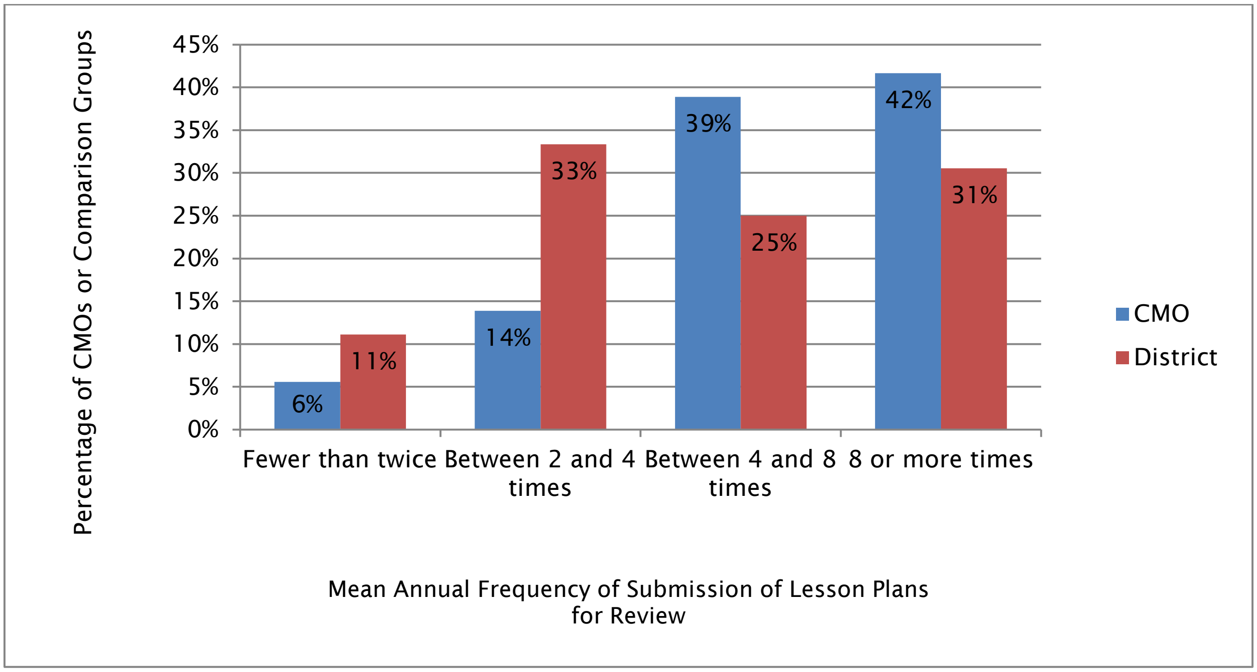
Figure III.10. More Frequent Observation of Teachers by Administrators in CMO Schools



Source: Principal Survey.

[15] The intraclass correlation coefficient for our composite measure of frequency of teacher coaching and monitoring is 0.30.

Figure III.11. More Frequent Submission of Lesson Plans for Review in CMO Schools



Source: Principal Survey.

CMOs are reportedly less likely than nearby districts to provide certain types of professional development support to schools. Specifically, CMO school principals report significantly less central office provision of professional development, such as workshops and in-school service programs, for teachers outside the classroom than do district school principals ($p<.01$). However, although CMOs do not explicitly provide this type of centralized professional development, they appear to seek to shape teacher practice in a different way—through intense teacher coaching and monitoring to develop like-minded staff.[16]

Our analysis of the associations among the practices that comprise our primary hypotheses indicates that CMOs that provide a substantial amount of coaching may fall into two categories. As shown in Table III.3, CMOs that emphasize teacher coaching are also more likely to encourage or require school staff to frequently review and analyze formative assessment data. We also find a strong and significant relationship between the intensity of teacher coaching and the number of instructional hours in the school year.

[16] We find a positive and marginally significant association between intensity of teacher coaching and monitoring and the provision of centralized professional development support (correlation: 0.29, $p=.08$), however, which indicates that the two approaches could be complementary.

Table III.3. Intense Teacher Coaching Highly Correlated with Formative Assessment Use and Instructional Time Correlations between Teacher Coaching and Other Practices (Pearson Correlation Coefficient)

	Frequent Use of Formative Assessments	Use of School- Wide Behavior Policies	Instructional Hours in School Year
Intense Teacher Coaching	0.51***	0.31*	0.58***

Source: Principal survey.

** p< .10; **p<.05; ***p<.01.*

C. Ways of Categorizing CMOs

CMOs that tend to adopt one practice often adopt other practices that are consistent with their overall educational strategy. For example, as noted above, some CMOs that provide intensive coaching also employ formative assessments frequently to help teachers determine which skills or topics to prioritize. In addition to correlating pairs of practices, we used cluster analysis to explore whether CMOs can be categorized based on the mix of practices they employ; examining whether CMOs bundle certain practices may provide insight into their strategic approach. Below we discuss two ways to categorize CMOs, based on (1) the prescriptiveness of their policies and (2) the core practices that define our primary hypotheses regarding the drivers of impacts.[17]

1. Four groups of CMOs based on extent and form of CMO prescriptiveness

We explored the extent to which CMOs centralize decision-making or delegate authority to school principals using several items from the principal survey. Specifically, we used responses to survey questions that address whether CMO or school staff typically make key decisions relating to three broad areas: (1) curriculum/instructional approach, assessment, and professional development; (2) teacher evaluation and compensation; and (3) behavior policy.[18] With respect to these three broad dimensions of centralization, cluster analysis indicates that CMOs fall into four categories.[19] Group A CMOs tend to be very centralized in decision making only about policies and practices related to instruction (that is, educational approach, professional development, and formative assessments). Group B CMOs tend to be highly prescriptive or centralized in decision-making across all dimensions measured. Group C CMOs are very prescriptive with regard to behavior policy

[17] We conducted a hierarchical cluster procedure in SAS using Ward's minimum variance method, in which the distance between clusters is defined as the ANOVA sum of squares between two clusters, summed over the variables specified. At each step in the procedure, clusters from the previous step are merged so as to minimize the within-cluster sum of squares. Although we were interested in the potential grouping of CMOs on dimensions of prescriptiveness and school-level practices, we found no pattern of correlations of our defined core practices with prescriptiveness. Therefore, we explored CMO centralization as a separate domain for the purposes of our cluster analysis.

[18] We created five composite measures of prescriptiveness using principal survey items pertaining to decisions made or resources provided by the CMO to its schools: (1) CMO-prescribed educational approach, (2)centralized provision of professional development, (3) centralized policy on teacher evaluation and compensation, (4) CMO-prescribed formative assessment system, and (5) centralized behavior policy. From these composite measures, we developed three vectors of prescriptiveness or centralization using principal components analysis: (1) prescriptiveness on instructional policies, (2) centralized policy on teacher evaluation and compensation, and (3) centralized behavior policy.

[19] One outlying CMO does not appear to fit other patterns of prescriptiveness.

and decentralized elsewhere, on average. Group D CMOs appear to be very decentralized in decision making across all dimensions, on average.

These group typologies are depicted in Table III.4. Each row indicates one of the dimensions of principal-reported prescriptiveness, while the mean values and corresponding rankings across these dimensions are reported by the group in each column. The CMO and district comparison means are provided for reference. The column for Group A, for example, illustrates that those CMOs tend to be highly prescriptive and rank first in their degree of centralization of teacher evaluation criteria and compensation on average, while they are substantially less apt to mandate practices in the other areas measured.

Table III.4. Prescriptiveness Varies Across CMOs and Across Dimensions Within CMOs[20,21]

Prescriptiveness Components	Group A: Tight Instruction	Group B: Tight	Group C: Tight Evaluation and Compensation	Group D: Loose	CMO Mean	Comparison Mean
Instructional vector (educational approach, professional development, and formative assessments)	3rd (- 0.68)	1st (0.71)	2nd (0.25)	4th (- 2.07)	- 0.43	0.42
Teacher evaluation criteria and compensation	1st (0.42)	2nd (0.32)	3rd (- 1.07)	4th (- 1.96)	- 0.45	0.39
Behavior policy	2nd (- 0.44)	1st (1.21)	3rd (- 1.12)	4th (- 1.43)	- 0.31	0.31
N	13	5	8	4		

Source: Principal Survey.

2. Four clusters of CMOs defined by core practices

In addition to examining groupings of CMOs based on the tightness of their management approach, we also explored the extent to which CMOs fall into groups defined by their bundling of particular practices. Based on implementation of the core measures of school-level practices corresponding to six of our primary impact hypotheses, the cluster analysis suggests that CMOs can be categorized into four clusters:[22]

1. "Incremental Innovation" Cluster, the largest cluster, is composed of CMOs that deviate the least from the district means across the six core practices, on average.[23] This large group appears to include two subgroups that, while relatively similar on most

[20] The values for each of these composite measures are standardized across CMOs and district comparison groups such that the mean is 0 and the standard deviation is 1.

[21] The outlier CMO not included here has the following mean values: -0.98 on the instructional vector, -4.66 on teacher evaluation and compensation, and 1.87 on behavior policy.

[22] The tree diagram depicted the hierarchical clustering procedure indicated that the number of clusters could range from three to five. An examination of the means by cluster and the CMO values within each cluster indicated that a specification of four clusters was the most meaningful, with a combination of separation between clusters and consistency within clusters.

[23] That is, this cluster includes the fewest dimensions that deviate from the district mean by more than 1 standard deviation.

dimensions measured, diverge in the extent to which they centralize their educational approach.

2. "Data Driven" Cluster, the next largest cluster, is composed of CMOs that are most likely to implement performance-based compensation and evaluation and use formative assessment data most frequently. They also emphasize teacher coaching and tend to centralize instruction materials and methods.
3. "Time on Task" Cluster is composed of CMOs that tend to emphasize comprehensive behavior policies at the school level and require the most instructional hours for their students. In addition to ranking first in those dimensions, Table III.5 shows that they also tend to coach and monitor teachers most frequently.
4. "Alternative Approach" Cluster is composed of CMOs that tend not to consistently emphasize the key practices measured by the principal survey, ranking last among the clusters on these dimensions; these CMOs may focus on other strategies we have not captured.

The specific practices differentiating these clusters are shown in Table III.5. As in Table III.4, each row indicates one of the core practice measures, while the mean values and corresponding rankings across these dimensions are reported by the group in each column.

Table III.5. Core CMO Practices: Rankings and Means by Cluster

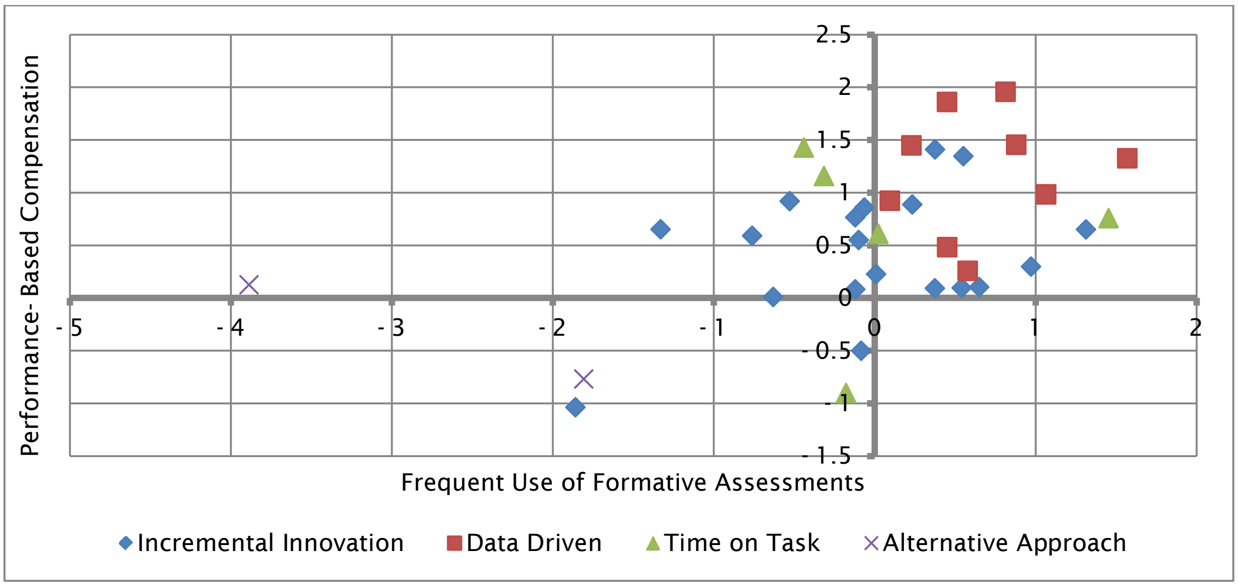
	Cluster 1 Incremental Innovation	Cluster 2 Data Driven	Cluster 3Time on Task	Cluster 4 Alternative Approach	CMO Mean	Comparison Mean
Consistent behavior policy	2nd (0.35)	3rd (0.16)	1st (1.26)	4th (- 2.02)	0.30	- 0.30
Frequent use of formative assessments	3rd (- 0.03)	1st (0.68)	2nd (0.11)	4th (- 2.85)	0.03	- 0.03
Consistent educational approach	3rd (- 0.47)	1st (0.64)	4th (- 1.36)	2nd (- 0.45)	- 0.31	0.30
Frequency of teacher coaching/mentoring	3rd (- 0.05)	2nd (0.91)	1st (1.14)	4th (- 1.55)	0.30	- 0.30
Performance- based compensation	3rd (0.42)	1st (1.19)	2nd (0.61)	4th (- 0.32)	0.60	- 0.62
Yearly instructional hours	3rd (- 0.16)	2nd (0.76)	1st (2.28)	4th (- 0.45)	0.41	- 0.42
Number of CMOs in cluster	19	9	5	2		

Source: Principal Survey.

Although Table III.5 shows that patterns in the implementation of key practices help to define the clusters of CMOs *on average*, scatter plots illustrate that individual CMOs within each cluster are indeed clearly differentiated on a few particular dimensions. In other words, these scatter plots offer evidence that, on the key dimensions, the means used to define the clusters are not masking substantially different individual CMO values within those clusters. Figures III.12, III.13, and III.14 show that clusters of CMOs are most differentiated in their emphasis on performance-based compensation and formative assessment use, behavior policy and instructional time, and teacher coaching and monitoring.

CMOs within the Data Driven cluster are consistent in their focus on formative assessment data and performance-based compensation. Figure III.12 plots CMO values on our composite measure of the use of formative assessment data versus values on our composite measure of the emphasis on performance-based compensation and evaluation; values farther to the right on the x-axis indicate greater frequency in reviewing formative assessment data, while values higher on the y-axis indicate a stronger emphasis on performance-based compensation and evaluation. The points shown as red squares depict the consistency of the Data Driven CMOs in emphasizing both practices. We can see CMOs from other clusters drifting to the upper right-hand corner of the plot, showing that an emphasis on these more data-focused practices is not entirely absent among other CMOs. However, the CMOs in the Data Driven cluster are most alike in their bundling of the dimensions shown.

Figure III.12. Data Driven CMOs Emphasize Frequent Use of Formative Assessment Data and Performance- Based Compensation

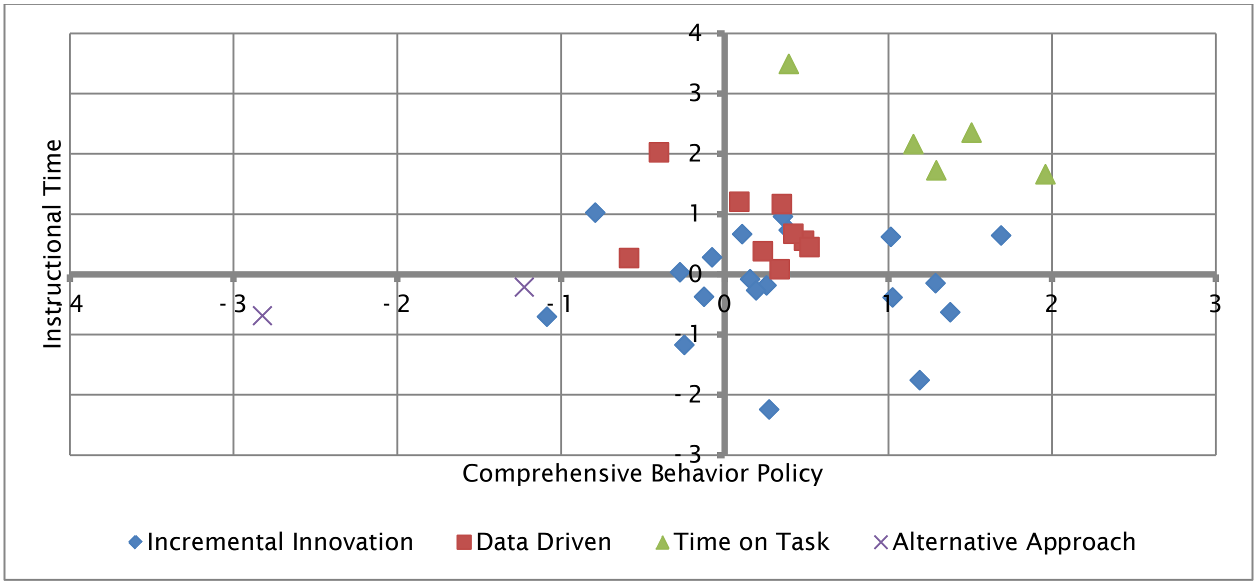


Source: Principal Survey.

CMOs in the Time on Task cluster are all consistently differentiated in their emphasis on comprehensive school-wide behavior policies and longer instruction time. Figure III.13 plots CMO values on our composite measure of a comprehensive school-level behavior policy versus values on instructional time. The Time on Task CMOs (identified by the green triangles) are concentrated in the upper right-hand quadrant, while CMOs in other clusters tend to fall below and to left.

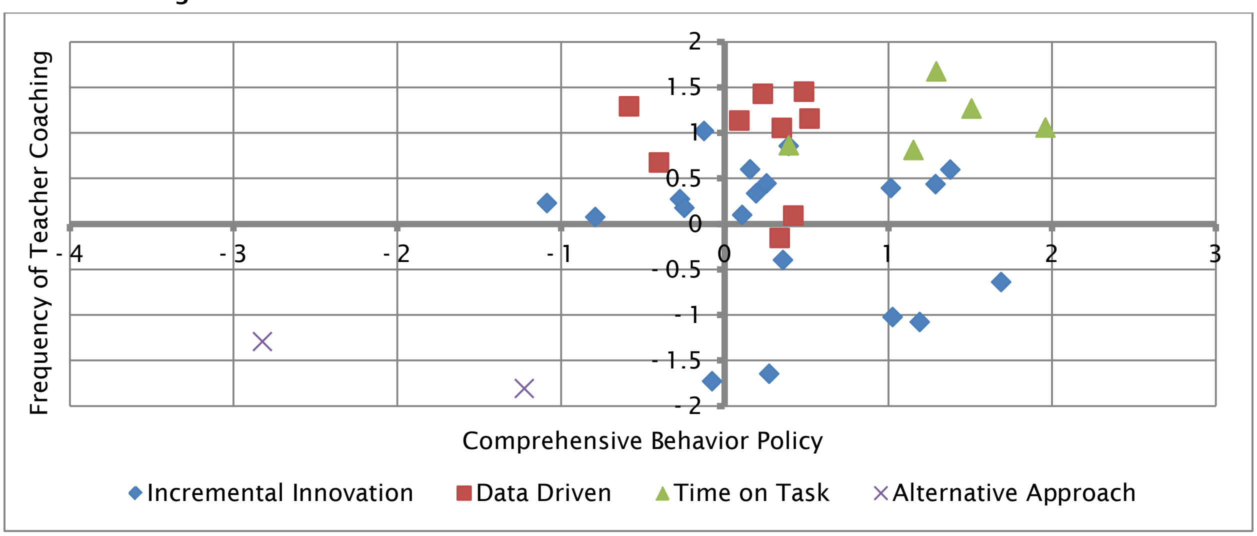
As noted earlier, both the Data Driven and Time on Task CMOs emphasize coaching. Figure III.14 plots CMO values on our composite measure of a comprehensive school-level behavior policy versus values on our composite measure of intensity of teacher coaching and monitoring. The red squares and green triangles identifying the Data Driven and Time on Task CMOs are arrayed towards the top because of the emphasis CMOs in these clusters place on teacher coaching. The Time on Task CMOs are concentrated to the right because of their emphasis on behavior policy.

Figure III.13. Time on Task CMOs Maximize Instructional Time and Emphasize Comprehensive Behavior Policies



Source: Principal Survey.

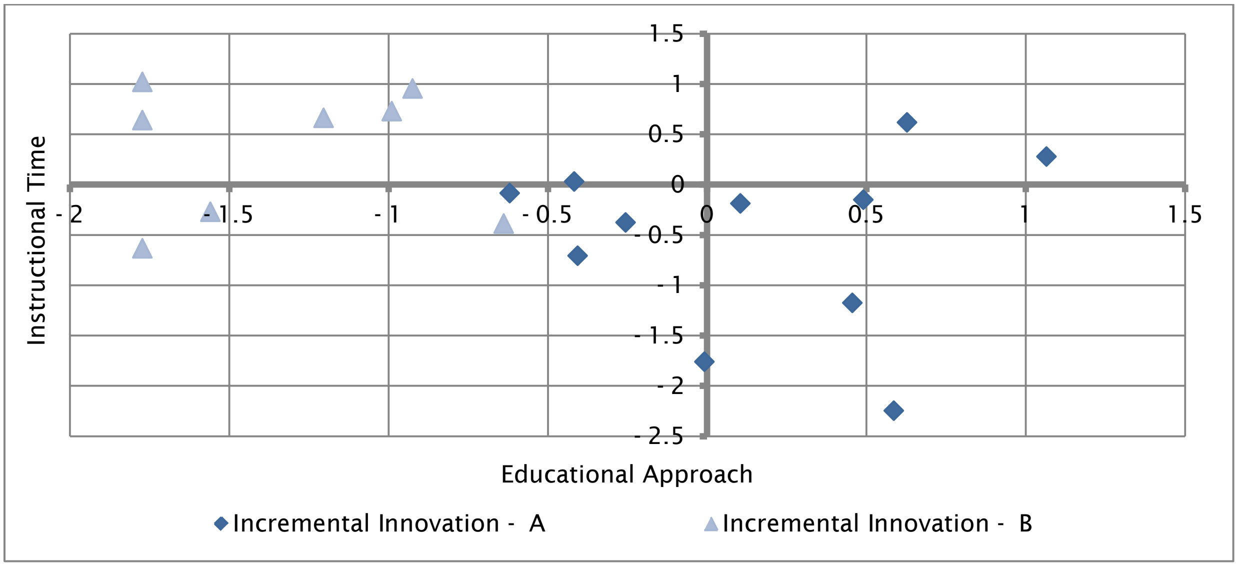
Figure III.14. Both Data Driven and Time on Task CMOs Engage in More Frequent Teacher Coaching and Monitoring



Source: Principal Survey.

Although the CMOs within the Data Driven and Time on Task clusters are very consistent in their emphasis on the dimensions that typify their clusters, there is substantially more variation within our largest cluster—the Incremental Innovation cluster. Our largest group of CMOs is least distinguishable on any practices; these CMOs deviate the least from districts on these dimensions, on average. That said, the plotting of specific data points for each CMO in the group indicates substantial variation within this cluster on one particular practice: the extent to which schools rely on CMOs to centralize instruction and curriculum. In Figure III.15, the subgroup designated as Incremental Innovation - B consists of CMOs that tend toward decentralization or more school-level autonomy in educational approach, in contrast to the CMOs in the Incremental Innovation - A subgroup, which take a more centralized CMO-level approach.

Figure III.15. Variation Within Incremental Innovation Cluster on Educational Approach



Source: Principal Survey.

A central question for this study is whether and how CMO practices contribute to greater student achievement. Chapter V examines the relationship between CMO practices and impacts on student achievement. The mechanism through which practices might affect students could be either direct (for example, extending the time allocated for academic instruction) or indirect (for example, coaching teachers on ways to collaborate with one another or make use of student assessments). In the next section, we examine two intermediate school level outcomes—instructional coherence and organizational health—that could mediate the indirect relationships between CMO practices and CMO impacts.

D. Instructional Coherence and Organizational Health of CMO Schools

If CMOs have a positive effect on student achievement, the effect must be mediated by something happening in their schools, which have direct contact with students. What is it about affiliation with a CMO that could increase school performance? Our descriptive analysis of CMO school practices suggests that at least some CMOs encourage or assist their schools in coaching and evaluating teachers, extending instructional time, and fashioning school-wide behavioral strategies. In addition, previous research indicates that new charter schools often struggle to unite all the parts of their instructional programs, and that principals are sometimes pulled away from instructional leadership by crises involving facilities, finance, compliance, and staffing (Hill and Rainey 2010).

At the outset of the study, we hypothesized that successful CMOs might address the challenges faced by new charter schools in at least two ways. First, CMOs could promote instructional coherence—helping schools ensure that the different aspects of their instructional program complement and reinforce one another. Some studies have suggested that instructional coherence is related to student achievement (Newman et al. 2001). Second, by reducing the administrative duties of principals and stabilizing the working environment for teachers, CMOs could increase schools' organizational health.

In order to examine instructional coherence and organizational health in CMO schools, we first developed measures of these outcomes using our survey data. The instructional coherence measures rely entirely on the teacher survey and hence are available only for the CMO schools. The organizational health measures, with one exception, rely on the principal survey, which allows us to compare CMO and district schools. Using these survey-based measures, we explored whether specific CMO practices are associated with higher levels of either instructional coherence or organizational health in schools.

Building on work done by the Consortium on Chicago Schools Research (Newman et al. 2001), we developed two multi-item measures of instructional coherence:

- **Use of a common instructional framework with consistent focus, pacing, standards, and formative student assessments**, based on teacher survey items addressing perceptions of the extent to which instruction and curriculum are aligned within and across grades, clarity and consistency of learning standards, and modification of lesson plans based on formative assessment data
- **A cooperative and supportive environment for teachers**, based on perceptions of staff cooperation and administrative support as reported in the teacher survey

We also developed and analyzed five sets of organizational health measures:

- **High teacher job satisfaction**, as reported in the teacher survey
- **Administrative staff stability**, based on principals' reports of turnover in the last three years and the need for recent leadership on an emergency basis
- **Average number of applications per open teaching position** during the past two years, as reported in the principal survey
- **Low student turnover**, based on school attrition rates as reported by principals
- **Minimal legal, financial, and other administrative challenges**, based on principal-reported time spent addressing issues related to finances/payroll, facilities leases, and building or grounds maintenance

In this section, we examine which CMO strategies and actions, if any, are correlated with instructional coherence and organizational health. We also explore differences between CMO and district schools on some of our measures of organizational health subcomponents.

1. Instructional coherence appears to be high when teachers are observed frequently

As established via our case studies, CMOs encourage intensive teacher monitoring and coaching, intended to help teachers implement the CMOs' classroom management strategies and focus instruction on skill areas identified as priorities via formative assessment of students.

A strategy of intensive teacher coaching and monitoring, as reported in our surveys, appears to be associated with high levels of instructional coherence. CMO-affiliated schools whose teachers report that they are observed frequently score high on both composite measures of instructional coherence, and schools whose teachers report receiving frequent guidance from their principals also score high on our measure of a collaborative and supportive environment. These findings are confirmed when we use measures of coaching from the principal and CMO central office staff surveys.

2. CMO principals appear to spend less time on administrative duties than do district principals

One pattern common in charter schools that CMOs seek to address is the need for principals to devote so much time to addressing administrative, compliance, and human resource management challenges that they cannot be instructional leaders. CMO central offices negotiate leases and manage compliance reporting in order to minimize principal burden. They also provide day to day help with hiring, personnel management, and data analysis, which is intended to allow principals to focus on classroom monitoring and assistance to teachers.

Indeed, surveys of CMO and district principals indicate that CMOs may be successful in minimizing administrative challenges for their principals. Compared to principals of district-run public schools serving similar students, CMO principals report that they spend less time resolving issues related to payroll and building and grounds maintenance (an average of less than once per month by CMO principals compared to one to five times per month by district principals, $p<.01$).

3. Principal turnover is lower where CMOs provide professional development and higher where CMOs prescribe the curriculum

Principal turnover is a major challenge for charter schools and can be highly disruptive to instruction, staff morale, and parent confidence. Some CMOs have sought ways to make the school leadership positions more manageable in order to minimize principal turnover. In addition, some CMOs have tried to build leadership pipelines and identify possible successors in an effort to preemptively reduce the impact of principal turnover.

CMO actions are linked to principal turnover, in both predictable and puzzling ways. CMOs that support their schools by providing access to professional development, including workshops or in-service training programs, have lower principal turnover compared to other CMOs, on average.

However, CMOs that are more prescriptive about the educational approach a school must take, as reported by principals, also experience more principal turnover, on average. This finding is one of a set of surprising results about the links between CMO actions and organizational health. Organizational health is negatively correlated with teacher coaching and professional development support provided by the CMO, frequency of CMO staff visits to the school, and the use of sample lessons in teacher hiring. These results may reflect a reversal of cause and effect: CMOs may undertake these actions when they believe a school is in trouble. Alternatively, the association could be driven by a third unobserved factor.

IV. CMO SCHOOLS' IMPACTS ON STUDENTS

Key Findings

- Impacts on student achievement and high school graduation vary a great deal across CMOs, ranging from substantially positive to substantially negative. Achievement and graduation impacts of individual CMOs are more often positive than negative, and in some instances are large.
- Among the four CMOs with college enrollment data, two have large positive impacts on enrollment; impacts for the other two are not distinguishable from those of nearby schools.
- The variation in school-level impacts is largely attributable to differences between rather than within CMOs, with some CMOs systematically outperforming others.
- Averaging across CMOs, overall impacts on all outcomes appear to be positive, but none are statistically significant.
- Several CMOs appear to have larger math and reading impacts for Hispanics than for other students, but impacts do not seem to vary appreciably for subgroups defined by gender, baseline test scores, or free and reduced-price lunch eligibility.

A. Introduction

As noted in the introductory chapter, an extensive body of research suggests that variation in the performance of charter schools is wide but that high-performing charter schools can produce substantial positive achievement effects for their students. CMOs represent an attempt to produce the effects of high-performing charter schools on a larger scale. To date, little research has examined their success in doing so.

In this chapter, we report our estimates of the effects of CMOs on test scores, high school graduation, and postsecondary enrollment, examining not only average effects across all CMOs but also the variation in effects among CMOs. Although much of the existing research on charter schools focuses on math and reading impacts, there is also interest in the test score impacts for other subjects that typically have not been the focus of high-stakes state accountability systems. We therefore estimate impacts on science and social studies tests (where available) as well as reading and math tests, and we examine the extent to which the impacts of CMOs are correlated across subjects.

The CMO (rather than the school) is the key unit of analysis for the study, and variation of impacts among CMOs is of particular interest given the variation in CMO practices reported in the survey results described in the preceding chapter. A key goal of CMOs is to produce consistent results across all of their schools, so we also examine within-CMO variation in impacts.

The rapid growth and substantial philanthropic investments in CMOs discussed in Chapter II beg the question: do successful CMOs attract more funding, which then leads to organizational growth? We test this hypothesis by looking at the associations between impacts and measures of CMO size and growth. A related question is whether CMOs are able to expand without compromising the effectiveness of both their existing and new schools. To answer this question, we conduct a series of within-CMO analyses to gauge whether individual CMOs' impacts tend to get larger (or smaller) as they grow.

Following up on prior studies of charter schools (Gleason et al. 2010; Angrist et al. 2011) that have found suggestive evidence of greater benefits for low-income minority students in urban areas, we examine whether CMOs benefit certain subgroups of students differentially. Finally, where data are available we estimate the impacts of CMOs on high school graduation and enrollment in college.

More specifically, this chapter addresses the following research questions:

1. How much do CMOs vary in their effects on student test scores in math, reading, science, and social studies?
2. What are the average effects of CMOs on student test scores in math, reading, science and social studies?
3. Do larger CMOs tend to have more substantial positive impacts?
4. Are CMOs able to sustain their impacts as they grow?
5. Are CMO test score impacts correlated across subjects within CMOs?
6. Do CMOs produce consistent impacts among their schools?
7. Are particular subgroups of CMO students differentially impacted by CMOs?
8. What are CMOs' impacts on high school graduation and college enrollment?

As we describe in detail in the methods section below, our ability to estimate CMO achievement effects depends on having "pre-treatment" test scores for CMO students and comparison students. This means, unfortunately, that we cannot estimate the achievement effects of CMO elementary schools. No data are available on the math and reading achievement of five-year-olds before they enter kindergarten. In most states, standardized testing does not begin until the end of third grade, a full four years after the beginning of elementary school. As a result, there are no proven and generalizable methods for estimating impacts of charter elementary schools. (Indeed, we are not aware of a good nonexperimental method for estimating impacts of *any* elementary schools—a problem that has not been sufficiently recognized by researchers or policymakers.)

High school impact analyses can, in contrast, make use of pre-entry achievement data, but they pose other challenges related to a limited number of standardized tests, high rates of grade repetition, and (in many places) tests that are course-specific rather than grade-specific. For these reasons, our analysis of CMO high school impacts focuses on two long-term attainment measures: high school graduation and enrollment in college.

We begin the chapter by describing our analysis sample, specifying the CMOs and the number of students included in our analyses, and defining the outcomes of interest. We then describe our approach to estimating impacts. We also discuss several threats to the validity of our impact estimates and how we address them. We then present impact findings in math, reading, science, and social studies achievement for middle school grades in 22 CMOs across the country, and conclude with a discussion of impacts on high school graduation and college entry for 6 CMOs.

B. Data and Scope of CMOs and Students Included

This chapter estimates impacts separately for CMO middle and high schools. The middle school analyses focus on achievement (test score) impacts, while the high school analyses focus on attainment (graduation and college entry) impacts. Among 26 CMOs across the country that have middle schools and met the eligibility criteria for the study (described in Chapter 1), the study obtained sufficient state and district school records data to estimate achievement impacts for 22 (85 percent). The achievement impact estimates cover 68 CMO middle schools (81 percent of all of the eligible middle schools).[1] The analyzed CMO schools span a total of eight states (including 16 metropolitan areas and two rural districts) located in the West, Southwest, Midwest, and Mid-Atlantic regions.

For a smaller number of CMOs, the study obtained data on high school graduation and college enrollment. Data on these outcomes were often not available, both because statewide data systems do not consistently track attainment outcomes at the student level and because many CMO high schools have not been operating long enough for us to observe students graduating and entering college. Among the 23 CMOs across the country that include high schools and met the study's eligibility criteria, we obtained sufficient data to estimate impacts on high school graduation for 6 CMOs and college enrollment for 4 CMOs. These attainment impact estimates cover 13 CMO high schools in three states, located in the West, Southwest, and Midwest regions. (We do not focus on test scores in high school because they are not consistently available and because they are often course-specific rather than grade-specific, creating a selection problem.)[2]

The impact analyses use school records data obtained from states and districts. The achievement outcomes are grade-specific statewide assessments in math, reading, science, and social studies. All pre-baseline, baseline, and outcome test scores were converted to standard deviation units, also known as z-scores.[3] All jurisdictions provided data on test scores (in at least reading and math and sometimes in science and/or social studies), race, gender, and school enrollment. In addition, a majority of jurisdictions also provided data on English language learner (ELL) status, special education status, and free and reduced-price lunch status. For the high school graduation outcome, administrative records were used to create an indicator variable showing whether students graduated within four years after entering grade 9. To create a college entry indicator, we used

[1] As discussed in Chapter I, CMOs were required to operate a total of at least four schools (at any grade level) to be included in the study. Some of the CMOs whose impacts are analyzed in this chapter operated fewer than four middle schools. Of the 22 CMOs in the middle school sample, estimates for 12 CMOs include three or more middle schools, 6 CMOs include two middle schools, and 4 CMOs include one middle school.

[2] Although data limitations prevented us from estimating achievement test impacts for most CMO high schools, we were able to secure high school test data for three CMOs; we report impacts on these outcomes in Appendix H.

[3] The z-score was calculated as the student's scaled score minus the mean scaled score for all students in a given population taking the test in the same year and grade and then divided by the standard deviation of scores for that same group. For 11 of the 22 middle school CMOs, z-scores were calculated using the statewide mean and standard deviation of scores for each test. In cases where the statewide mean and standard deviation could not be derived from the data or obtained through published sources, z-scores were calculated using the distribution of scores in the district-level data files provided to the study.

administrative data to determine whether students enrolled in a two- or four-year college within four years of their first grade 9 semester. We defined attainment using these four-year indicators (rather than definitions permitting graduation or college-entry five or more years after grade 9) to increase the number of student cohorts with observable data.

To be included in the analysis sample, CMO students had to enter the relevant CMO school in the normal intake grade,[4] have a baseline test score in at least reading or math, and have follow-up data on the relevant achievement or attainment outcome. The total number of CMOs, CMO students, and matched comparison students associated with each outcome is shown in Table IV.1.[5]

Table IV.1. Achievement and Attainment Analysis Sample Sizes for CMOs

Outcome	Measure	After 1 year of treatment	After 2 years of treatment	After 3 years of treatment	After 4 years of treatment
Math	Number of CMOs	22	22	14	N. A.
	Number of CMO students	18,606	13,434	5,747	N. A.
	Number of comparison students	321,296	237,490	121,050	N. A.
Reading	Number of CMOs	22	22	20	N. A.
	Number of CMO students	18,769	13,674	8,131	N. A.
	Number of comparison students	325,063	242,946	159,945	N. A.
Science	Number of CMOs	N. A.	N. A.	11	N. A.
	Number of CMO students	N. A.	N. A.	3,803	N. A.
	Number of comparison students	N. A.	N. A.	72,121	N. A.
Social Studies	Number of CMOs	N. A.	N. A.	9	N. A.
	Number of CMO students	N. A.	N. A.	3,529	N. A.
	Number of comparison students	N. A.	N. A.	69,751	N. A.
High School Graduation	Number of CMOs	N. A.	N. A.	N. A.	6
	Number of CMO students	N. A.	N. A.	N. A.	2,659
	Number of comparison students	N. A.	N. A.	N. A.	33,302
College Entry	Number of CMOs	N. A.	N. A.	N. A.	4
	Number of CMO students	N. A.	N. A.	N. A.	2,150
	Number of comparison students	N. A.	N. A.	N. A.	25,860

Source: State, district, and CMO school records.

[4] The analytic sample did not include CMO students who enrolled for the first time after the normal intake grade in a CMO school. However, for the sample of analyzed middle school and high school CMOs, students who enrolled in the normal intake grade consistently represent the majority of total student enrollment.

[5] Table IV.1 shows the number of unique CMO and matched comparison students analyzed for a given outcome. To account for data attrition, whenever a CMO student was missing data for a given outcome, all of the matched comparison observations for that student were also dropped from the sample for that outcome.

For middle school CMO impacts, we focus most of our attention on the cumulative achievement impacts for two years after CMO entry. These have the advantage of including more than just one year of treatment while still including all 22 CMOs. Science and social studies estimates, however, must be based on impacts after three years because most states do not test those subjects in earlier grades and so we cannot examine impacts after two years.[6] Similarly, data on high school graduation and college entry are only available at least four years after entry into grade 9. Because of this, the science, social studies, and attainment impacts are all based on smaller numbers of CMOs.

C. Methods for Estimating Impacts of CMOs

Producing rigorous and broadly applicable measures of the effects of CMOs (or of charter schools more generally) is methodologically challenging. The challenge is inherent in the intervention: Charter schools are by definition schools of choice, which means their students may differ from students in conventional public schools in ways that are not readily apparent (for example, in student or parent motivation). Failure to account for these differences could lead researchers astray and thus produce biased estimates of CMO impacts.

The best way to ensure the validity of impact estimates is to conduct a randomized experiment. Properly designed and implemented randomized experiments produce impact estimates that support stronger causal conclusions than any other method does by ensuring that the treatment and control groups are similar on all observed and unobserved characteristics prior to receiving an intervention. Thus, any significant difference between group outcomes can be attributed to the impact of the intervention. In the charter school context, we can sometimes perform randomized experiments using the lotteries that oversubscribed schools conduct to determine who will be admitted (see, e.g., Angrist et al., 2011; Dobbie and Fryer, 2011; Gleason et al., 2010; Hoxby and Murarka, 2009).

However, not all charter schools are oversubscribed, not all oversubscribed schools use lotteries, and not all schools using lotteries keep good records of winners and losers. In conducting several charter-school studies (see, for example, Gleason et al. 2010; Tuttle et al. 2011), Mathematica researchers have found that admissions lotteries can be used for experimental analysis in only a small proportion of charter schools nationwide. Of the 292 CMO schools initially targeted for inclusion in our study, data adequate for an experimental analysis were available in only 16 schools and only for select grades, cohorts, and achievement outcomes at those schools. Moreover, the oversubscribed schools where experimental analysis is possible might be different from the schools where it is not possible (Tuttle et al. 2010; Abdulkadiroglu et al. 2009). Oversubscribed charter schools might be oversubscribed because they are more effective than non-oversubscribed schools. In sum, admissions lotteries can be used to conduct experiments producing strong causal inferences about a small subset of CMO schools, but they cannot be used to examine the impacts of large numbers of CMO schools across the country. (Experimental impact estimates for the small subset of CMO schools and cohorts with usable lotteries are reported in Appendix B.)

Two primary aims of this study are to determine the average effectiveness of CMOs and to assess and understand the diversity of impacts among CMOs. To address these questions it is critical to estimate the effects of as many CMOs and CMO schools as possible. We therefore require a

[6] Most states in our sample administer grade-specific science and/or social studies tests in grade 8, although there are some states that also administer these tests in other middle and high school grades.

method that, unlike experimental or lottery-based methods, can be used for large numbers of schools. We use a nonexperimental panel (NXP) approach, which involves first identifying a matched comparison group of non-CMO students who are similar to the CMO students immediately prior to CMO enrollment and then following the achievement trajectories and attainment outcomes of individual students in both groups for several years. NXP methods for estimating CMO impacts rely on longitudinally-linked data on individual students before and after they enter CMO schools. Our preferred NXP approach, propensity score matching (described below), involves comparing the achievement of CMO students to non-CMO students who have been identified based on the similarity of their baseline achievement (prior to entering the CMO school) and other characteristics. Although NXP methods lack the strong causal validity of randomized experiments because matched students may differ on unobserved characteristics, they allow us to include many more schools and CMOs.

Test scores measured prior to CMO entry are critical to our NXP approach, especially when estimating test score impacts. Prior research in various topical areas has demonstrated that NXP methods can replicate the findings of randomized experimental studies if the researchers have a good pre-treatment measure of the outcome of interest (Glazerman et al. 2003; Cook et al. 2008). In addition, our study provided an opportunity to directly test the NXP methods against experimental test score impacts in the subset of CMO middle and high schools for which admissions lottery data are available and usable. The success of this test—described later in this chapter and in more detail in Appendix C—confirms that in the CMO context, nonexperimental methods can reproduce rigorous experimental test score impacts if they can make use of pre-treatment measures of relevant outcomes along with other student-level covariates that are widely available in administrative data.

Unfortunately, none of the CMO high schools with admissions lottery data have lottery records that go back far enough to observe graduating cohorts. As a result, we were not able to test our NXP approach against benchmark experimental attainment impacts. However, Deming et al (2011) find that NXP measures of school quality—using methods similar to those of our NXP analysis—do a good job of predicting experimental impacts on high school graduation and enrollment in four-year colleges (using data from oversubscribed high schools in Charlotte-Mecklenburg). This suggests that our NXP approach may produce internally valid attainment impact estimates similar to those generated using experimental methods.

1. Propensity score matching identifies a nonexperimental comparison group

To obtain a matched comparison group, we use a propensity score matching (PSM) procedure. The central concept of the PSM approach is to estimate the probability of being in the treatment group using observed data for treatment and potential comparison group students (Rosenbaum and Rubin 1983). This probability is known as the propensity score—here, the likelihood of enrolling in a CMO middle school. Theoretically, if the propensity score succeeds in removing all differences between the two groups that are related to test scores during the follow-up period, matching on that propensity score would result in an unbiased estimate of the impact of treatment. We use various student characteristics available in school records—most prominently, a student's prior test scores—to estimate each student's propensity to enter a CMO middle or high school. For details on the propensity model selection, see Appendix D.

After estimating the propensity scores, the next step is to select a matched comparison group of students whose estimated propensity scores are similar to those of treatment group students. To improve statistical precision, we selected multiple matches for each treatment student, thus increasing the total sample size. To ensure the quality of the matches and reduce bias, we matched with replacement (allowing each comparison student to match to more than one CMO student) and implemented caliper matching, whereby a given treatment student is matched to all comparison students with estimated propensity scores within a specified range (caliper) rather than merely a specified number of nearest neighbors. The matching procedure is implemented separately for each grade, cohort, and district combination. For each CMO, we were able to match between 64 percent and 100 percent of CMO enrollees in the primary intake grade who had at least one valid outcome, with a match rate of at least 90 percent for most of the test scores and attainment measures that we rely on as key outcomes.[7] We were able to obtain matched comparison groups of students who were equivalent in terms of baseline test scores, gender, race, and free or reduced-price lunch status to students in the treatment groups. [8] (For details on the matching procedure see Appendix D.)

2. Statistical regression controls for remaining differences in estimating impacts

Following the creation of matched samples, we employ a regression model when estimating impacts to improve statistical precision and to control for any remaining differences in baseline characteristics. Our CMO-specific impact regression models include pre-baseline (i.e., two grades prior to CMO entry) and baseline test scores in reading and math; corresponding missing test score indicators; and other student characteristics, including race/ethnicity, poverty status (as measured by eligibility for free or reduced-price lunches), disability status, and ELL status.[9]

The impact for the average CMO is calculated by averaging the CMO-specific impacts for a given outcome, and by treating the CMOs in our analysis sample as representing a broader population of CMOs. We also examine whether CMO-specific impacts vary by student's sex, race/ethnicity, eligibility for free or reduced-price lunch, prior achievement, and the number of schools that CMO operates in a given year by including an interaction between the subgroup indicator and treatment group status in the CMO-specific impact model.

3. We account for selective attrition and grade repetition

Even if PSM successfully identifies comparison students who are similar to CMO students in relevant respects prior to CMO entry, estimates of impact in the long term could be biased if

[7] Match rates were 90 percent or better for 17 of 22 CMOs on two-year reading and math outcomes, for 8 of 11 CMOs on the three-year science outcome, and for 6 of 9 CMOs on the three-year social studies outcome. Match rates were also above 90 percent for 5 of 6 CMOs with graduation data and all 4 CMOs with postsecondary enrollment data.

[8] The sample of analyzed CMOs and CMO students varies by outcome and outcome year for two reasons. First, some cohorts are observed for only one or two follow-up years. As a result, the sample for three-year impacts includes fewer student cohorts than do the samples for one-year or two-year impacts. Second, individual students may have missing outcome data for other reasons. Approximately five percent of the middle-school sample observed in the first follow-up year is missing data in the second year, and eight percent is missing data in the third year on average. When a CMO student disappears from the data, the matched comparison students are also dropped (and vice versa). As a result, the study's CMO and non-CMO samples maintain baseline equivalence in their pre-entry test scores for all years and outcomes, and most maintain equivalence on race, gender, and free and reduced-price lunch status as well (see Appendix D for a more detailed description of baseline equivalence results for both middle schools and high schools).

[9] Not all student variables are available in all jurisdictions. We use weights that adjust for the distribution of treatment and comparison students across matching strata that are defined by grade, cohort, and district so that groups of students belonging to each unique grade, cohort, and district combination contribute equally to our impact estimates across treatment and comparison groups. We use robust standard errors that adjust for clustering of students within schools.

students who are struggling academically are more likely to exit the CMO schools. We address this possibility by keeping all CMO students permanently assigned to the treatment group even if they exit the CMO after their first year (or later). This is analogous to an "intent-to-treat" analysis in the experimental context.[10] In essence, we avoid the possibility of selective attrition by preventing transferring students from leaving the treatment condition. This means that our impact estimates will be conservative, in the sense that they underestimate the full effect of CMO enrollment on the students who remain enrolled.

Student grade repetition also creates analytical challenges: A student who repeats a grade takes different state end-of-year tests than other students in her/his original cohort take. Most state assessments are not vertically scaled, which prevents us from using observed test scores in a given academic year to compare the achievement of retained students to students in the same cohort who were promoted. If a CMO has a policy of retaining low-performing students at a higher rate than the other public schools, ignoring the attrition of students due to grade retention may result in biased CMO impacts. Following Tuttle et al. (2010), we address this issue by assuming that the retained students will perform at the same level relative to other students in their cohort as in the year prior to being retained. If a CMO has a higher rate of grade repetition and positive impacts (and if the impacts are positive even for the grade repeaters), this assumption will tend to underestimate true impacts. For more information on grade retention, see Appendix F.

4. The matching method successfully replicates rigorous experimental test score impact estimates

As mentioned earlier, we validated our PSM approach against benchmark experimental test score estimates in the same CMOs. This validation effort, conducted for a subset of CMO middle and high schools in which experimental analyses could also be conducted, finds that PSM produces test score impact estimates that are very similar to the benchmark experimental test score impact estimates.[11] In both reading and math, PSM estimates are statistically indistinguishable from experimental estimates, and the point estimates differ by very small amounts (0.01 to 0.03 standard deviation units). Site-specific estimates across seven lottery sites indicate that PSM results correlate with experimental results at levels of 0.9 or higher. These results provide evidence supporting the rigor of nationwide NXP/PSM test score impact estimates for the larger set of CMO schools where experimental methods are not feasible (see Appendix C).

As noted above, for the high school analysis of attainment outcomes, we were not able to validate our NXP approach against benchmark experimental impacts. However, we conducted a sensitivity analysis of our attainment impacts using an alternate method that addresses selection by comparing CMO students to other students who chose to enroll in charter schools during middle or high school grades (following the example of Booker et al. 2011), with the rationale that students who ever enrolled in a charter school may be more similar to one another than to students who have never chosen to enroll in a charter school. The sensitivity analysis produced results that differed in some cases from our primary estimates, but the overall average impacts on high school graduation

---

[10] The analogy is imperfect because some students may exit the CMO schools in their first year of enrollment, prior to testing and before we capture them in the treatment group. In general, we cannot identify transfers that occur prior to testing.

[11] Refer to Appendix B Section III.B for more details on how the diversity of CMOs in the validation effort compares with the diversity of CMOs included in the impact analyses.

and college enrollment remained statistically insignificant (see Appendix G for results and estimation details.)

D. CMOs' Impacts on Middle School Test Scores

1. After two years of enrollment, CMO impacts on students' reading and math achievement are more often positive than negative

The extent of variation in CMO impacts is potentially important for two reasons. First, policymakers and funders are interested in the extent to which CMOs show uniformly positive impacts because this could inform whether and how they seek to encourage the development of the most successful CMOs. Second, variation in impacts provides a basis for exploring factors that might contribute to this variation, a topic examined in the next chapter.

The test score impacts discussed in this chapter are cumulative impacts. In other words, two-year middle school CMO impacts, for example, refer to the total CMO impact on students who entered the CMO middle school two years earlier, and not to the incremental impact of the second year of CMO enrollment. We focus most of our attention on the two-year impacts, because two years is the longest period we can examine with the full sample of 22 CMOs.

In both reading and math, after two years of enrollment in a CMO middle school, positive impacts are more common than negative impacts. Of the CMOs covered by the impact analysis, half (11 of the 22) have positive impacts in math or reading while nine have negative impacts in one or both subjects (see Table IV.2). Ten of the CMOs have positive impacts in both subjects while only 4 have negative impacts in both subjects.

Table IV.2. Number of CMOs with Significant Positive and Negative Impacts in Math and Reading

	Math Impacts		
Reading Impacts	Significant Positive	Insignificant	Significant Negative
Significant Positive	10	0	0
Insignificant	1	2	3
Significant Negative	0	2	4

Source: State, district, and CMO school records.

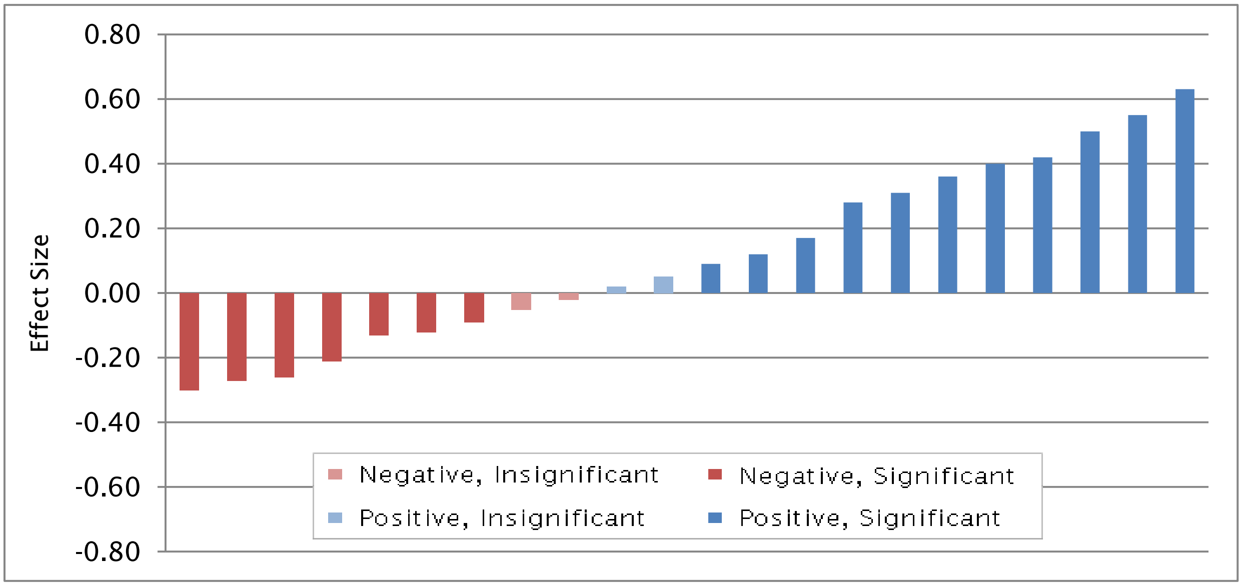
2. The range of achievement impacts for individual CMOs is wide, especially in math, for which the positive impacts of the highest performing CMOs are large

In addition to counting the number of CMOs with positive or negative impacts, it is also important to examine the size of those impacts. In specifying the size of impacts, we first use the typical approach of calibrating them in terms of the overall standard deviation in test scores in the entire state or district. Figure IV.1 below shows the distribution of estimated two-year math impacts, and Figure IV.2 shows the corresponding distribution for reading.[12] Each of the 22 bars in the figure

[12] We adjusted for multiple comparisons using the Benjamini-Hochberg correction and found that all impacts that were statistically significant remained significant post-adjustment (see Appendix F for details). To test if these impacts

represents a single CMO impact, ordered left to right, from the most negative to the most positive. Statistically significant (insignificant) impacts are illustrated using darker (lighter) shades of red (negative impacts) and blue (positive impacts). One-year and three-year impacts for math and reading can be found in Appendix H.

Figure IV.1. Distribution of Test- Score Effect Sizes After Two Years in Math



Source: State, district, and CMO school records.

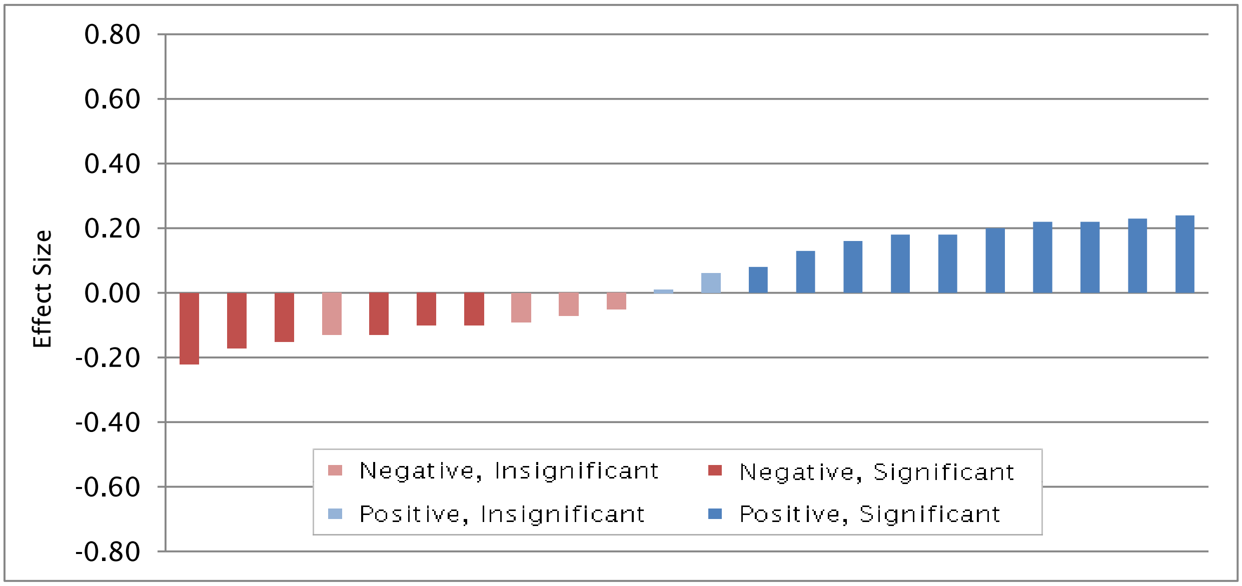
Note: "Significant" is defined as statistically significant at the .05 level, two-tailed test.

For the lowest performing CMOs, two-year[13] impacts are between -0.2 and -0.3 of a standard deviation in both math and reading. But the largest positive (and statistically significant) impacts in math exceed 0.6 of a standard deviation, twice the size of the negative math impacts of the lowest performing CMOs. In addition, these positive impacts for math are more than twice as large as the largest positive impacts in reading. Larger impacts in math than in reading were also observed in recent studies of charter schools such as the KIPP Lynn study (Angrist et al. 2010), the KIPP middle school study (Tuttle et al. 2010), the Harlem Children's Zone Promise Academy study (Dobbie and Fryer 2011), the Boston charter schools study (Abdulkadiroglu et al. 2009), and the New York City charter schools study (Hoxby et al. 2009).

(continued)
are homogenous, we performed a Q-test and were able to reject the null of homogeneity (Q-statistic is 1043.4 and 457.3 for math and reading, respectively).

[13] The CMO-level impacts in both subjects are highly correlated across years (0.89 between one-year and two-year math impacts; 0.94 between two-year and three-year math impacts).

Figure IV.2. Distribution of Test- Score Effect Sizes After Two Years in Reading



Source: State, district, and CMO school records.

Note: "Significant" is defined as statistically significant at the .05 level, two-tailed test.

3. The differences between high- and low-performing CMOs are large enough to produce substantial differences in student outcomes

To get a better sense of the variation and magnitude of the CMO achievement impacts, one can compare them to several other policy-relevant benchmarks. Data from the National Assessment of Educational Progress (reported in Bloom et al. 2008) show black-white and Hispanic-white achievement gaps ranging from 0.76 to 1.04 standard deviations in math and reading in grades 4 and 8. These numbers suggest that the CMOs at the high end of the scale have the potential to measurably reduce achievement gaps, especially in math. Another relevant benchmark is the typical effect of a year of schooling. A few of the CMOs are producing impacts that appear to be sufficient to generate three years of learning gains within two years (Bloom et al. 2008).[14] Those CMOs could be seen as producing three years worth of learning over just two years in the classroom. By the same token, the lowest performing CMOs are producing negative achievement effects that are nearly as large as the effect of a year of schooling—that is, their students have achieved not much more than one year of learning after two years in the classroom.

All of the impact estimates are, of necessity, measured against the performance of other local public schools. Our ability to make useful comparisons among CMOs is therefore constrained because each of them has a different counterfactual: Some of the variation in impacts among CMOs might be related to variation in the quality of the comparison schools rather than in the CMOs' own

[14] Note that the effect sizes in Bloom et al. (2008) are based on a national population, whereas the effect sizes in this study are based on a single state or district. Because test score standard deviations are likely to be bigger nationwide than at the state or district level, our impact sizes may be slightly inflated as compared to those in Bloom et al. (2008). Bloom's estimates indicate that the impacts required to achieve an extra year of learning within two years are about 0.2 to 0.3 in reading and about 0.3 to 0.4 in math at middle school grades.

performance. There is no way to directly test this possibility because we can't relocate CMOs to different communities. Nonetheless, we can examine the extent to which CMO impacts appear to be driven primarily by the achievement trajectories of CMO students or by the trajectories of comparison students.[15] It is theoretically possible that CMO students are experiencing achievement trajectories that are consistent across CMOs and that the variance in impacts is attributable to large variations in the trajectories of comparison students. If so, this would complicate our interpretation of the large variation in CMO impacts.

It appears that most of the variation in CMO impacts is attributable to variation in CMO students' own achievement trajectories (adjusted via regression for their characteristics). For two-year middle school math impacts, for example, the net achievement gains of CMO students range from -0.26 to 0.55, whereas the net achievement gains of the matched comparison groups (for each CMO) fall in a narrower range around zero, from -0.22 at the low end to 0.10 at the high end. Consequently, the estimated impact of each CMO is correlated very highly (at 0.95) with the adjusted average achievement gain of its students. Meanwhile, the correlation of CMO impacts with the adjusted achievement gains of comparison students is much closer to zero, at -0.21.[16] Thus it appears that the wide variation in test score impacts among CMOs is largely due to variation across CMOs rather than being primarily driven by local context.

4. Estimated CMO achievement effects are broadly consistent with effects measured for other charter schools

In a few jurisdictions, we were able to conduct parallel analyses of the achievement effects of independent charter schools (that is, charter schools not affiliated with a CMO) alongside our CMO analyses.[17] We found no patterns in the relative impacts of independent charter schools versus CMOs across the jurisdictions. Results varied across CMOs and across jurisdictions: Some CMOs outperformed independent charters; others did not. In general, the magnitude of impacts estimated for independent charters was in the same range as the impacts we estimated for CMOs. Details are available in Appendix I.

The most rigorous recent studies of charter schools have reported a range of achievement impacts. If one combines these studies, the span of impacts reported is broadly consistent with the range of our CMO impact estimates:

- In a national sample of oversubscribed charter middle schools, Gleason et al. (2010) found school-level variation in impacts ranging from -0.43 to 0.33 in reading and from -0.78 to 0.65 in math after two years.

---

[15] However one must be cautious in interpreting these findings since the CMO impacts are based in part on the baseline test scores of students entering CMOs from district schools. Hence the CMO impacts incorporate some of the variation in the quality of district schools.

[16] The correlation of CMO impacts with the achievement *levels* of comparison students is likewise low.

[17] These comparisons relied on ordinary least squares (OLS) regression analyses rather than propensity match analyses because OLS analyses also successfully replicated experimental estimates in our validation exercise and because they can be conducted with far less labor and computing time.

- A study of New York City charter schools estimated annual impacts which, if accumulated over two years, would imply effect sizes of 0.12 in reading and 0.18 in math (Hoxby et al. 2009).
- In oversubscribed charter middle schools in Massachusetts, Angrist et al. (2011) found annual impacts which, if accumulated over two years, would amount to 0.13 in reading and 0.42 in math.

Similarly, a study of 22 KIPP middle schools (Tuttle et al. 2010) examined the variability in achievement impacts among schools. KIPP is not itself a CMO, because it does not have direct operational authority over schools; some of KIPP's regions, however, are CMOs. The schools examined in the KIPP study would be in the upper half of the distribution of CMO impacts found here, but the magnitudes are broadly comparable. Two-year impacts for the 22 KIPP schools in reading ranged from -.12 to .43, whereas two-year impacts in math ranged from zero to 0.75.

5. Among the CMOs in our study, large CMOs are more likely than smaller ones to have positive impacts

We looked at whether there is a relationship between CMO size, as measured by the number of schools operated by the CMO in fall 2009, and two-year math and reading impacts. Figure IV.3 shows the distribution of estimated two-year math and reading impacts, respectively, on the x and y-axes, while the size of the bubbles represents CMO size.

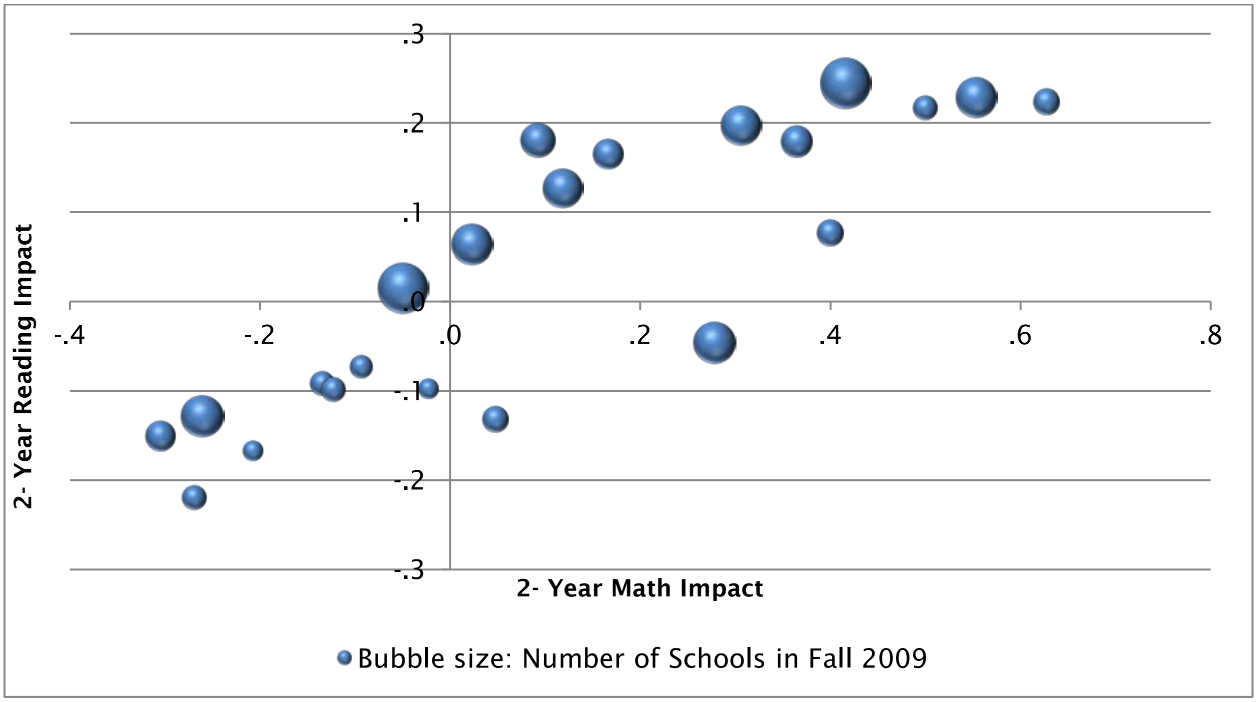
Large CMOs in our sample tend to have positive impacts, while small CMOs are more likely to have negative impacts. This might indicate that funders have had some success in supporting the expansion of CMOs that are more effective. In particular, eight of the 12 large CMOs (those operating more than 8 schools in 2009-10) have significant positive impacts in at least one subject, while only 3 of the 10 small CMOs (those operating 8 or fewer schools in 2009-10) have significant positive impacts in at least one subject. Meanwhile, only 2 of 12 large CMOs have significantly negative impacts in at least one subject, while 7 of 10 small CMOs have significantly negative impacts in at least one subject.[18] CMOs that have positive impacts in both reading and math operate an average of 12 schools, while those with negative impacts in both subjects operate an average of 6 schools. Despite this pattern, effectiveness is not related to size in a linear way: Correlations between math and reading CMO impacts and CMO size are not statistically significant.

As a result, among all of the elementary, middle, and high schools operated by these 22 CMOs in fall 2009, most are in CMOs with positive impacts in math or reading (at least at the middle school level). Approximately 57 percent of these schools are managed by CMOs with significant positive middle school impacts in either math or reading. Only 26 percent of the schools are managed by CMOs with significant negative impacts in either subject, and the remaining 17 percent

[18] Among the 12 large CMOs, 7 have significant positive impacts in both subjects, 1 has significant negative impacts in both subjects, and 4 have a mix of significant and insignificant impacts (both positive and negative) in both subjects. As for the 10 small CMOs, 3 have significant positive impacts in both subjects, 3 have significant negative impacts in both subjects, and 4 have a significant negative impact in 1 subject and an insignificant impact in the other subject.

are affiliated with CMOs with impacts that are not statistically distinguishable, positively or negatively, from those of nearby schools.[19]

Figure IV.3. Comparing Test Score Effect Sizes After Two Years in Math and Reading and CMO Size



Source: State, district, and CMO school records.

We also looked at whether absolute CMO growth (change in the number of schools operated by the CMO between fall 2004 and fall 2009) and relative CMO growth (the number of schools operated by the CMO in fall 2009 divided by the number of schools operated by the CMO in fall 2004) are associated with two-year impacts in math and reading. In both of these cross-sectional analyses, we found no statistically significant associations.

6. In many CMOs, reading impacts decline as the CMO adds more schools; math impacts do not consistently decline with growth

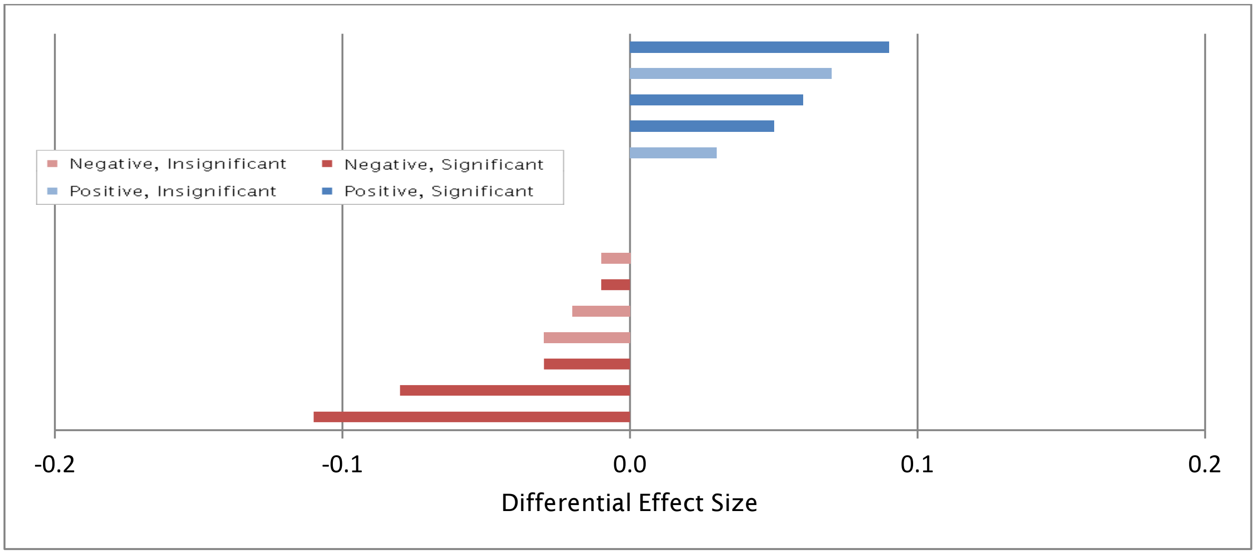
In order to explore the question of whether CMOs are able to expand and maintain their effectiveness on student test scores, we examined within-CMO changes over time in size and impacts. Specifically we gauged whether, as individual CMOs grow, their impacts tend to get larger (or smaller). To be eligible for this analysis, we require that CMOs have at least three different numbers of schools during the years covered by the analysis data. This eligibility criterion excluded 7 out of the 22 CMOs in our sample.

---

[19] Ignoring statistical significance, approximately two thirds of these CMO schools are managed by CMOs with positive math impacts and the same fraction are managed by CMOs with positive reading impacts. One obtains a similar breakdown if one focuses exclusively on the schools covered by the impact analysis rather than all schools affiliated with these CMOs (see section 6 below).

Based on the results of these longitudinal analyses, we conclude that some CMOs show declining impacts as they grow while others do not. Our within-CMO analyses suggest that CMO expansion often diminishes student impacts in reading: we found smaller reading impacts due to an additional school for nine out of twelve CMOs with statistically significant differential impacts. However, there is no clear pattern of changes in math impacts as CMOs grow. Figures IV.4 and IV.5 show the difference in magnitudes of two-year impacts due to an additional CMO school for math and reading, respectively.

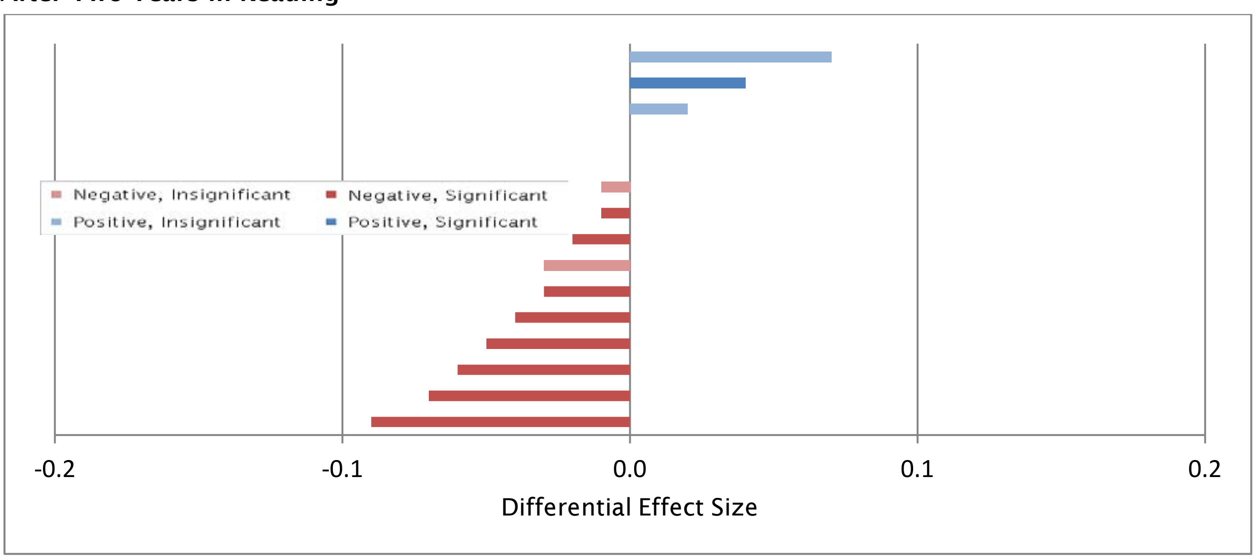
Figure IV.4. Distribution of *Differential* Test Score Effect Sizes Due to an Additional CMO School After Two Years in Math



Source: State, district, and CMO school records.

Note: "Significant" is defined as statistically significant at the 0.05 level, two-tailed test.

Figure IV.5. Distribution of *Differential* Test Score Effect Sizes Due to an Additional CMO School After Two Years in Reading



Source: State, district, and CMO school records.

Note: "Significant" is defined as statistically significant at the 0.05 level, two-tailed test.

Nonetheless, despite the attenuation of reading impacts for many CMOs as they grow, it remains true that the larger CMOs in our sample more often have positive impacts, as noted above. In other words, successful CMOs may not fully sustain the impacts they produce in their first schools, but even after their impacts decline with growth, they tend to remain positive.

7. The variation in school-level impacts is mostly due to differences between CMOs

One of the primary aims of CMOs is to promote consistent results across multiple schools. Our analysis of middle school achievement impacts sheds light on the extent to which the CMOs are achieving this objective. We estimated impacts separately for each CMO middle school. Specifically, we calculated each school's effect size after two years in reading and math, relative to the CMO's overall matched comparison group (see Appendix D for a description of the school-level impact estimation methods). In total, we obtained school-level impacts for 67 of the 68 middle schools included in the analysis of student achievement.[20]

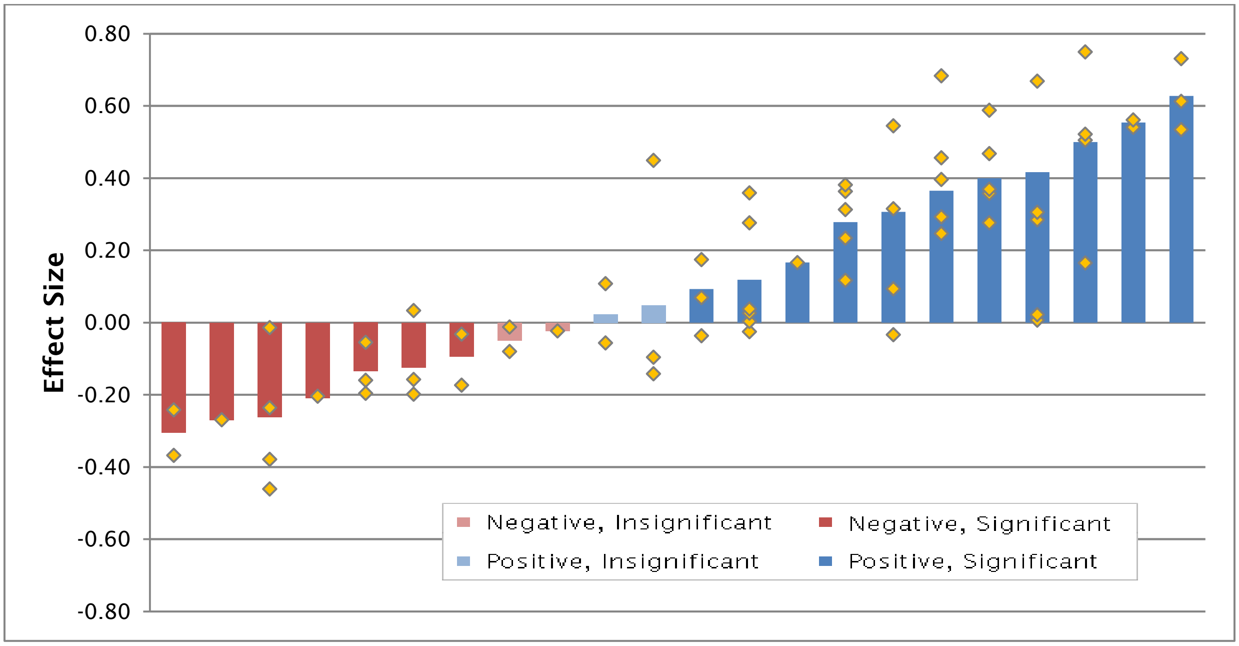
Most of these schools have positive impacts. About 64 percent have positive effects in math, and 52 percent have effects that are both positive and statistically significant. Similarly, in reading, 66 percent of schools have positive effects and 40 percent have effects that are significantly positive.

Most of the variation in school-level impacts occurs between rather than within CMOs. Figure IV.6 (for math) and Figure IV.7 (for reading) plot our point-estimate for each school's impact (a yellow diamond) against the overall average impact of that school's CMO (a blue or red vertical bar). As the figures indicate, within a CMO, there is often variation in the impacts of individual schools. However, this within-CMO variation is substantially smaller than the variation *between* CMOs. Specifically, in math we find that the between-CMO differences account for 87 percent of the variation in school-level impacts (with the remaining 13 percent being due to within-CMO variation in school impacts). In reading, CMO-level impacts account for 73 percent of the variation in school-level impacts.[21] In other words, after between-CMO variation is taken into account, differences within CMOs account for less than one seventh of the school-level variation in math impacts and less than a third of the school-level variation in reading.

[20] For one CMO middle school, the sample size of enrollees was too small to estimate a reliable school-level impact.

[21] These percentages refer to intra-cluster correlation coefficients (ICCs), which compare the variation in CMO-level impacts to the total variation in school-level impacts. For a discussion of how ICCs were calculated, see Appendix C.

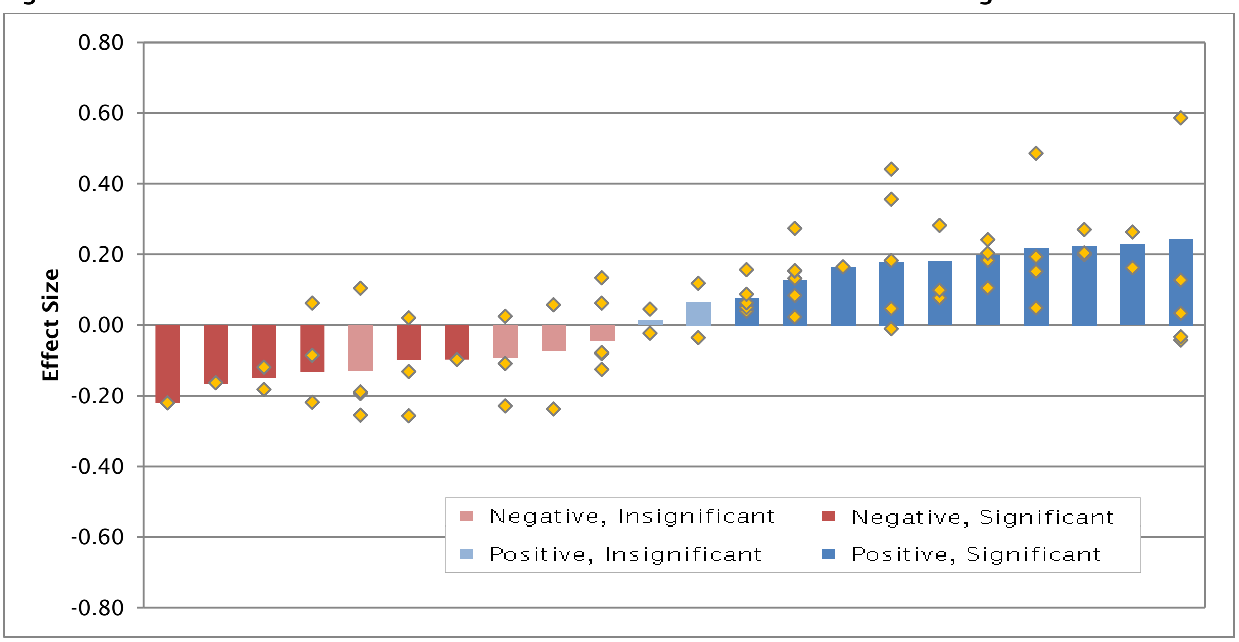
Figure IV.6. Distribution of School- Level Effect Sizes After Two Years in Math



Source: State, district, and CMO school records.

Note: Each yellow diamond shows the effect size of one middle school managed by the CMO whose total impact is represented by the associated vertical bar. "Significant" is defined as statistically significant at the 0.05 level, two-tailed test.

Figure IV.7. Distribution of School- Level Effect Sizes After Two Years in Reading



Source: State, district, and CMO school records.

Note: Each yellow diamond shows the effect size of one middle school managed by the CMO whose total impact is represented by the associated vertical bar. "Significant" is defined as statistically significant at the 0.05 level, two-tailed test.

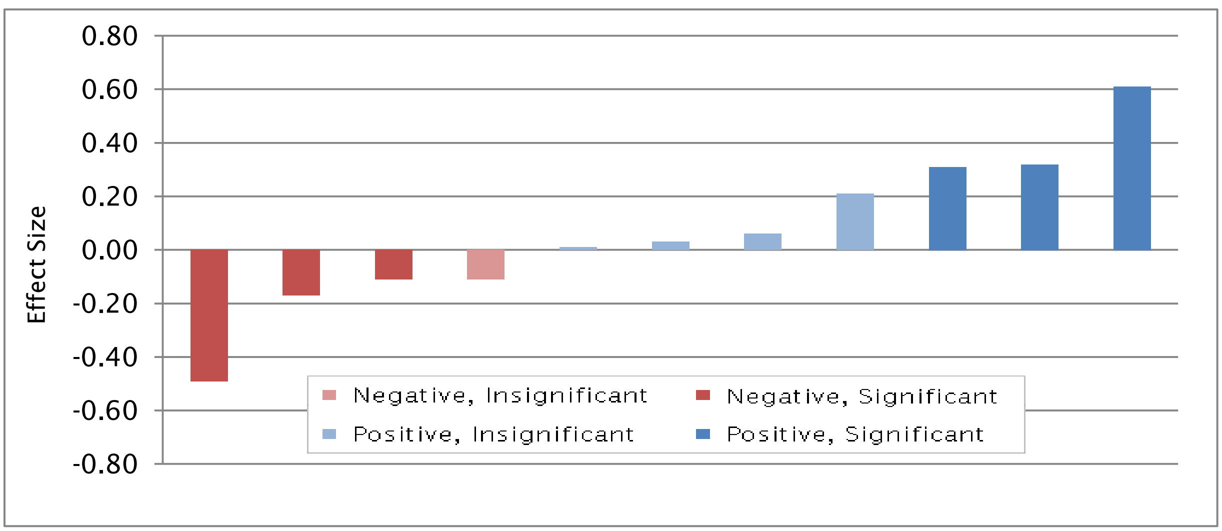
We can also use school-level results to assess whether *individual* CMOs are successful in producing consistently positive (or negative) impacts in their schools. In our middle school sample, there are 18 CMOs with estimates for two or more schools. In math, 7 of these 18 CMOs have uniformly positive school-level math impacts, 5 have uniformly negative impacts, and 6 have mixed effects.[22] In reading, 7 CMOs have uniformly positive reading effects, 1 has uniformly negative impacts, and 10 have mixed results.

8. CMOs also show substantial variation in impacts on science and social studies tests

States typically do not require science and social studies assessments in every grade. CMO impacts in science and social studies can most frequently be examined three years after students enroll in a CMO middle school and only for a subset of our CMO sample—11 CMOs for science and 9 CMOs for social studies. Nonetheless this is a large enough sample to provide useful information about the range of impacts generated by many CMOs.

As in math and reading, there is substantial variation in CMO impacts on science and social studies achievement three years after enrollment. Figures IV.8 and IV.9 below show the distribution of estimated three-year test score impacts for science and social studies, respectively. For science, the number of CMOs with significant positive impacts is equal to the number of CMOs with significant negative impacts (three each), and effect sizes range between -0.49 and 0.61. Estimated impacts for three-year social studies are positive and statistically significant for five out of the nine CMOs where we were able to estimate social studies impacts; only one CMO had a significantly negative impact on a social studies assessment. Three-year social studies impacts range between -0.48 and 0.41.

Figure IV.8. Distribution of Test Score Effect Sizes After Three Years in Science

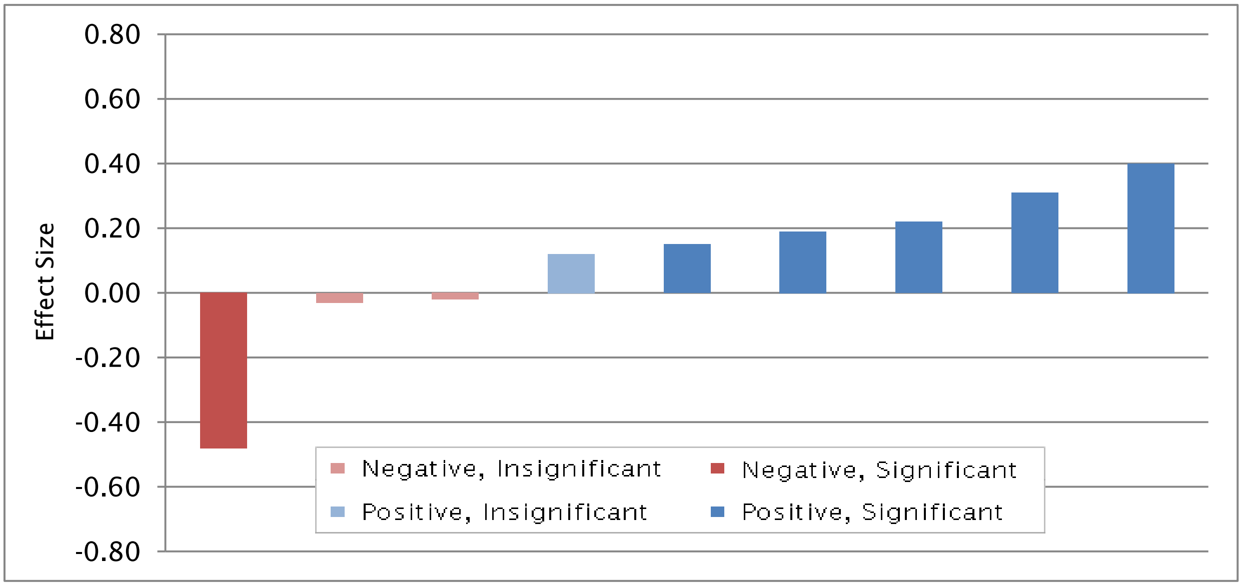


Source: State, district, and CMO school records.

Note: "Significant" is defined as statistically significant at the .05 level, two-tailed test.

[22] These counts disregard the statistical significance of each school-level effect.

Figure IV.9. Distribution of Test Score Effect Sizes After Three Years in Social Studies



Source: State, district, and CMO school records.

Note: "Significant" is defined as statistically significant at the .05 level, two-tailed test.

9. Although overall average two- and three-year test score impacts are positive in all four subjects, they are not statistically significant

We also estimated the impact of the "average" CMO on achievement in each of the four subjects. Average impacts provide a sense of the overall contribution of CMOs to efforts to improve student achievement. With only 22 CMOs in the sample (and fewer for science and social studies impacts), however, average effects would need to be substantial in order to achieve statistical significance.

Estimated middle school test score impacts for the average CMO are presented in Table IV.3 below. Although average two-year impacts are positive, they are not statistically significant at the five percent level. The average CMO's two-year math impact is 0.11 and is marginally significant (p=0.08). Average impacts across CMOs for two-year reading and three-year science and social studies are positive and not statistically significant.

If we were to estimate the effect of these CMOs on the average student (rather than the average CMO impact), thereby giving the larger CMOs more weight, the math impact would be significantly positive. Because the larger CMOs tend to have larger and more positive impacts, the cross-CMO impact on the average student is larger than the impact of the average CMO. Furthermore, if the impact of these specific CMOs on the average student is the question of interest, then the standard errors of the impacts are smaller, making the impacts more statistically significant.[23] (We do not

[23] When estimating the impact of these CMOs on their average student, the standard errors are driven largely by the large number of students in the analysis rather than the small number of CMOs. In this case we are focusing exclusively on the CMOs in the sample rather than attempting to generalize to a broader universe of CMOs.

focus on impacts for the average student here because the study is about CMOs, making it important to be able to generalize results to other CMOs not included in the sample. Most of our analysis therefore focuses on the effect of the average CMO).

Estimated middle school test score impacts for the average CMO are presented in Table IV.3 below. Although average two-year impacts are positive, they are not statistically significant at the five percent level. The average CMO's two-year math impact is 0.11 and is marginally significant (p=0.08). Average impacts across CMOs for two-year reading and three-year science and social studies are positive and not statistically significant.

Table IV.3. Average CMO Test Score Impacts, by Year After CMO Enrollment

	1-Year Impact	2-Year Impact	3-Year Impact
Math Effect Size	0.06	0.11^	0.15
(Standard Error)	(0.05)	(0.06)	(0.09)
Number of CMOs	22	22	14
Reading Effect Size	-0.01	0.03	0.05
(Standard Error)	(0.02)	(0.03)	(0.04)
Number of CMOs	22	22	20
Science Effect Size	N.A.	N.A.	0.06
(Standard Error)			(0.09)
Number of CMOs			11
Social Studies Effect Size	N.A.	N.A.	0.09
(Standard Error)			(0.09)
Number of CMOs			9

Source: State, district, and CMO school records.

^Significantly different from zero at the 0.10 level, two-tailed test.

10. Impacts are highly positively correlated within CMOs among academic subjects

Another question of interest is whether certain CMOs produce positive impacts in only certain subjects or whether CMOs tend to produce impacts of similar direction and magnitude across all subjects. Do CMOs appear to focus their efforts on increasing achievement in some subjects—such as math and reading, which often receive more attention from policymakers—at the expense of others?

We find that CMOs that produce positive impacts in one subject also tend to produce positive impacts in other subjects. The correlations are as high as 0.86 for (two-year) math and reading impacts; the lowest correlation is between (two-year) math impacts and (three-year) social studies impacts at 0.60. This suggests that CMOs either do not place more emphasis on particular subjects, or if they do, that there are positive spillover effects from skills students acquire for one subject to another. Table IV.4 contains the within-CMO correlations of impacts across the four subjects we examined for the years with the best data coverage.

Table IV.4. Within-CMO Correlations of Impacts Across Math, Reading, Science, and Social Studies

	2-Year Math	2-Year Reading	3-Year Science	3-Year Social Studies
2-Year Math	1.00 (N=22)			
2-Year Reading	0.86 (N=22)	1.00 (N=22)		
3-Year Science	0.81 (N=11)	0.82 (N=11)	1.00 (N=11)	
3-Year Social Studies	0.60 (N=9)	0.73 (N=9)	0.79 (N=9)	1.00 (N=9)

Source: State, district, and CMO school records.

11. CMO impacts do not generally differ by subgroup, but several CMOs have larger two-year math and reading impacts for Hispanic students

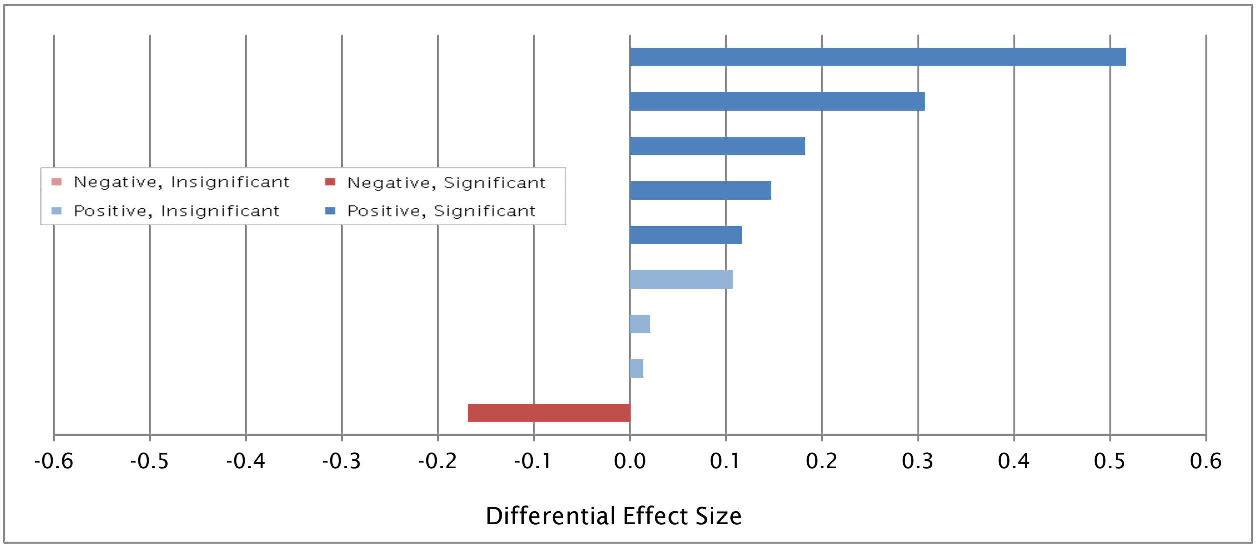
Prompted by prior studies of charter schools (Angrist et al. 2011; Gleason et al. 2010) that have found suggestive evidence of greater benefits for low-income minority students in urban areas, we examined whether two-year math and reading impacts were different for particular subgroups of students. The subgroups we focused on were Hispanic students, African American students, low-achieving students[24] at baseline, students receiving free and reduced-price lunch, and male students. We were not able to examine all of these subgroups across our sample of 22 CMOs either because of data restrictions (we did not have free and reduced-price lunch information for all our jurisdictions) or because the student population in a particular CMO's analysis sample was too homogenous.[25] Appendix J contains the impact estimates for each of these subgroups.

There is some evidence of larger two-year math and reading impacts for Hispanic students. Figure IV.10 shows the difference in the magnitude of two-year math impacts between Hispanic students and all other students in each of the nine CMOs where we were able to estimate subgroup impacts for Hispanics. We found larger math impacts for Hispanic students for five out of six CMOs with statistically significant differential impacts. We found similar results for Hispanic students in reading. Other subgroups did not show clear patterns of differential impacts, positively or negatively, in reading or in math.

[24] We define low-achieving students as those who had a baseline test score below the 50th percentile of students in the relevant district or state.

[25] We require between 15 and 85 percent of students in our analysis sample to be part of a subgroup of interest (that is, Hispanic, African American, low achieving, students receiving free or reduced-price lunch, or male) before that particular subgroup is eligible for analysis.

Figure IV.10. Distribution of *Differential* Test Score Effect Sizes for Hispanic Students Relative to Effects for All Other Students After Two Years in Math



Source: State, district, and CMO school records.

Note: "Significant" is defined as statistically significant at the .05 level, two-tailed test.

E. High School Graduation and Postsecondary Enrollment Impacts

In addition to examining CMOs' impacts on middle school test scores, we also estimate the longer-term impacts of CMO high schools on graduation and postsecondary enrollment for a subset of CMOs.[26] For many students, graduating from high school or enrolling in college may ultimately represent a more important outcome than the academic achievement measured by test scores. While data on these attainment outcomes were available only in a small number of CMOs, we still find substantial variation in the direction and magnitude of CMO effects, and the magnitude of CMO impacts on these outcomes is sometimes large.

The attainment effects discussed in this section represent percentage-point impacts on the likelihood of either graduating from high school or enrolling in college. Specifically, each estimate refers to the total CMO impact on students who enrolled in a CMO high school in grade 9, four years before the graduation or college enrollment outcome was observed. As discussed in Section C, each impact was calculated using a regression model controlling for remaining differences between CMO students and the propensity-score-identified matched comparison group.[27] For further discussion of these estimation methods—including results of an alternate attainment analysis that

[26] Due to data limitations, we could not consistently distinguish between enrollment in different types of postsecondary programs in all jurisdictions. As a result, our postsecondary enrollment outcome variable indicates whether a student was recorded as enrolling in any postsecondary program within four years of the semester they first entered grade 9.

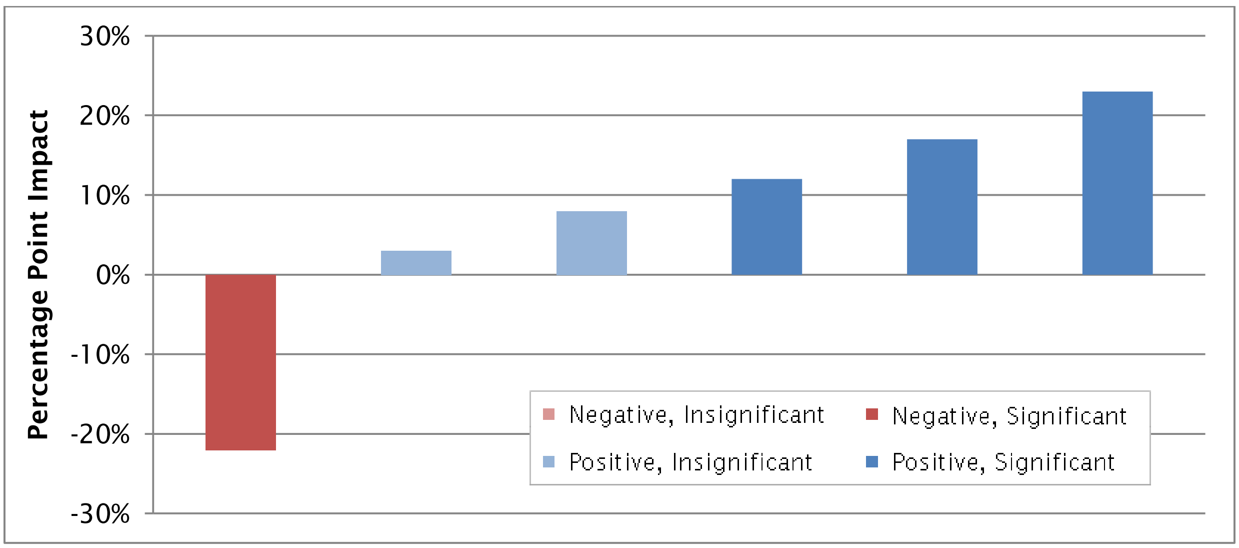
[27] Impacts for these binary attainment outcomes were calculated using logistic (logit) regressions. Each CMO's estimated regression model was used to calculate a percentage point impact, defined as the difference between the mean model-predicted graduation or college enrollment rate of CMO students and matched comparison students. We also calculated percentage point impacts using ordinary least squares (OLS) regressions for the same sample of CMOs and matched comparison students: in all cases, the OLS estimates were nearly identical to the logit estimates shown here. For additional details, see Appendix C.

limited the analyzed sample to students who attended charter schools—see Appendix G. Results of the alternate analysis differ somewhat from those reported below for individual CMOs, but the overall average impacts remain statistically insignificant, regardless of which method is used.

1. Among the six CMOs where high school graduation data are available, three have large positive impacts on graduation, one has a large negative impact, and two have insignificant impacts

For high school graduation, CMOs' impacts are more often positive than negative. Figure IV.11 shows the distribution of impacts for the six CMOs with graduation data. Impacts at half of the analyzed CMOs are positive and statistically significant, and only one CMO has a negative impact on graduation (in the remaining two cases, impacts were positive but insignificant). The magnitude of some of these impacts is substantial, ranging from positive 23 percentage points to negative 22 percentage points. Put differently, at the extremes we find that one CMO increases by 23 percentage points the likelihood of ninth graders graduating from high school four years later, whereas a different CMO reduces its ninth graders' chances of graduating by 22 percentage points.

Figure IV.11. Distribution of Percentage Point Impacts on High School Graduation



Source: State, district, and CMO school records.

Note: "Significant" is defined as statistically significant at the .05 level, two-tailed test.

2. Among the four CMOs where college enrollment data are available, two have large positive impacts on enrollment and the other two have insignificant impacts

CMO staff often place a high priority on preparing students for college. Indeed, nearly 9 out of 10 surveyed CMO principals said that helping students prepare for college is very important.[28] While we were able to secure college enrollment data for only four of the CMOs in the study, the results are nonetheless useful in gauging these CMOs' success.

[28] Nearly all CMO principals also said that helping students exceed state academic standards is very important.

The two CMOs with the largest impacts on high school graduation also have significant, positive impacts on enrollment in a two-year or four-year college (Table IV.5). The magnitudes of these effects are large: 21 percentage points and 23 percentage points, respectively. Two other CMOs had insignificant impacts on postsecondary enrollment; one of these CMOs had a significant positive impact on high school graduation while the other had an insignificant positive impact on graduation. (No college enrollment data were available for either the CMO that has large negative impacts on high school graduation or the remaining CMOs with an insignificant positive effect on graduation.)

Table IV.5. CMO Impacts on High School Graduation and Postsecondary Enrollment (In Percentage Points)

	High School Graduation	Postsecondary Enrollment
CMO #1	23%**	23%**
CMO #2	17%**	21%**
CMO #3	12%*	3%
CMO #4	8%	4%
CMO #5	3%	N/A
CMO #6	-22%**	N/A
CMO Average	**7%**	**13%**

Source: State and district school records.

*Significantly different from zero at the .05 level, two-tailed test.
**Significantly different from zero at the .01 level, two-tailed test.

3. For CMOs with data, average graduation and college enrollment impacts are positive but not significant

There is relatively little past research on the attainment impacts of charter schools. One study examining impacts in Chicago and Florida suggests that charter schools may have positive effects on both high school graduation and college attendance (Booker et al. 2011). Since we have only a limited sample of CMOs with the relevant data available, we are not able to provide comprehensive assessment of the overall average CMO impact on these attainment outcomes. Even so, the average impacts for the CMOs with available data are a useful indicator of the efficacy of some CMOs.

The average impacts of these CMOs on the attainment outcomes are positive but not statistically significant. As shown in Table IV.5, the average CMO in our sample has a 7 percentage point impact on high school graduation, but the effect is not statistically significant at the five percent level. For college entry, on average we find a positive impact of 13 percentage points, but the mean impact is not significant (p=0.10) for the limited number of CMOs in the sample. These average impacts correspond reasonably well with those reported in the study of Chicago and Florida charter schools, which found a 7 to 15 percentage point impact on high school graduation and an 8 to 10 percentage point impact on college attendance (Booker et al. 2011).

We have ninth grade high school test score impact estimates alongside high school attainment impacts for a small subset of three CMOs. For these three CMOs, impacts on test scores do not always correspond to impacts on attainment (see Appendix G for details). This finding is consistent with two other studies which found discrepancies between achievement impacts and attainment impacts for some charter high schools and voucher programs (Booker et al. 2011; Wolf et al. 2010). Although we cannot draw strong conclusions from a sample of three CMOs, the discrepancies in test-score impacts and attainment impacts in this small sample suggests the potential importance of examining attainment outcomes alongside test scores in future studies, to obtain a more complete picture of charter schools' long-term effects on students.

V. STRUCTURES AND PRACTICES ASSOCIATED WITH STUDENT IMPACTS

Key Findings

- Among CMOs, school-wide behavior policies and intensive coaching of new teachers are positively associated with student impacts in both math and reading.
- At the CMO level, we do not find impacts to be associated with use of a uniform curriculum, extended instructional hours, frequent formative student assessment, or performance-based compensation.
- Intensive teacher coaching in CMOs may increase student achievement in part by increasing the frequency with which teachers modify their lesson plans using the results of student assessments.

A. Introduction

Might the variation in the practices of CMOs explain why CMO impacts vary substantially? To inform policy and practice, we examine the associations between CMO practices and impacts, with the aim of identifying practices that are associated with larger positive achievement effects on students. We also explore whether other factors, including CMO size and growth and state charter policies, are associated with impacts.

The results in this chapter are best considered exploratory. Most of the analysis is based on bivariate associations between a single CMO-level practice and student impacts in math or reading. Although we also conduct some multivariate regression analyses of the relationship between impacts and several practices, multicollinearity and limited sample sizes impede our ability to parse out the importance of each practice based on its association with impacts. And any observed associations between practices and impacts could be driven by other, unmeasured factors that are correlated with both the practice and impacts. Hence it is not possible to make causal inferences on the basis of these analyses.

The analysis is conducted at the CMO level, reflecting our primary interest in the policies and practices of CMOs rather than of individual schools. The analysis covers the six primary hypotheses, discussed in Chapter III, of associations between CMO practices and impacts. Appendix Table K.1 shows CMO-specific values for each of these measures, along with estimated impacts and baseline test scores. We also explore a longer list of 43 secondary hypotheses that relate CMO policies and practices to student impacts. These 43 secondary hypotheses include proposed mediators of the association between our primary hypotheses measures and impacts, alternative measures of our primary hypotheses, and other practices not captured by our primary hypotheses. All of the primary hypotheses and nearly all of the secondary hypotheses were defined before conducting the impact analysis. We performed multiple comparison adjustments for our primary hypotheses to test the robustness of our main results accounting for multiple tests.

Results for secondary hypotheses should be considered especially exploratory, given the large number of them. We provide complete results in Appendix Table K.2, Appendix Table K.3, and Appendix Table K.4 for all secondary hypotheses, but in the main text of this chapter, we focus predominantly on those secondary hypotheses can shed light on the findings related to primary hypotheses. For secondary hypotheses, we do not adjust for multiple comparisons.

In the remainder of the chapter, we first describe the methods and data in more detail. We begin our discussion of the results by presenting the associations between each primary hypothesis measure and impacts in math and reading. Next we summarize how impacts are related to the main primary hypotheses. We then elaborate findings related to the three primary hypothesis measures that are significantly associated with student impacts: school-wide behavioral policies, intensive teacher coaching, and extended instructional time. The discussion of the teacher coaching findings also includes a summary of our analysis of the interrelationship among coaching, measures of instructional coherence, and impacts. We then summarize our findings related to a number of staffing practices, including the use of Teach For America (TFA) and Teaching Fellow teachers. Finally, we present average impacts for the categories of CMOs discussed in Chapter III, including those defined by the primary hypotheses and by the prescriptiveness of CMOs. The methods employed in the chapter are summarized in Appendix J and the detailed results from all the analyses are included in Appendix K.

B. Methods Overview

Our primary results make use of bivariate ordinary least squares (OLS) models with robust standard errors to gauge the correlation between student impacts and CMO practices (see Appendix J for more detail). The outcome variables are the estimated two-year middle school impacts of the CMO on students' achievement in reading and math. The independent variables are measures of our primary and secondary hypotheses, the majority of which are constructed from responses to the principal survey (N = 19 CMOs, 294 principals). For practices measured by the principal survey, the responses of CMO principals and those of principals in matched district comparison schools are differenced to construct a measure of disparity of practices between CMO schools and nearby district schools. For some secondary hypotheses we make use of measures constructed from responses to the CMO central office staff survey (N = 17 CMOs) and the teacher survey (N = 12 CMOs, 384 teachers). For these items, there are no comparison measures from districts or district schools.

Any significant bivariate associations between CMO practices and student impacts may be spurious. A practice that appears to be positively associated with impacts may simply be correlated with other practices that are the real drivers of student outcomes. In addition to bivariate associations, we conducted a multivariate analysis that includes any measures of primary hypotheses that were significantly associated with CMO impacts in bivariate models. Because our sample size is limited and collinearity among practices inflates standard errors in multivariate models, we must also remain cautious in drawing conclusions that practices that are not significantly associated with impacts in multivariate models are necessarily not related to student impacts. To balance the concerns raised by both bivariate and multivariate models, we present the results of both for our primary hypotheses.

C. Overview of Primary Hypotheses

Below we focus on the significant associations between CMOs' impacts and two of our six primary hypotheses (see Table V.1). Comprehensive school-wide behavior policies and an emphasis on coaching new teachers are at least marginally significantly associated with positive impacts in both subjects.[1]

Table V.1. Correlations Between Six Primary CMO Practices and Impacts

	Math	Reading
Consistent educational approach	-0.08 (0.12)	-0.04 (0.06)
Comprehensive behavior policy	0.18** (0.05)	0.08* (0.03)
Emphasize formative assessment	0.15 (0.12)	0.07 (0.05)
Emphasize intensive teacher coaching	0.19* (0.07)	0.08^ (0.04)
Emphasize performance-based compensation	-0.02 (0.10)	0.01 (0.05)
Instructional hours per year	0.15^ (0.07)	0.07 (0.04)

Source: State, district, and CMO school records, and Principal Survey

^Significantly different from zero at the .10 level, two-tailed test.
*Significantly different from zero at the .05 level, two-tailed test.
**Significantly different from zero at the .01 level, two-tailed test.

We found no significant relationship between impacts and three of our primary hypotheses measures: consistent educational approach, use of formative assessment, or performance-based compensation. Consistent educational approach was measured by whether all schools in the CMO use a single curricular model, whether the CMO provides support for selecting books and other instructional materials, and whether the CMO is responsible for selecting curricula and instructional materials. The lack of association between a centralized educational model and CMO impacts suggests that effective CMOs need not implement a common, centralized curriculum across all schools. The lack of association between performance-based teacher compensation and student impacts is consistent with some previous research (Fryer, 2011; Glazerman and Seifullah, 2010).

[1] We supplemented our main, CMO-level analysis with school-level tests of our six primary hypotheses. In this analysis, we found significant associations between school-level behavior policy and school-level student impacts in both subjects, but no other statistically significant associations (see Appendix Table K.5). The weaker school-level associations between practices and impacts may be due to measurement error that attenuates the coefficients in the school-level analysis but has a lesser effect on the aggregate CMO-level measures. The school-level association between coaching and impacts may also be reduced compared to the CMO-level association if CMOs choose to allocate teacher coaching resources to schools in the CMO that are struggling; this "negative selection" of schools receiving coaching could offset any positive effect of coaching on student achievement.

Frequent review of student test results by teachers, principals, and CMO central office staff is also not significantly associated with student impacts. By itself, frequent testing of students may not translate to gains in student achievement. As discussed later, student impacts do tend to be higher when teachers frequently revise their teaching plans in response to the results of assessment data, suggesting that student assessment data may be useful only to the extent that it is associated with changes in instructional practices.

The marginally significant association between school-wide instructional hours and math impacts appears to be due to the fact that CMOs with more instructional hours also emphasize teacher coaching and school-wide behavior policies (see Appendix Table K.6). In the multivariate model, the magnitude of the associations between instructional hours and impacts is close to zero in both subjects. In comparison to school-wide behavior policies and intensive teacher coaching, the association between instructional hours and student impacts is not robust.

D. Findings

1. Comprehensive behavior policies are positively associated with student impacts

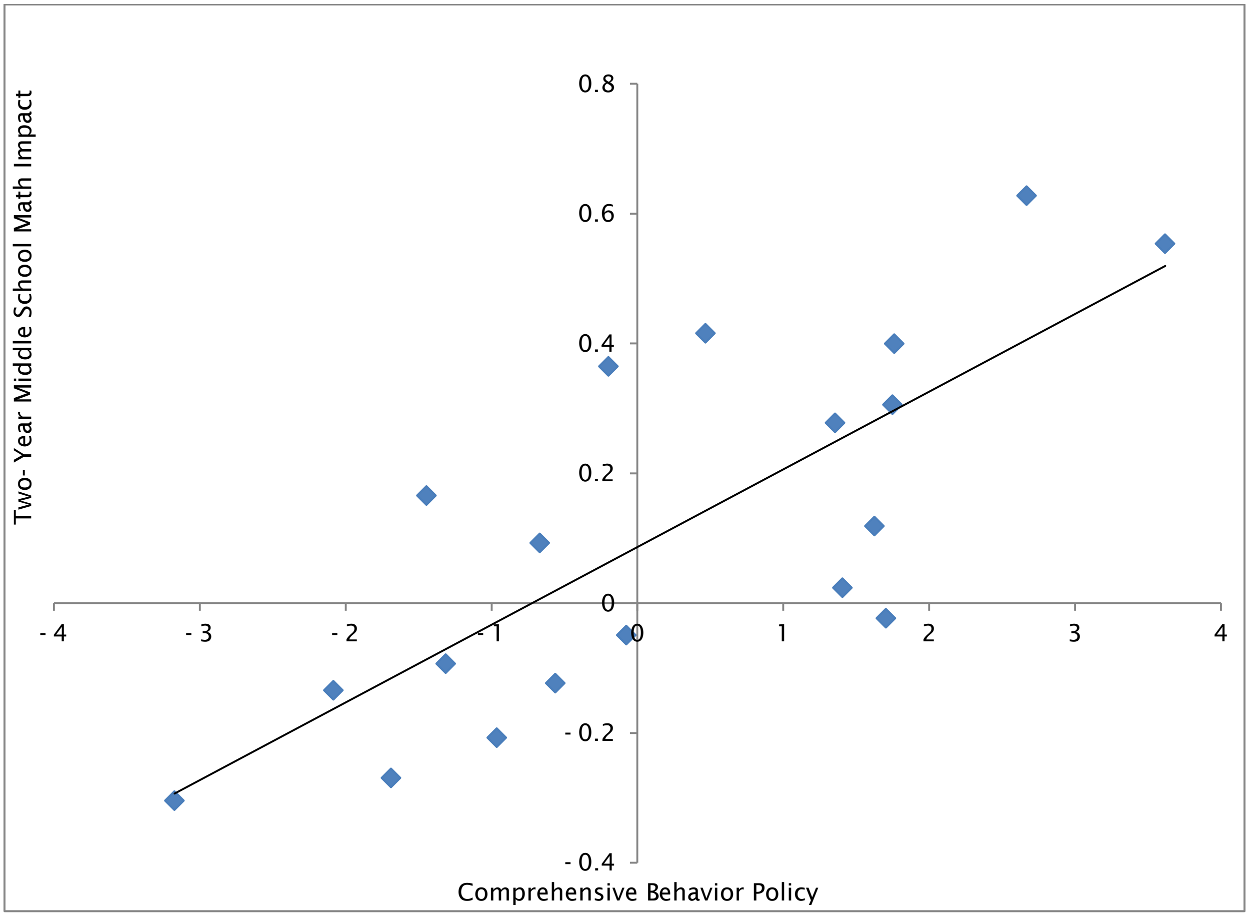
Behavior policies have the potential to affect student achievement if they encourage students to focus, reduce the amount of disruptions, and increase time on task. As discussed in Chapter III, comprehensive behavior policies within schools were measured by an index that combines principals' reports on five issues: (1) whether consistent behavioral standards and disciplinary policies are enforced, (2) whether schools have zero-tolerance policies for potentially dangerous behaviors, (3) whether schools have behavior codes with student rewards, (4) whether schools have behavior codes with student sanctions, and (5) whether the parent or student signs a responsibility agreement. There was substantial variation across CMOs in the extent to which CMO principals said their schools employed these policies.

CMOs with comprehensive behavior policies in their schools tended to have more positive impacts on math and reading achievement. This association does not appear to be driven by the results for a select group of CMOs. Figure V.1 plots CMOs' estimated student impact in math on the y-axis against the average comprehensiveness of behavior policies in that CMO's schools as compared to matched district comparison schools. Each point indicates a single CMO. The line of best fit is the linear regression line. There is a clear positive association between math impacts and comprehensive behavior policies across the full range of CMOs. The association between reading impacts and comprehensive behavior policies is not as strong, but again the results do not appear to be driven by outliers.

The association between the implementation of comprehensive school-wide behavior policies and student impacts is not driven by a single factor: each component of the composite is positively associated with impacts in both subjects (see Appendix Table K.7). Having school-wide zero-tolerance policies for potentially dangerous behaviors and having a behavior code with student sanctions are positively and at least marginally significantly associated with student impacts in both math and reading.

Although impacts are associated with the presence of school-wide behavior policies, they do not appear to be associated with whether the CMO centrally sets student disciplinary policies and provides a system of rewards and punishments for student behavior (see Appendix Table K.2). Thus, it is the comprehensiveness of behavior policies within schools, not the extent to which CMOs set these policies, that may be related to more favorable outcomes.

Figure V.1. Comprehensive Behavior Policy vs. Math Impacts



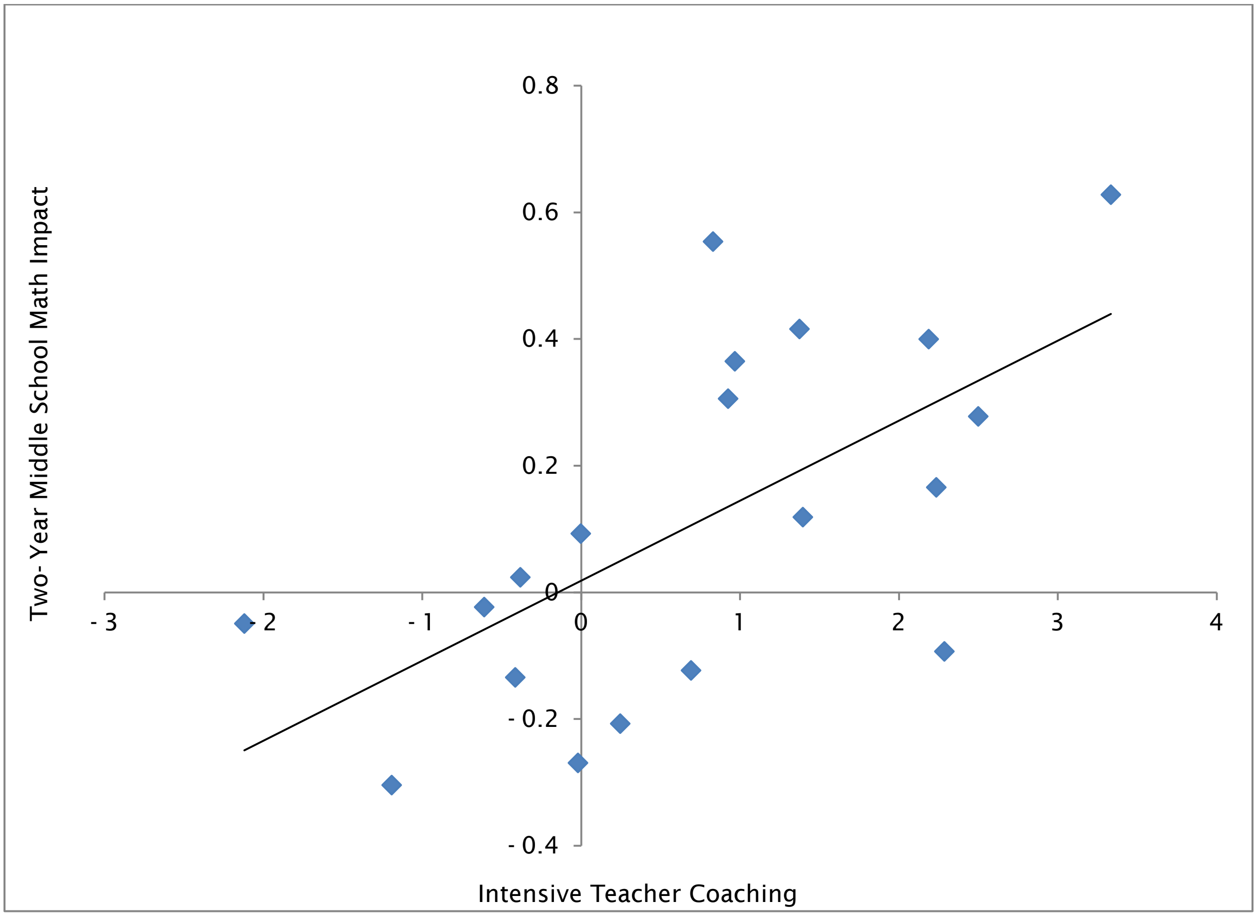
Source: State, district, and CMO school records and Principal Survey.

2. Intensive teacher coaching is positively associated with student impacts

Intensive coaching of teachers has the potential to increase student achievement by increasing the quality of instruction. Coaching may be particularly important for new teachers who are still developing their teaching skills and practices. The intensity of coaching for new teachers is captured with an index that measures the frequency with which new teachers (1) are observed by coaches, (2) are observed by principals or other administrators, (3) receive feedback from observers, and (4) must submit lesson plans for review.

More intensive teacher coaching is associated with more favorable student outcomes in both math and reading, although in reading the association is only marginally significant (p=0.07). Each component of the composite measure is positively associated with student impacts in both subjects (see Appendix Table K.8). As with behavior policies, the association is consistent across the range of math impacts and practices (see Figure V.2), and the results are similar in reading.

Figure V.2. Intensive Teacher Coaching vs. Math Impacts



Source: State, district, and CMO school records and Principal Survey.

Caution is merited in interpreting these results. When we include in a single model all three primary measures that are significantly associated with impacts—comprehensive school-wide behavior policies, intensive teacher coaching, and instructional hours—the effect of coaching is no longer statistically significant. Nonetheless, the magnitude of the association between teacher coaching and impacts still is non-trivial (the coefficient declines by only about a third) and our ability to isolate the independent contribution of coaching is limited by the small samples.

Several of the secondary hypotheses related to teacher coaching are also associated with impacts. The professional development and coaching resources provided by CMOs and the staff in their central office are consistently positively associated with impacts. Principals' reports that the CMO (or district) provides professional development support were positively and significantly associated with student impacts in both subjects. Additionally, the frequency with which CMO staff meet one-on-one with principals is positively and significantly associated with student outcomes in both math and reading.

If coaching of teachers improves student outcomes, we might expect it does so by improving instructional practice. Indeed, in chapter 3 we noted that the extent to which teachers are observed appears to be correlated with the frequency with which they modify their lesson plans using student assessments (which is one of the components of our instructional coherence measure). And the frequency with which teachers modify lesson plans is positively associated with impacts in both subjects and mediates the effect of coaching on impacts in math (see Appendix Tables K.9 and K.10). These results are consistent with the interpretation that teacher coaching improves student outcomes through altering teachers' instructional plans and encouraging them to make greater use of individual student assessments.

3. CMOs using TFA and teaching fellow teachers have higher impacts, but other staffing decisions are not associated with impacts

In addition to coaching teachers, CMOs must make many other decisions about how to staff their central office and affiliated schools. They must determine the number of office staff and teachers to hire as well as how to select and compensate these staff. These decisions may affect several important factors including the skills teachers initially bring with them to schools, class sizes, and other kinds of support CMOs need to provide schools. We examined whether various staffing decisions are associated with math and reading impacts.

Math impacts are higher among CMOs that rely more heavily on TFA and the Teaching Fellows programs as sources of new teachers. Specifically there is a statistically significant association between math impacts and the percentage of new teachers from these two sources, both of which tend to recruit and provide some training to recent graduates of highly selective colleges. One should be cautious about placing substantial weight on this finding because this is one of the many secondary hypotheses tested and the positive association could be due to random chance. However, it is possible that some promising CMO practices—such as adoption of school-wide behavior policies or longer instructional hours—are easier to implement in schools making use of TFA and Teaching Fellow teachers.

We find little evidence that other CMO hiring, staffing, or compensation practices are associated with student outcomes. CMOs' strategies of compensation for principals and teachers, including the use of performance-based compensation, are not associated with student impacts (see Appendix Table K.2). Impacts are not significantly associated with the weight given during hiring to the teacher's sample teaching performance or her commitment to the school's mission. There is also no evidence that student impacts are associated with whether teachers have opportunities for tenure. Similarly, the allocation of students to teachers is in general not associated with impacts, including instruction of students of similar ability together and teacher looping over grades, and the teacher-student ratio. Although we find a negative association between the ratio of CMO central office staff to teachers in both math and reading, these associations appear to be because of the results for two outlier CMOs and hence may be due to chance.

4. CMOs categorized as "data-driven" and "time on task" have larger impacts, on average, than two other categories of CMOs

The associations between CMO practices and student impacts may not be independent. CMOs may achieve more favorable student outcomes when they implement a package of complementary practices designed to improve student learning. In Chapter III, we identified four clusters of CMOs that implemented distinct groups of practices. CMOs classified as Time on Task emphasize school-wide behavior policies, lengthened instructional hours, and more intensive teacher coaching, whereas CMOs in the Data-driven group emphasize frequent formative assessments, performance-based compensation, and intensive teacher coaching. CMOs in the Incremental Innovation groups are similar to district schools in their policies, and Alternative Approach CMOs are least likely to implement each of the practices in our primary hypotheses.

Variation in impacts across these four clusters is significant in both subjects, based on results from an F-test for homogeneity of impacts across clusters. On average, CMOs in the Data-driven and Time on Task clusters achieve the most favorable outcomes for students in both subjects. In math, Time on Task CMOs have the highest average impacts, whereas in reading the largest impacts are from the Data-driven CMOs. Average impacts are positive and significant for the Data-driven cluster in both subjects, and positive but only significant in math for the Time on Task cluster. By contrast, the Incremental Innovation group has average impacts close to zero, which is perhaps unsurprising, given their similarity to district schools. The Alternative Approach group has significantly negative average impacts in both subjects.

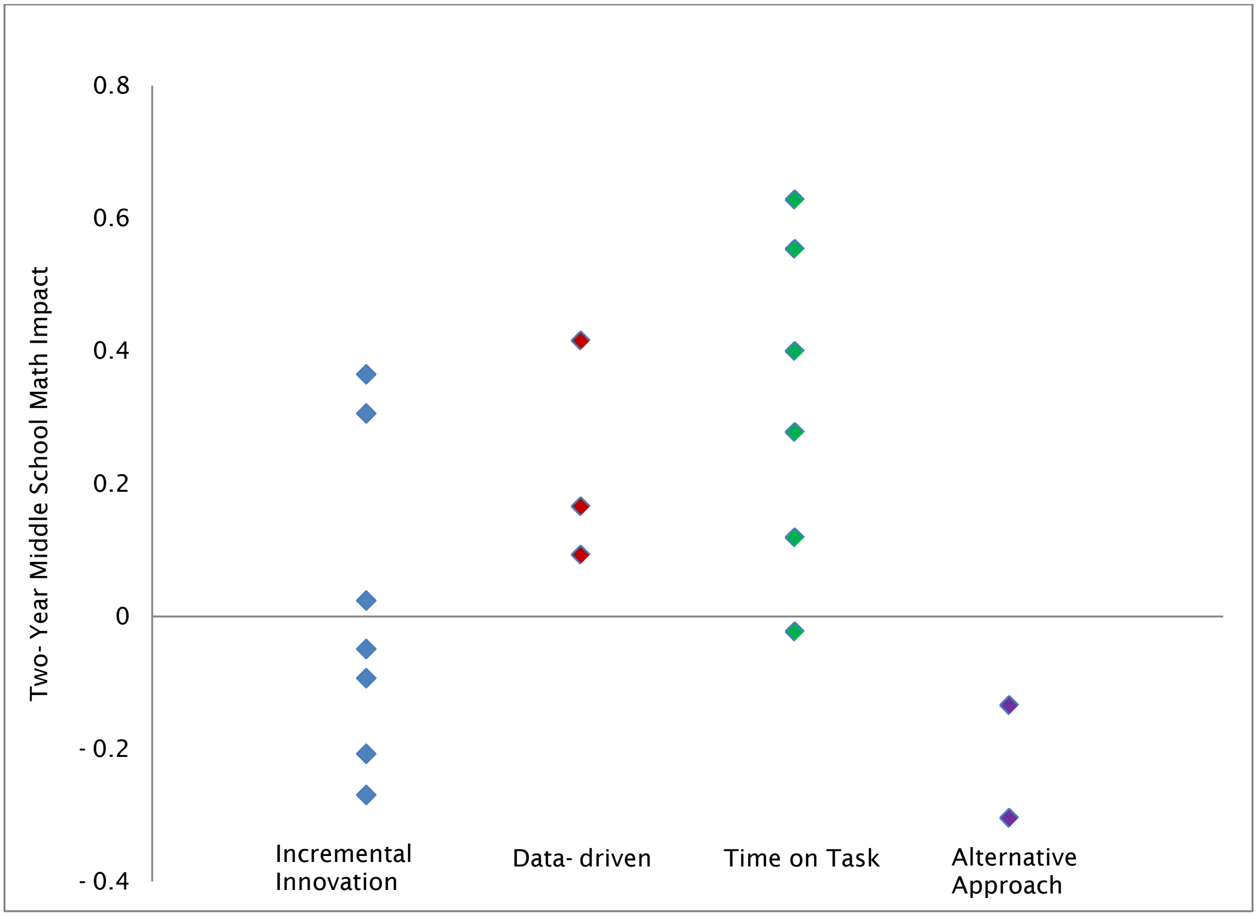
Although average impacts among the clusters can be distinguished, there is also considerable variation in impacts within some of the clusters. In Figure V.3, each dot represents the estimated impact for a single CMO, with CMOs grouped by cluster along the x-axis. The variation is particularly large in the Time on Task cluster, which includes CMOs with impacts ranging from less than 0 to more than 0.6 of a standard deviation.

The higher average impacts for the Time on Task and Data-Driven groups do not point towards any other promising practices aside from behavior policy and coaching. The relatively large impacts of the Time on Task cluster appear to be partly due to the emphasis these CMOs place on teacher coaching and behavior policy, but does not appear to be attributable to their longer instructional hours, a practice that is not associated with impacts in any of our multi-variate regressions including those that interact instructional time with other practices. The relatively large impacts of the Data-Driven cluster are also partially explained by the emphasis of these CMOs on teacher coaching. We find no evidence that the favorable impacts for this cluster are due to their emphasis on either formative assessments or performance compensation, since these practices do not appear to be associated with impacts (see Table V.1); moreover, we find no evidence of an interaction effect (or synergy) between coaching and either performance compensation or formative assessment (see Appendix Table K.11). However, even after taking into account their use of teacher coaching, the Data-Driven CMOs have somewhat higher impacts than expected. Therefore, some unmeasured attributes or practices of Data-Driven CMOs may be partly responsible for their large positive impacts.

5. Tightness of CMO management is weakly associated with impacts

In addition to variation in our core practice measures, CMOs exhibit considerable variation in the prescriptiveness of their management style. In Chapter III, three domains were identified in which CMOs might exhibit either a tight or loose management style: behavior policy, evaluation and compensation, and instructional approach. Based on their prescriptiveness in each of the three domains, CMOs were classified into four groups: Tight (all domains), Tight Evaluation and Compensation (loose instructional approach and behavior policy), Tight Instructional Approach (loose behavior policy and evaluation and compensation), and Loose (all areas).

Figure V.3. Math Impacts by Core Clusters

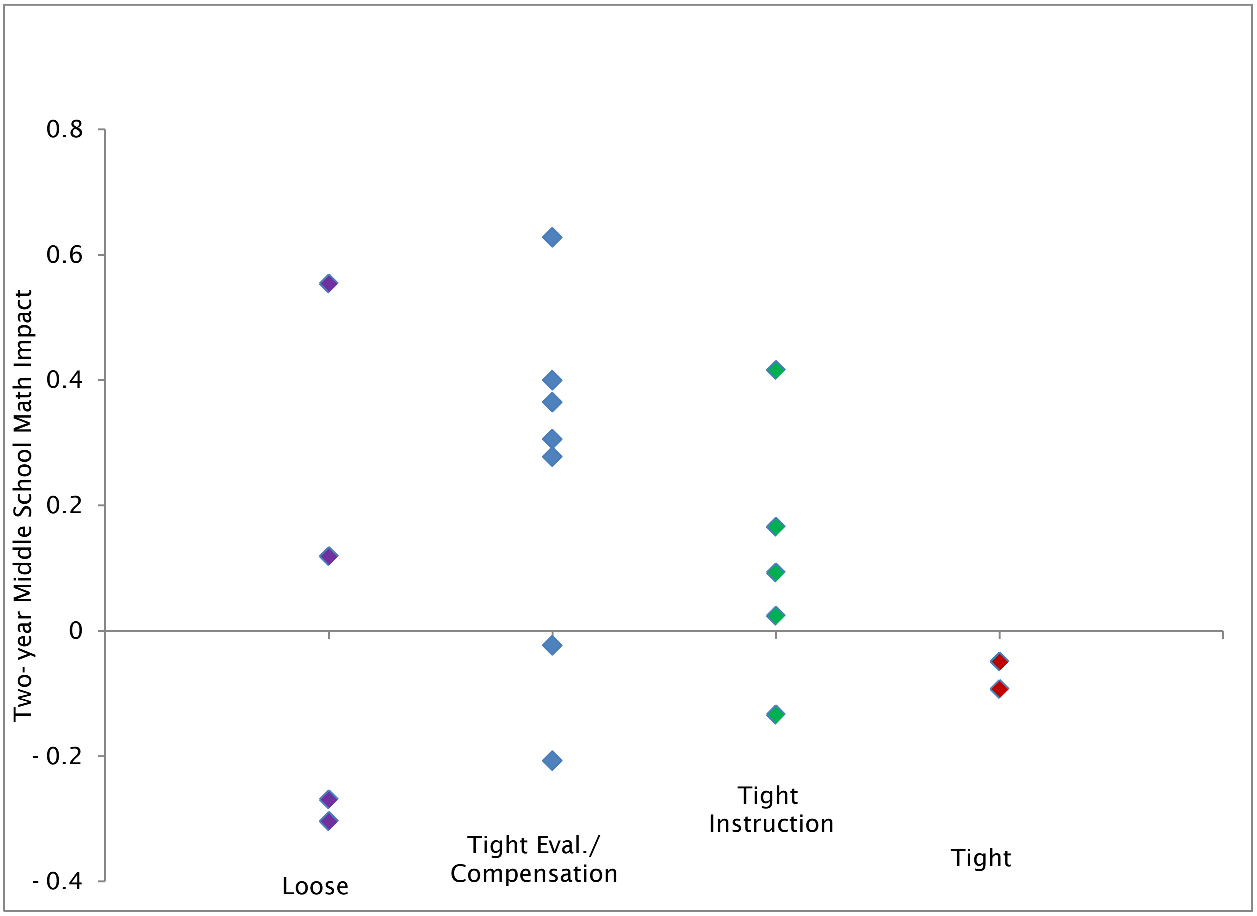


Source: State, district, and CMO school records and Principal Survey.

Compared to the core clusters discussed in the previous section, the tightness of CMO management is only weakly associated with impacts. In reading, the variation in impacts across all prescriptiveness groups is not significant at the five percent level. In math, the variation across groups is significant, but the variation in impacts within groups is large (see Figure V.4).

In both math and reading, average impacts are higher for the two groups that pursue a hybrid tight-loose approach than for the groups that are either tight in all domains or loose in all domains. However, few groups have average impacts that are significantly different from zero. Overall, CMO management style appears to be weakly associated with impacts.

Figure V.4. Math Impacts by Prescriptiveness Groups



Source: State, district, and CMO school records and Principal Survey.

VI. QUESTIONS FOR FUTURE RESEARCH

As is often the case in studies of this kind, some of the interesting findings raise other important questions. Here we discuss several questions that future studies might address.

What are the impacts of CMOs on students' long term outcomes? Most of the impact findings in this report focus on how CMOs affect academic achievement in middle school. Although many CMOs are focused on increasing academic achievement as measured by state student assessments, this is by no means their only objective. For example, nearly all CMOs have a long-term goal to prepare students for college. And many seek to cultivate students' broad intellectual, social, and emotional development and prepare them for successful careers. While chapter four reported CMO impacts on postsecondary enrollment, data on this outcome was available for only four CMOs. And a truly comprehensive evaluation of CMOs would entail an analysis of how they affect postsecondary degree completion, civic behavior, and earnings.

Are some CMOs selecting the wrong models to replicate or having difficulty replicating promising school models? The development of CMOs was intended to scale up the most promising charter schools. But over 40 percent of the CMOs covered by our analysis are falling short of the performance of nearby district schools in math or reading. This raises questions about the models CMOs choose to replicate and the ways in which they implement them. For example, some CMOs may be scaling up the wrong models. Or perhaps they originally identified a promising model but have had difficulty replicating it. (In addition, as noted above, some CMOs may also be focused on other outcomes aside from increasing academic achievement as measured by test scores).

The larger CMOs in our study have somewhat larger impacts on average than the smaller ones, which suggests that some of them have succeeded in replicating promising models. These CMOs have attracted sufficient amounts of funding and families to expand. Perhaps their funders and families have succeeded in confirming that the schools managed by these CMOs are effective.

But we also found evidence that several CMOs have become less effective as they have grown, at least in terms of their impacts on reading skills, which declined for eight CMOs as they grew and rose for only one CMO. (In math, by contrast, most CMOs' impacts did not change appreciably as they grew.) These findings suggest that CMO expansion can pose challenges. Our case studies suggest that CMOs encounter a number of issues as they grow, including difficulties finding teachers with the same skills as those in their first schools, providing principals with the right mix of direction and flexibility, and finding new facilities.

Questions remain about how CMOs should seek to overcome these hurdles. For example, how can CMOs recruit and train staff effectively and provide them with effective incentives, guidance, and other forms of support? There are also questions about whether and how conventional public schools and school districts can replicate promising CMO strategies. Districts may encounter challenges implementing these strategies that are quite different from those faced by CMOs. A first step in helping others decide whether and how to adopt any CMO strategies is to describe them in more detail, an issue we discuss below.

Which promising strategies should CMOs and school districts implement, how should they implement them, and to what extent must specific strategies be bundled together to be effective? Our findings on promising practices are all tentative because they are based on correlations of practices and impacts. Even correlations that are statistically significant could be spurious. Nonetheless these findings point to important issues that can be explored in the future.

Researchers could examine specific student behavior strategies to identify those that are effective in promoting better behavior and higher student achievement, and look for ways to implement them in more schools. A quarter-century ago, James Coleman and Thomas Hoffer (1987) argued that effective schools created communities in which behavioral expectations reinforced academic expectations. Our data indicate that the policies associated with CMO impacts include those that provide sanctions and rewards for specific student behaviors and, to a lesser extent, agreements with parents and students. Among the issues that should be explored are the best ways to specify and enforce sanctions and rewards and to train staff to implement them.

Frequent teacher coaching is also associated with positive impacts. Presumably the content and intensity of coaching depend in part on a school's educational and behavior policies. Previous experimental studies suggest that coaching programs do not necessarily pay off in the first year but sometimes do after two years of coaching individual teachers (Glazerman et al. 2010). Future research could explore what form of coaching is most effective.

To what extent do CMOs add value compared to independent charter schools? CMOs seek to take advantage of both the autonomy associated with charter status and the scale implicit in a larger organization, which could provide benefits in several areas, including curriculum development, teacher training, and various administrative tasks. Whether CMOs can take advantage of scale without losing the benefits of charter status (that is, becoming indistinguishable from school districts) is a key question. Our study was not designed to precisely measure the extent to which CMOs improve student outcomes relative to independent charter schools. Indeed, most of our data collection and analysis focus on how CMOs perform relative to the most common nearby alternative—regular district schools. However, in four large districts we also estimated how CMO schools performed relative to independent charter schools in terms of their contribution to students' academic achievement. Some CMOs did better than independent charters and others did worse. These analyses are insufficiently comprehensive to estimate overall how CMOs perform relative to independent charter schools. Moreover, we did not collect information on how the practices of CMOs compare with those of independent charter schools, so we could not ascertain how differences in practices might be related to their relative performance. These important issues should be explored in the future.

Are the newest generation of CMOs using the same strategies and producing the same impacts as the established CMOs in our study? Many more CMOs are operating today than could be included in our study. Since our study began four years ago, we focused on CMOs with four schools as of fall 2007. Hence many newer CMOs did not meet our criteria for inclusion. Some of these—for example, those in New Orleans—have arisen in response to specific needs and contexts that differ substantially from those encountered by the older CMOs. And some may have had opportunities to learn from the experiences of the older CMOs. It remains to be seen whether these new CMOs are more or less effective than the older ones.

What other factors might contribute to CMO impacts? A number of other differences between CMOs and districts, or among different CMOs, may contribute to CMO impacts. Some of these are difficult to observe or model and were not analyzed in this report. Other factors were measured but they did not show sufficient variation across CMOs to detect whether they contributed to impacts. Among the hypotheses we could not fully test in this study are ways in which impacts are related to high expectations in the classroom, funding, peer effects, grade configuration, and specific approaches to classroom instruction. We discuss these factors below.

First, positive impacts might be channeled through high expectations for student achievement—sometimes described as a "no excuses" approach—as manifested in the intensity of instruction in the classroom and in systems that hold teachers and principals accountable. These are factors that some field research suggests are central in high-performing charter schools (see, for example, Angrist et al. 2011). However, surveys are not ideal for measuring whether a "no excuses" approach is in fact in use. Intensive observations of classroom practices and central office interventions could produce more knowledge about the importance of expectations.

Second, differences in available funding—public and private, for operations and facilities—merit additional attention as a factor that might matter. Although we did not detect a relationship between funding and impacts, the tax forms of CMOs on which we had to rely are not likely to fully capture variation in resources available to different organizations. In addition, we could not develop comparable estimates of district and CMO per-pupil expenditures. An in-depth analysis of finances, ensuring comparability among CMOs and between CMOs and districts, could shed more light on the importance of resources.

Third, because CMOs operate schools of choice, the families they attract are different in both measurable and unmeasurable ways, which may give rise to peer effects. The selection process of students is driven in part by who learns about and chooses to apply to CMO schools. It is possible that the parents or students who end up enrolling in some CMO schools are more motivated or have other assets. In addition, CMOs can encourage certain families to apply or enroll in their school; even those with random lotteries can target their recruitment efforts and ask students to sign agreements to attend regularly and do their homework. An individual student may benefit from being in the same school and classroom with other students with higher levels of motivation or parental support. If peer effects are contributing to CMO impacts, this does not mean that our impacts are improperly measured. Indeed, our experimental results suggest the impacts are accurate. But it could affect our understanding of the mechanisms behind the impacts: Peer effects may explain why CMO students do better than they would have had they been placed in a school or classroom where there are fewer students like themselves. If that turns out to be true, it would also have important implications for policy: Similar effects might not be achieved, for example, if CMO practices were directly applied to conventional public schools that are not schools of choice. While peer effects can be challenging to estimate, future research should explore their importance.

Fourth, charter schools often employ grade configurations that are less common in conventional public schools, including K-8 and K-12 configurations. Some research suggests that schools using longer grade configurations that eliminate the elementary-to-middle school transition produce better outcomes for their students (Jacob and Rockoff, 2011). A study of charter schools in Chicago found substantial positive effects of charter schools on educational attainment, and noted that this might have been attributable to the fact that many of the schools had eliminated the transition from middle to high school (using grade configurations such as K-12 or 6-12) (Booker et al. 2011). Our study was not able to take a close look at the effects of varying grade configurations, because the absence of pre-entry baseline test scores precluded the examination of all schools beginning in kindergarten. In the future, historical data covering longer periods should at least permit the examination of CMO schools that eliminate the transition from middle school to high school.

Finally, our data provided only limited understanding of the classroom dynamics within high-performing CMOs. CMOs with large positive impacts must be doing something in the interaction between teacher and student that leads to such impacts. As previously mentioned, we do not know exactly how many teachers succeed in implementing a "no excuses" approach or exactly what this entails. We do not have much information on the types of assignments and homework assigned, or the specific curricular programs teachers are using in math, reading, and other subjects. We do not know the details of the strategies teachers employ to engage students and manage their classrooms. Follow-up research involving intensive data collection in classrooms could help answer these questions.

REFERENCES

Abdulkadiroglu, Atila, Josh Angrist, Sarah Cohodes, Susan Dynarski, Jon Fullerton, Thomas Kane, and Parag Pathak. "Informing the Debate: Comparing Boston's Charter, Pilot and Traditional Schools." Boston, MA: Boston Foundation, January 2009.

Angrist, Joshua D., Susan Dynarski, Thomas Kane, Parag A. Pathak, and Christopher R. Walters. "Inputs and Impacts in Charter Schools: KIPP Lynn." *American Economic Review: Papers and Proceedings*, vol. 100, no. 2, May 2010, pp. 1-5.

Angrist, Joshua, Susan Dynarski, Thomas Kane, Parag Pathak, Christopher Walters. (2010). Who Benefits from KIPP? NBER Working Paper No. 15740.

Angrist, Joshua D., S.R. Cohodes, S. Dynarski, J.B. Fullerton, T.J. Kane, P.A. Pathak, and C.R. Walters. "Student Achievement in Massachusetts' Charter Schools" Cambridge, MA: Center for Education Policy Research at Harvard University, January 2011

Angrist, Joshua D., Parag A. Pathak, and Christopher R. Walters. *Explaining Charter School Effectiveness.* National Bureau of Economic Research, Working Paper 17332. Cambridge, MA: NBER, 2011.

Angrist, Joshua, Imbens, Guido, and Rubin, Donald. (1996). Identification of causal effects using instrumental variables. Journal of the American Statistical Association, 91: 444-455.

Batdorff, Megan, Larry Maloney and Jay May, with Daniela Doyle and Bryan Hassel. May 2010. Charter School Funding: Inequity Persists. Available at:http://cms.bsu.edu/Academics/CollegesandDepartments/Teachers/Schools/Charter/CharterFunding.aspx.

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." New York: MDRC Working Paper, October 2008.

Booker, Kevin, Tim R. Sass, Brian Gill, and Ron Zimmer. "The Effects of Charter High Schools on Educational Attainment." *Journal of Labor Economics,* vol. 29, April 2011, pp. 377-415.

Coleman, James S. and Thomas Hoffer (1987). Public and Private High Schools: The Impact of Communities. New York: Basic Books.

Center for Research on Education Outcomes (CREDO). (June 2009). *Multiple choice: Charter school performance in 16 states.* Stanford, CA

Cook, T., W. Shadish, and V. Wong. "Three Conditions Under Which Experiments and Observational Studies Produce Comparable Causal Estimates: New Findings from Within-Study Comparisons." *Journal of Policy Analysis and Management*, vol. 27, no. 4, 2008, pp. 724–750.

Dobbie, Will, and Roland G. Fryer. "Are High-Quality Schools Enough to Increase Achievement Among the Poor? Evidence from the Harlem Children's Zone." *American Economic Journal: Applied Economics*, vol. 3, no. 3, 2011, pp. 158–187.

Fryer, R., [Use Lower case:] "Teacher Incentives and Student Achievement: Evidence from New York City Public Schools" Working Paper 16850, National Bureau of Economic Research, Cambridge, MA, 2011

Glazerman, S., A. Seifullah. "An Evaluation of the Teacher Advancement (TAP) in Chicago: Year Two Impact report", Mathematica, Washington D.C., 2010.

Glazerman, S., D. M. Levy, and D. Myers. "Nonexperimental versus Experimental Estimates of Earning Impacts." *The Annals of the American Academy of Political and Social Science*, vol. 589, 2003, pp. 63–93.

Gleason, P., Clark, M., Tuttle, C. C., and Dwoyer, E.. "The Evaluation of Charter School Impacts: Final Report (NCEE 2010-4029)." Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education, June 2010.

Hill, Paul T., and Lydia Rainey. "Charter School Maturation as a Factor in Performance Assessment and Accountability." In *Taking Measure of Charter Schools: Better Assessments, Better Policymaking, Better Schools,* edited by Julian C. Betts and Paul T. Hill. Los Angeles: Rowman and Littlefield, 2010.

Hoxby, Caroline M., Sonali Murarka, and Jenny Kang. "How New York City's Charter Schools Affect Student Achievement: August 2009 Report." Second report in series. Cambridge, MA: New York City Charter Schools Evaluation Project, September 2009.

Hoxby, Caroline and Sonali Muraka. (2006). Methods of assessing the achievement of students in charter schools. National Conference on Charter School Research.

Jacob, Brian A. and Jonah E. Rockoff (2011, September). "Organizing Schools to Improve Student Achievement: Start Times, Grade Configurations, and Teacher Assignments." Washington DC: The Hamilton Project at the Brookings Institution.

Lake, Robin, Brianna Dusseault, Melissa Bowen, Allison Demeritt, and Paul Hill. "The National Study of Charter Management Organization (CMO) Effectiveness: Report on Interim Findings." Center on Reinventing Public Education, June 2010.

Miron, Gary., & Urschel, J.L. (2010). Profiles of nonprofit education management organizations: 2009-2010. Boulder, CO: National Education Policy Center. Retrieved October 7, 2011 from http://nepc.colorado.edu/publication/EMO-NP-09-10

Murnane, Richard and John Willet. (2011). *Methods Matter: Improving Causal Inference in Educational and Social Science Research*. Oxford University Press: New York.

Newman, Fred, BetsAnn Smith, Elaine Allensworth, and Anthony C. Bryk. *Instructional Program Coherence: Benefits and Challenges*. Chicago: Consortium on Chicago Schools Research, 2001.

Rosenbaum, P.R., and D. B. Rubin. "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, vol. 70, no.1, 1983, pp. 41–55.

Shadish, William, Thomas Cook, and Donald Campbell. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Houghton Mifflin Co: Boston.

Tuttle, C.C., P. Gleason, and M. Clark. "Using Lotteries to Evaluate Schools of Choice: Evidence from a National Study of Charter Schools." *Economics of Education Review*, forthcoming, 2011. doi: 10.1016/j.econedurev.2011.07.002

Tuttle, Christina Clark, Philip Gleason, and Melissa Clark (2011). "Using Lotteries to Evaluate Schools of Choice: Evidence from a National Study of Charter Schools." Economics of Education Review, July online.

Tuttle, C.C., B. Teh, I. Nichols-Barrer, B.P. Gill, and P. Gleason, "Student Characteristics and Achievement in 22 KIPP Middle Schools." Washington, DC: Mathematica Policy Research, June 2010.

U.S. Department of Education. *What Works Clearinghouse Procedures and Standards Handbook Version 2.1*. Washington, DC: U.S Department of Education, December 2008.

Wolf, Patrick et al. (2010), "Evaluation of the DC Opportunity Scholarship Program." U.S. Department of Education, Institute of Education Sciences

Zeibarth, Todd. "2007 Charter Schools Dashboard." National Alliance for Public Charter Schools, 2007

APPENDICES

APPENDIX A

CONSTRUCTION AND ANALYSIS OF MEASURES USED IN CHAPTER III

Several of our measures of CMO practices are composite variables. These composites were created by combining closely related survey items into a single measure, reducing measurement error and capturing the breadth of a construct. The composite variables and the items they each combine are listed in Table A.1.

The process for creating these composite measures included a number of steps to maximize reliability and reduce dimensionality. We first identified all the items from the surveys that were conceptually related to a specific construct. Next we standardized the values for each of the items such that the overall mean for both CMO and comparison schools had a value of 0 and a standard deviation of 1.[1] We used principal components analysis to confirm that the composite was unidimensional, excluding items that were not related to the underlying construct. We then computed the standardized Cronbach's alpha, an estimate of the internal consistency or reliability of a composite measure, and rejected composites with alphas smaller than 0.6.[2] All of the composites passed these tests.

To measure the CMO characteristics described in Chapter III, we chose to rely primarily on the principal survey, rather than on our central office staff survey or teacher survey, for several reasons. First, relative to the central office survey, the principal survey reflects the perceptions of respondents who are closer to the implementation of CMO practices in schools. Second, because we surveyed the principals of both CMO and nearby district schools, we could compare the responses of these two groups to gauge how CMO schools are distinctive. Moreover, the analysis of how impacts are related to practices described in chapter V, makes use of our measures of the differences in practices between each CMO school and its matched district school. Finally, the sample of 36 CMOs (including 221 responding CMO principals and 171 responding district principals of 292 matched pairs) covered by the principal survey is larger than the sample of 23 CMOs covered by the teacher survey, which also included only CMO teachers. Thus, while teacher respondents may be more attuned to the implementation of CMO practices in classrooms, relying on the principal survey enabled us to measure practices in a substantially larger sample of CMOs, as well as to draw contrasts with district schools.

Our unit of analysis is the CMO.[3] Some of our measures incorporate survey questions that ask explicitly about policies and activities of the CMO central office. However, we also rely on principal survey responses about school activities to infer CMO-level characteristics, averaging principal

---

[1] We summed the items in each composite at the school level and restandardized the composite values after aggregation to the CMO level, such that each overall mean across CMOs and district comparison school groups had a value of 0 and a standard deviation of 1.

[2] Conventionally, alphas greater than 0.7 are considered reliable, but because some of our composites had a small number of items, we accepted composites with alphas slightly smaller than this threshold.

[3] The unit of analysis for the comparison sample corresponds to the group of district schools matched with the schools within a CMO. Therefore, for CMOs with schools located in more than one district, this comparison unit aggregates responses from multiple districts, weighted by the number of CMO schools located in each district.

responses within each CMO and within the group of comparison schools associated with that CMO using weights to adjust for nonresponse.

Of 292 CMO principals eligible for the study, 76 percent responded to the survey. Among the 292 matched comparison principals, the response rate was 59 percent. Five CMOs eligible for the study declined to participate in the survey.

Among participating CMOs, weights to adjust for principal nonresponse were created within each CMO and within each corresponding group of matched schools. Specifically, we created weighting cells within the CMO and district groups based upon CMO, school district, and a three-level indicator of whether the school was (1) a middle or high school, (2) an elementary school, or (3) contained elementary, middle, and high school grades (such as K-12). Schools that included elementary and middle grades were considered "elementary" for this indicator. Adjustments were calculated by summing the total sample of eligible schools within the weighting cell and dividing by the total of these weights for respondents only.

The intraclass correlation coefficients for our primary measures of CMO practices indicate that a substantial portion of the total variation we observe in CMO school-level practices across our primary measures is due to variation across CMOs (Table A.2). As such, we assume that these school-based measures capture the emphasis the CMO central office places on a specific practice or some characteristic of the CMO, even if the CMO is not explicitly promoting the strategy in its schools. To estimate the intraclass correlation coefficients we began with the following random effects model:

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij},$$

where i indexes group (CMOs); j indexes principals within each CMO; μ is an unobserved overall mean; α_i is an unobserved random effect shared by all values in group I; and ε_{ij} is an unobserved error term.

Next the intraclass correlation coefficients were estimated as:

$$\frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\epsilon^2}.,$$

where σ_α^2 corresponds to the variance of αi and σ_ε^2 corresponds to the variance of ε_{ij}.

Table A.1. Survey Items Used in Composite Measures of Primary Practices

Composite Measure	Principal Survey Items Included
Consistent educational approach, including curriculum and instructional materials	1. All schools in CMO (district) must use singular instructional model 2. CMO (district) provides support selecting some books/instructional materials 3. CMO (district) responsible for selecting curricula/instructional materials
Student behavior policies that include specific rewards, sanctions, and commitments	4. School enforces consistent behavioral standards and disciplinary policy 5. School has zero tolerance policy for potentially dangerous behaviors 6. School has behavior code with student rewards 7. School has behavior code with student sanctions 8. Parent or student required to sign responsibility agreement
Frequent review and analysis of student formative assessment data	1. How often diagnostic test results are given to new teachers 2. Frequency with which teachers meet to analyze student test score data 3. Frequency with which principal meets with CMO (district) central office to review test scores
Intensive teacher coaching and monitoring	1. Frequency with which new teachers are observed by coaches 2. Frequency with which new teachers are observed by principals/administrators 3. Frequency with which new teachers receive feedback from observers 4. Frequency with which new teachers must submit lesson plans for review
Performance-based teacher evaluation and compensation	1. Relative importance of student test scores to teacher education and seniority/tenure in determining teacher pay 2. Relative importance of administrator assessments to teacher education and seniority/tenure in determining teacher pay 3. Use of student test scores to evaluate teachers

Table A.2. Intraclass Correlation Coefficients for Primary Measures

Primary Practice	ICC
Consistent educational approach, including curriculum and instructional materials	0.51
Student behavior policies that include specific rewards, sanctions, and commitments	0.30
Frequent review and analysis of student formative assessment data	0.28
Intensive teacher coaching and monitoring	0.30
Performance-based teacher evaluation and compensation	0.40
Amount of instructional time	0.64

APPENDIX B

EXPERIMENTAL IMPACTS ON STUDENT ACHIEVEMENT

I. Introduction

CMO schools with more applicants than available seats often admit students using a random lottery. Records from these lotteries can be used to conduct a randomized experiment as students who participated are offered CMO admission completely by chance. Randomized experiments provide the most internally valid impact estimates for the small group of CMO schools with oversubscribed admission lotteries (Shadish et al., 2002; Murnane and Willet, 2011). Unfortunately, only a small fraction of CMO schools had oversubscribed lotteries, only a few cohorts at those schools were admitted by randomized lottery, and sufficient lottery records could only be obtained for a fraction of those cohorts. To illustrate, the middle school experimental analysis in this study includes seven CMO schools and five CMOs, while the primary middle school quasi-experimental analysis includes 68 CMO schools, and 22 CMOs. Clearly, the experimental impact estimates are inferior to the rigorous quasi-experimental estimates for estimating *overall* CMO impacts. However, the experimental estimates, the most causally valid estimates, may still interest researchers (Hoxby, 2006).

In this study, experimental treatment and control status was determined by whether the student received an admission offer *at the time of the lottery*. Some of the students offered admission ultimately decline the offers and the school then offers admission to lottery "losers." These subsequent offers are designed to be made in the random order generated at the time of the lottery. The CMO enrollment rates of students admitted at the time of the lottery are substantially higher than students rejected at the time of the lottery, meaning that this measure provides enough random enrollment differences to fairly test CMO impacts. Angrist et al., (2010) used a similar approach. This approach does mean that some control students attend a CMO school (that is, high crossover or contamination) obscuring intent-to-treat estimates. However, a traditional experimental approach where assignment is based on whether students were ever offered admission to a CMO school was only possible for two lotteries; there were two reasons this approach was not possible at other schools. First, many schools with oversubscribed lotteries ultimately offered admission to all students not admitted at the time of the lottery, meaning there were no randomized students who never received an offer (that is, no control students). Second, a few schools did not have good records on whether the randomization order was followed when admitting students after the lottery.

The experimental results presented in this appendix use a slightly larger sample of CMO schools than the experimental analysis presented in Appendix C. This sample includes an extra middle school lottery where data was obtained too late to be included in the validation sample plus three elementary schools not part of the validation (students applying in kindergarten do not have a baseline test score necessary for a rigorous non-experimental analysis). This appendix primarily reports impacts by school district and school-level (elementary/middle/high), but also reports an overall experimental impact estimate. Note that this overall experimental estimate includes many fewer schools than the study's primary propensity-score estimates and that the overall estimates are calculated slightly differently.

II. Recruiting Schools, Lottery Records, and Student Data

a. Eligible Schools

Many CMO schools with lotteries were not sufficiently oversubscribed and documented to support a rigorous experimental impact analysis. In the summer of 2008, we began identifying schools with two types of lotteries: a *prospective sample* that involved the study team to observing and independently documenting lotteries occurring the following spring, and a *retrospective sample* that utilized schools' earlier lotteries documentation with no independent verification by the study team. CMOs and schools were identified by reviewing data provided by project funders and other publicly available information. As with the non-experimental impact analysis, CMO schools must have opened by fall 2007 to be eligible. This process identified 207 CMO schools.[4] Prospective eligibility was often assessed before retrospective eligibility because prospective lotteries required more preparation (such as obtaining parental consent).

Table B.1 displays school counts at each stage of the process. While 207 CMO schools were initially identified, only 109 had students reaching a tested grade in prospective sample and we were only able to obtain lottery and student data for 26 schools. In the retrospective sample, we obtained lottery and student data for 14 schools. Other studies have also found low rates of usable charter school admission lotteries (Tuttle, Gleason, and Clark, 2011).

Table B.1. Number of CMO Schools in Each Stage of Sample Process

	Prospective	Retrospective
Identified CMO Schools	207	207
Cohort Reaching Tested Grade	109	187
Inadequate Lottery Records (retrospective only)	NA	123
Adequate Records and Lottery Participants	33	15
Able to Obtain Student Data	26	14

Ninety-eight schools were excluded from the prospective sample because students would not have outcome data. Specifically, elementary schools were not eligible for the prospective lottery sample because students would be completing kindergarten or first grade, and standardized tests, the outcome measures, are not administered in those grades. Similarly, since many states do not administer tests to ninth graders, high schools in some states were ineligible for the prospective lottery sample.

Eligible CMOs and schools were then contacted by phone and email to obtain more information on lottery oversubscription. To screen schools for the prospective lottery sample, we gauged whether spring 2009 lotteries were likely to be eligible by asking CMOs about their schools' spring 2008 lotteries. The following information was requested for each school:

[4] Eight-five additional CMO schools were identified later (the final frame was 292 CMO schools).

- Whether they had at least 20 lottery winners and losers at the time of the lottery; the losers were generally placed on a waiting list and could receive an offer in the future;
- Details on any lottery stratification, such as preference given to particular groups of students (for example, living in-district or qualifying for free/reduced price lunch);
- A description of the waitlist procedure to examine whether it was random

We also asked CMOs about potential changes to the admissions or lottery processes between 2008 and 2009 that might influence eligibility for inclusion in an experimental analysis.

Seventy-six schools were excluded based on responses to these questions. Most schools were excluded because there were too few lottery winners or losers. Some schools were excluded if they did not hold lotteries in 2008 or were unresponsive to our contacts. We also excluded some schools that held lotteries but did not appear promising for other reasons, such as excessive lottery stratification.

For the retrospective sample, only 20 schools were excluded for lacking outcome data. Elementary schools could be included if they were open by fall 2006 and held a kindergarten lottery that year such that these students would reach 3rd grade, the first tested grade, by fall 2009. High schools could be included if a cohort had reached a tested graded by the 09-10 school year.

Of the remaining 187 potential retrospective schools, 58 were excluded based on prospective lottery responses that indicated a lack of oversubscription, exhausting waitlists, or making offers first-come/first-admitted. For the 129 schools that were left, the retrospective screening process began by asking about the records schools retained from their lotteries prior to 2009, and about the quality of any retained records. For example:

- Do you have accurate documentation on the randomized lottery results from this year—which students were offered admission based on a favorable lottery draw and which students were not offered admission?
- Do your records distinguish between students who applied at the time of the lottery and those who applied after the lottery?
- Do your records distinguish between students who were accepted at the time of the lottery and those who were accepted after the lottery?
- Were siblings or any other types of students exempt from the lottery? If so, do your records identify those students?
- Was preference given to particular groups of students in the lottery? If so, do your records identify those students?

Ninety-two schools were deemed ineligible for the retrospective sample based on answers to our questions and further review of available records. The majority of these, 64 schools, were dropped because they were unable to provide sufficient documentation of their lotteries. The remaining schools maintained records but could not answer key questions such as distinguishing between students who won a seat at the time of the lottery and those made admission offers later. Finally, we excluded 31 schools that were unresponsive to our attempts to collect lottery information for the study. As we did not observe or document the lottery process for the retrospective lotteries, it is possible that the retrospective information reported by schools is not fully accurate.

b. Observing and Obtaining Consent for Prospective Lotteries

Working with CMOs and schools, we secured parental consent for participation and student record collection. Consent was obtained prior to the lottery meaning that treatment status could not affect or be related to the likelihood of consenting. We also attempted to observe many prospective lotteries to verify randomization and obtain a list of students admitted and waitlisted during the lottery process. Obtaining a list at the time of the lottery provides confidence that students classified as treatment were not admitted after the lottery in a non-random way. To maximize our chances of obtaining a valid list, the lottery observer procured a copy of the lottery list generated by the school while in attendance, and wherever possible, the observer recorded his or her own version of the lottery drawing. Of the 33 prospective lotteries, we observed and documented 14. For the remaining lotteries, we asked several questions to verify the validity of the process, and tried to obtain lottery records as soon as possible after the lottery.

c. Student Records Data

The experimental analysis uses student-level administrative data provided by state departments of education, school districts, and CMOs. Unlike the non-experimental analysis, the experimental student data needed to be identified (that is, include student IDs or names) to match with the lottery records. For prospective lotteries, we obtained student records only for students whose parents had consented. For the retrospective lotteries, we signed agreements with the district to provide experimental impact estimates in return for identified data. In addition to the identifiers, the following student characteristics were requested: school, grade level, baseline reading and math test scores (including missing test score indicators), sex, race/ethnicity, baseline free- or reduced-price lunch (FRPL) eligibility, baseline English language learner (ELL) indicator, and baseline special education status indicator. Three districts or states were not responsive to data requests and schools in those districts or states were dropped from the analysis.

III. Sample

a. Sample

Once student records were obtained, we excluded certain student types from our analysis sample for ethical, practical, or validity reasons. Each of these exclusions used data that was determined in advance of the lottery and is thus uncorrelated with assignment (that is, these exclusion do not bias the internal validity of our estimates). First, the baseline sample also excluded students if we were unable to obtain parental consent to participate in the study before the lottery. Second, we excluded the few students who had more than three entries in a single lottery because the sample weighting to account for different admission probabilities was extremely complicated for such students. Third, we excluded students whose administrative records indicated they were in the wrong grade at time of application; for example, students in the third grade applying to a fifth grade admission lottery. Often these students applied to the lottery by mistake.[5] Finally, we excluded students for whom we could not obtain achievement data in any subject in the baseline year, the school year before lottery winners enrolled (Gleason et al., 2010 also excluded these students). This exclusion eliminated all applicants from private schools, since private school students are not in district or state administrative records. The motivation for this last exclusion was to reduce differential attrition that could bias impact estimates, as private school applicants rejected at the admission lottery were more likely to return to private schools than admitted private school applicants, meaning that observable outcomes were less likely for control students than treatment students. While this exclusion also eliminated other kinds of applicants without baseline data, students from out of district for example, the goal of the experimental analysis is not representative results, but the most valid causal inference.

We also dropped entire sites[6] to ensure the validity and power of the impact estimates. Experimental sites had to meet each of the following criteria to be included in the analysis:

1. *Low student attrition*. Specifically, the overall and differential attrition rates must be lower than the What Works Clearinghouse maximum thresholds (liberal attrition standard);
2. *Valid randomization*. If we did not observe the lottery and consequently were unsure of the randomization validity, any difference between treatment and control average baseline test scores must be less than 0.25 standard deviations and demographic differences must be less than 25 percentage points;[7]
3. *Higher treatment group CMO enrollment*. The difference in CMO school enrollment between treatment and control groups must be at least 20 percent.

The first restriction aimed to reduce bias associated with attrition. Five sites had high attrition.[8] The second restriction was implemented to reduce the likelihood of including sites with faulty randomization or recordkeeping. Ten sites where we did not observe lotteries had one baseline test score difference of at least 0.25 standard deviations. The final restriction served to make sure there were enough differences in treatment and control CMO enrollment to provide a reasonable test of CMO impacts given the substantial control student crossover; if both treatment and control groups had similar rates of enrollment, the observed differences between the groups would be small regardless of whether CMOs had impacts. Four sites did not have large enough enrollment rate differentials.

---

[5] Students in the "wrong" grade during the outcome year would have a different outcome measure—state achievement tests vary by grade—than other students in the analysis. We could not include the subset of students in the correct grade during the outcome year, because being admitted to a CMO school could have affected whether students were in the correct grade.

[6] Experimental sites are groups of schools that shared applicants. Grouping schools into sites was necessary as treatment was defined as receiving an offer to any CMO school.

[7] The effect size measure for test was Hedge's g. Relatively large baseline differences were allowed because some of the sites were small and could have moderate baseline differences even if the randomization was sound.

[8] Some sites were eliminated for multiple reasons.

b. Sample Sizes and Attrition

After these exclusions, there were 687 treatment and 943 control students with baseline data (16 schools in 10 sites) who were eligible for the reading analysis, and 425 treatment and 667 control students with baseline data (10 schools in 8 sites) who were eligible for the math analysis.[9] Some applicants did not have student records data in the outcome year—attrition—leaving a final reading analysis sample size of 609 treatment and 819 control students and a final math analysis sample size of 395 treatment and 605 control students. Baseline and analytic sample sizes by school district and school-level are reported in Table B.2. For each district/level combination, overall and differential attrition are low by What Works Clearinghouse attrition standards, indicating that attrition is unlikely to substantially bias impact estimates.

Table B.2. Baseline and Analytic Condition Sample Sizes by District/Level

	A				B				C		D	
District Level	Middle		High		Elementary		Middle		High		Middle	
Condition	T	C	T	C	T	C	T	C	T	C	T	C
Baseline	23	87	140	170	82	94	305	423	122	106	15	69
Analytic	23	79	111	124	64	54	295	407	103	90	13	65

Notes: T indicates the treatment group sample size and C indicates the control group sample size. The sample sizes for the high schools in Districts A and C are for the reading analysis only.

c. Baseline Equivalence

Consistent with minimal bias, baseline statistics of observable characteristics indicate that the final treatment and control groups for each district/level combination were very similar for both reading and math impact analyses (see Table B.3). Only one of the differences was statistically significant, the percentage of applicants with an IEP in District D.

[9] The math analysis also excludes 396 students applying to a high school in one state because the math test taken in the ninth grade in that state could be affected by treatment.

Table B.3. Baseline Treatment And Control Means by District/Level

	A				B				C		D	
District Level	Middle		High		Elementary		Middle		High		Middle	
Condition	T	C	T	C	T	C	T	C	T	C	T	C
Read Score	.32	.08	-.02	-.15	-	-	.09	.08	.24	.30	.15	.17
Math Score	.32	.18	-	-	-	-	.05	-.02	-	-	.19	.22
% Male	52	39	40	46	44	53	53	47	57	67	46	52
% Black	0	0	1	0	85	72	71	73	12	10	38	29
% Hispanic	100	100	99	99	12	22	26	24	41	55	62	29
% FRPL	65	63	54	54	-	-	-	-	50	38	85	77
% ELL	17	14	21	31	0	3	4	3	10	5	-	-
% IEP	4	4	5	9	-	-	8	8	6	6	0	8
% CS enroll	-	-	-	-	-	-	5	2	54	54	0	2
Sample Size	**23**	**79**	**111**	**124**	**64**	**54**	**295**	**407**	**103**	**90**	**13**	**65**

Notes: CS stands for charter school. For some districts the sample sizes are slightly smaller for baseline reading test score (2 students had missing data), baseline math test score (2 student had missing data), and percent LEP (21 students had missing data). Hyphens indicate reliable data was not available from that district. Means are weighted to account for admission probabilities.

IV. Methods

a. Impact Analysis

Two common analyses in the charter school literature are estimating the effect of receiving an offer of admissions to a charter schools (an intent-to-treat or ITT estimate) or the effect of attending a charter school (an effect of treatment-on-the-treated or TOT estimate). These estimates are not identical, because some lottery winners decline their offers, and some losers at the time of the lottery end up enrolling in the schools as a result of wait-list admissions or other post-lottery changes. Both estimates are of interest as a policy matter: the ITT impact is relevant in assessing the likely effects of making more CMO enrollment spaces available; and the TOT impact is relevant in assessing the effect that can be expected for a student who enrolls. These different analyses estimate impacts for different groups of students and make different assumptions.

We chose to use ITT estimates, because experimental ITT estimates make fewer assumptions than experimental TOT estimates and therefore have stronger causal validity. Specifically the validity of the experimental TOT impact estimates depends on the assumption that crossover students—lottery losers who enroll in the CMO schools—experience the same impacts as lottery winners who would have enrolled in CMO schools regardless of whether they won the lottery.[10] Since the assignment variable for our lotteries is winning a lottery at the time of the lottery, the ITT estimates include substantial crossover of control students into treatment, as seen in Table B.4. As crossover students receive offers later, it is possible that their enrollment in the CMO begins after the start of the school year, meaning crossover students experience a different treatment and have different impacts from the lottery winners. (Unfortunately, the data do not indicate how often this occurs.) Smaller impacts for crossover students would produce TOT estimates that are biased upward. As the goal of the experimental analysis is to provide the most causally valid estimates, we report results of ITT analyses.

Table B.4. Treatment and Control CMO Enrollment Rates by District/Level

	A				B				C		D	
District Level	Middle		High		Elementary		Middle		High		Middle	
Condition	T	C	T	C	T	C	T	C	T	C	T	C
Rate (%)	69	4	68	24	67	34	60	30	71	10	79	3

b. Impact Model

To estimate an intent-to-treat impact by school district and school level, we compare outcomes of applicants offered admission at the time of the lottery to those of applicants rejected at the time of the lottery, controlling for students' previous test scores and demographic characteristics.[11] The impact estimation model for each district/level is:

$$(1)\quad y_i = \alpha + X_i\beta + \delta T_i + S_i\theta + \epsilon_i,$$

where y_i is the reading or math test score outcome after one year[12] of potential exposure to a CMO school for student i; α is the intercept; X_i is a vector of baseline achievement and demographic characteristics (see Table F.2);[13] T_i is a binary variable for treatment status, indicating whether student i was admitted at the admission lottery; S_i is a vector of indicators identifying which site the student applied to; ε is a random error term that reflects the influence of unobserved factors on the outcome; and β, δ, and θ are vectors of parameters or parameters to be estimated. The estimated coefficient on treatment status, δ, represents the impact of admission to a CMO school at the time of the lottery. The model assumes that the treatment indicator and covariates influence the

[10] In other words, among the students who would attend the CMO schools regardless of whether they won or lost the lottery (that is "always-takers"), impacts experienced by the lottery losers must be equivalent to impacts experienced by the lottery winners.

[11] As student admission to CMO schools was randomly determined, we could simply compare the mean outcomes of the treatment and control groups. However, to obtain more precise impact estimates, we adjust for baseline student characteristics in a regression model.

[12] For elementary students, it is four years, the difference between kindergarten and third grade.

[13] In some jurisdictions/levels there was no variation in a characteristic (for example, all students were Hispanic). In these cases, those characteristics were dropped from the analysis.

outcomes in the same way across all sites within a district/level. We assume fixed site effects when the jurisdiction/level includes multiple sites.[14]

To estimate an average experimental impact, we estimate site-level impacts (using Equation 1 without $S_i\theta$) and then regress middle and high school site-level impacts on an intercept assuming fixed site effects. The intercept coefficient is the average impact estimate, with each site weighted equally. Elementary school impacts are excluded from this average because elementary outcomes occur four years after the lottery, rather than one year for middle and high schools.

Student observations are weighted to account for admission probabilities. Weighting is needed because different types of applicants have different probabilities of admission to a CMO school or site (based on sibling preferences, multiple entries, and so on); thus applicant types with different admission probabilities will be over- or underrepresented in each condition. Since admission probabilities could be correlated with outcomes, this is problematic. Weighting rectifies this problem by giving applicant types with low admission probability proportionately more representation in the treatment condition and proportionately less representation in control condition (and vice versa for high admission probability types). Additionally, some students applied to more than one experimental CMO school, and in these cases, we treat all schools sharing applicants as a single site with the weights adjusted accordingly. For more information on the weighting approach, see the supplement to this appendix.

c. Missing Data

As student observations were required to have at least one baseline-year achievement score to be included in the analysis sample, there was minimal missing data for any of the covariates. Only two students were missing a baseline math score and two students were missing baseline reading scores; for these students we set each missing test score to the state or district-level mean, which is zero since the scores are standardized. Rates of missing pre-baseline test scores—two school years prior to outcome year—were higher (151 students were missing pre-baseline math scores and 111 were missing pre-baseline reading scores); for these students we set each missing test score to the state or district-level mean and included a missing data indicator. No observations were missing race, sex, free/reduced price lunch status, IEP indicator, or enrolled in a charter school at baseline indicator. For the 21 students with missing ELL indicators, we recoded the missing values to the mode across all students in the sample, not an English language learner. We did not impute outcome test scores; these students are considered attrition.

V. Results

For middle and high schools, the average one-year reading impact estimate is -.02 and the average one-year math estimate is .05; neither impact was statistically significant (Table B.5).[15] As noted, these are ITT estimates, meaning they represent the impact of being admitted to a CMO

---

[14] Assuming fixed site and school effects means these results cannot be generalized beyond the current sample of schools and we cannot cluster standard errors at the site level.

[15] Assuming random site effects or random school effects would increase the standard errors and accordingly the p-values; hence the impacts would remain statistically insignificant.

school at the time of the lottery, not the impact of actually attending a CMO school. Additionally, these estimates are not representative of general CMO impacts, as they are based on a small subsample of CMO schools and students—the propensity score estimates reported in this study are considerably more representative.

As with the non-experimental impact estimates, the experimental impact estimates are diverse (Table B.5). The middle and high school reading impact estimates range from -.20 to .11 and the math estimates range from -.15 to .23. Given the small to moderate impact estimates and small sample sizes, unsurprisingly none of the estimates is statistically significant at the 5 percent level.

The elementary four-year impact estimates, -.33 in reading and .10 in math, are of particular interest because we cannot estimate rigorous non-experimental impacts for elementary schools since there are no baseline test scores. Unlike the one-year middle and high school impact estimates, the elementary estimates are four-year ITT impact estimates (kindergarten to third grade). The elementary reading estimate is very sensitive to the inclusion of six students with large, negative standardized outcome scores of 3.67 (no elementary student had a math outcome larger than 2.38). All of these students were not admitted at a CMO lottery, although two of them actually attended CMO schools. Excluding these six outliers, the reading impact estimate is .04 with a standard error of .12.

Table B.5. ITT Achievement Impacts by District/Level

	Overall (middle and high school)		District A			District B				Dist. C	District D	
Level			Middle		High	Elementary		Middle		High	Middle	
Subject	Read	Math	Read	Math	Read	Read	Math	Read	Math	Read	Read	Math
Coeff.	-.02	.05	.01	.23	.11	-.33	.10	-.02	.07	-.09	-.20	-.15
SE	.04	.07	.11	.13	.07	.19	.16	.05	.04	.09	.12	.13
n	1310	882	102	102	235	118	118	702	702	193	78	78

Notes: This table reports the coefficients, standard errors (SE) and sample sizes (n) from a regression of standardized reading/math test scores on whether a student was admitted at the time of the lottery. Regression covariates include two years of baseline test scores in math and reading, as well as indicators for the demographic characteristics listed in Table F.3. The elementary regression does not include baseline test scores, and the elementary reading impact estimate is strongly influenced by six outliers.

Supplement: Sample Weighting To Account For Different Admission Probabilities

a. Introduction

Weighting is needed when different types of applicants have different probabilities of admission, otherwise applicant types with different admission probabilities will be over/underrepresented in each condition (that is, probability of being treatment is no longer determined just by chance—as required for an experiment—but by chance and applicant type). Weighting "fixes" this problem by giving applicant types with low admission probability proportionately more representation in the treatment condition and proportionately less representation in control condition (and vice versa for high admission probability types).[16]

The formula for calculating weights is always:

For students admitted, W^T: $\frac{1}{p_i}$, p_i is the probability of admission
For students not admitted, W^C: $\frac{1}{(1-p_i)}$, p_i is the probability of admission

b. Basic Lottery

Calculating p_i when no sibling preference, all one-entry applicants, and not part of a site:

When there is one stratum, $p_i = \frac{N_T}{N}$, N_T is treatment group size[17] and N is total size.

When there are multiple strata,[18] for each stratum, $p_{ij} = \frac{N_{Tj}}{N_j}$, N_{Tj} is treatment group size for stratum j and N is total size for stratum j.[19]

[16] Some lottery participants applied to CMO schools that are not part of the experimental component; we only identify these students if they ultimately ended up attending a CMO school and are thus identified in the student records data. We did not weight for applications to non-experimental component schools. This is crossover (that is, a control participant who receives treatment).

[17] Treatment status is indicated by whether the student was admitted at the time of the lottery.

[18] All strata are assumed to be mutually exclusive. This procedure is used when there are lotteries for multiple grade cohorts in an individual school.

[19] For a stratum to be eligible, it must have at least one student offered admission and one student not offered admission.

c. Some Applicants with Multiple Entries[20]

1. *Some applicants with two entries.*

For two-entry applicants:

$$p_i: \left[1 - \left(\left(\frac{N_{CE}}{N_E}\right) \times \left(\frac{N_{CE}-1}{N_E-1}\right)\right)\right] + \left[\left[\left(\frac{N_{CE}}{N_E} \times \left(\frac{N_{CE}-1}{N_E-1}\right)\right)\right] \times \left(\frac{\left[\left(\left(\frac{N_T}{N_E}\right)\times\left(\frac{N_T-1}{N_E-1}\right)\right)\times N_{2\ entries}\right]}{\left[\left(\frac{N_{CE}}{N_E}\right)\times N_{1\ entry}\right]+\left[\left(\left(\frac{N_{CE}}{N_E}\right)\times\left(\frac{N_{CE}-1}{N_E-1}\right)\right)\times N_{2\ entries}\right]}\right)\right]$$

, N_T is treatment group sample size, N_{CE} is the number of losing entries, N_E is total number of entries, and $N_{1\ entry}$ and $N_{2\ entries}$ are the number of applicants with one and two entries, respectively.[21]

For one-entry applicants:

$$p_i: \left(\frac{N_{TE}}{N_E}\right) + \left[\left(\frac{N_{CE}}{N_E}\right) \times \left(\frac{\left[\left(\left(\frac{N_T}{N_E}\right)\times\left(\frac{N_T-1}{N_E-1}\right)\right)\times N_{2\ entries}\right]}{\left[\left(\frac{N_{CE}}{N_E}\right)\times N_{1\ entry}\right]+\left[\left(\left(\frac{N_{CE}}{N_E}\right)\times\left(\frac{N_{CE}-1}{N_E-1}\right)\right)\times N_{2\ entries}\right]}\right)\right]$$

, N_T is treatment group size, N_{CE} is number of losing entries, N_E is total number of entries, and $N_{1\ entry}$ and $N_{2\ entries}$ are the number of applicants with one and two entries.

2. *Some applicants with two or three entries.*

For three-entry applicants:

p_i:

$$\left[1 - \left(\left(\frac{N_{CE}}{N_E}\right) \times \left(\frac{N_{CE}-1}{N_E-1}\right) \times \left(\frac{N_{CE}-2}{N_E-2}\right)\right)\right] +$$

---

[20] No more than six applicants at any school had multiple entries. If a student had more than one entry in the lottery, we know about the multiple entries regardless of whether they were admitted. The following process occurred when there were multiple entries: (1) a randomized list was created at the time of the lottery and (2) duplicates were deleted before admission offers were sent out (for example, if the school wanted to admit 100 students and 4 of the first hundred listed were duplicates, those 4 would be replaced by non-duplicates so that 100 different students received offers at the time of the lottery). Students with multiple entries always received their highest place on the list and were not punished for submitting multiple entries. The weighting approach is modeled as: (1) applicants have an initial admission probability—based on the number of entries not the number of students—and then (2) if applicants lose, they compete again for slots predicted to open when duplicates are removed.

[21] For applicants with two entries, p is the sum of winning at least one once (literally the probability of losing both lotteries and subtracting that from one) plus the probability of losing but then winning extra openings created when duplicate entries are eliminated by the school (the predicted number of extra slots created when some applicants win twice divided by the predicted number of losing applicants competing for those extra slots).

$$\left[\left[\left(\frac{N_{CE}}{N_E}\right)\times\left(\frac{N_{CE}-1}{N_E-1}\right)\times\left(\frac{N_{CE}-2}{N_E-2}\right)\right]\times \left(\frac{\left[\left(\left(\frac{N_T}{N_E}\right)\times\left(\frac{N_T-1}{N_E-1}\right)\right)\times N_{2\text{ entries}}\right]+\left[3\times\left(\left(\frac{N_T}{N_E}\right)\times\left(\frac{N_T-1}{N_E-1}\right)\times\left(\frac{N_{CE}}{N_E-2}\right)\right)\times N_{3\text{ entries}}\right]+\left[2\times\left(\left(\frac{N_T}{N_E}\right)\times\left(\frac{N_T-1}{N_E-1}\right)\times\left(\frac{N_T-2}{N_E-2}\right)\right)\times N_{3\text{ entries}}\right]}{\left[\left(\frac{N_{CE}}{N_E}\right)\times N_{1\text{ entry}}+\left(\left(\frac{N_{CE}}{N_E}\right)\times\left(\frac{N_{CE}-1}{N_E-1}\right)\right)\times N_{2\text{ entries}}+\left(\left(\frac{N_{CE}}{N_E}\right)\times\left(\frac{N_{CE}-1}{N_E-1}\right)\times\left(\frac{N_{CE}-2}{N_E-2}\right)\right)\times N_{3\text{ entries}}\right]}\right)\right]$$

, N_T is treatment group sample size, N_{CE} is the number of losing entries, N_E is total number of entries, $N_{1\text{ entry}}$, $N_{2\text{ entries}}$, and $N_{3\text{ entries}}$ are the number of applicants with one, two, and three entries, respectively.

For two-entry applicants:

p_i:

$$\left[1-\left(\left(\frac{N_{CE}}{N_E}\right)\times\left(\frac{N_{CE}-1}{N_E-1}\right)\right)\right]+$$

$$\left[\left[\left(\frac{N_{CE}}{N_E}\right)\times\left(\frac{N_{CE}-1}{N_E-1}\right)\right]\times \left(\frac{\left[\left(\left(\frac{N_T}{N_E}\right)\times\left(\frac{N_T-1}{N_E-1}\right)\right)\times N_{2\text{ entries}}\right]+\left[3\times\left(\left(\frac{N_T}{N_E}\right)\times\left(\frac{N_T-1}{N_E-1}\right)\times\left(\frac{N_{CE}}{N_E-2}\right)\right)\times N_{3\text{ entries}}\right]+\left[2\times\left(\left(\frac{N_T}{N_E}\right)\times\left(\frac{N_T-1}{N_E-1}\right)\times\left(\frac{N_T-2}{N_E-2}\right)\right)\times N_{3\text{ entries}}\right]}{\left[\left(\frac{N_{CE}}{N_E}\right)\times N_{1\text{ entry}}+\left(\left(\frac{N_{CE}}{N_E}\right)\times\left(\frac{N_{CE}-1}{N_E-1}\right)\right)\times N_{2\text{ entries}}+\left(\left(\frac{N_{CE}}{N_E}\right)\times\left(\frac{N_{CE}-1}{N_E-1}\right)\times\left(\frac{N_{CE}-2}{N_E-2}\right)\right)\times N_{3\text{ entries}}\right]}\right)\right]$$

, N_T is treatment group sample size, N_{CE} is the number of losing entries, N_E is total number of entries, $N_{1\text{ entry}}$, $N_{2\text{ entries}}$, and $N_{3\text{ entries}}$ are the number of applicants with one, two, and three entries, respectively.

For one-entry applicants:

$$p_i:\left[\frac{N_T}{N_E}\right]+\left[\left(\frac{N_{CE}}{N_E}\right)\times \frac{\left[\left(\left(\frac{N_T}{N_E}\right)\times\left(\frac{N_T-1}{N_E-1}\right)\right)\times N_{2\text{ entries}}\right]+\left[3\times\left(\left(\frac{N_T}{N_E}\right)\times\left(\frac{N_T-1}{N_E-1}\right)\times\left(\frac{N_{CE}}{N_E-2}\right)\right)\times N_{3\text{ entries}}\right]+\left[2\times\left(\left(\frac{N_T}{N_E}\right)\times\left(\frac{N_T-1}{N_E-1}\right)\times\left(\frac{N_T-2}{N_E-2}\right)\right)\times N_{3\text{ entries}}\right]}{\left[\left(\frac{N_{CE}}{N_E}\right)\times N_{1\text{ entry}}+\left(\left(\frac{N_{CE}}{N_E}\right)\times\left(\frac{N_{CE}-1}{N_E-1}\right)\right)\times N_{2\text{ entries}}+\left(\left(\frac{N_{CE}}{N_E}\right)\times\left(\frac{N_{CE}-1}{N_E-1}\right)\times\left(\frac{N_{CE}-2}{N_E-2}\right)\right)\times N_{3\text{ entries}}\right]}\right]$$

, N_T is treatment group sample size, N_{CE} is the number of losing entries, N_E is total number of entries, $N_{1\text{ entry}}$, $N_{2\text{ entries}}$, and $N_{3\text{ entries}}$ are the number of applicants with one, two, and three entries, respectively.

3. As noted in Step #1, applicants with more than three lottery entries should not be listed on the Baseline Sample sheet.

d. Sibling Preference[22]

Charter schools typically use one of three approaches for dealing with siblings who apply together to schools (Gleason et al., 2010):

1. *Sibling applicants are treated no differently than other applicants.* Since siblings have the same probability of admission as non-siblings, the normal weighting approach can be used.
2. *All siblings are treated as a single unit in the lottery* (e.g., all sibling names are written on a single piece of paper; if that piece of paper is selected all students are admitted and if not none are admitted). Each sibling has the same probability of admission as a non-sibling so there is no need for special weighting. In this case $p_i = \frac{N_T}{N}$, N_T is total number of winning entries at the stratum/school and N is total number of entries at the stratum/school *(i.e., a sibling pair counts as one entry, as does one non-sibling student).*[23]
3. *Each sibling has their own lottery entry, but if one sibling "wins" both siblings are offered admission.* Following the Gleason et al. (2010) approach, the probability of admission is calculated differently depending on the number of sibling pairs.
   a. *One sibling pair.*
      - For each sibling: p_i: $\left(\frac{2\times N_T}{N}\right) - \left[\frac{N_T}{N} \times \left(\frac{(N_T-1)}{(N-1)}\right)\right]$, N_T is treatment group sample size at the stratum/school and N is total sample size at the stratum/school.
      - For each non-sibling, p_i: $\left\{\left(\frac{N_T}{(N-2)}\right) \times \left[1 - \left(\frac{2\times N_T}{N}\right) + \left(\frac{N_T}{N} \times \frac{(N_T-1)}{(N-1)}\right)\right]\right\} + \left\{\left(\frac{(N_T-2)}{(N-2)}\right) \times \left[\left(\frac{2\times N_T}{N}\right) - \left(\frac{N_T}{N} \times \left(\frac{(N_T-1)}{(N-1)}\right)\right)\right]\right\}$, N_T is treatment group sample size at the stratum/school and N is total sample size at the stratum/school.

   b. *More than one sibling pair.* An assumption is made on the number of siblings that would be offered a spot despite losing the lottery.[24] Specifically, estimate the expected number of siblings offered admission to the school despite a losing lottery draw ("win-by-sibling" or WBS). The probability of WBS for each sibling pair is: $\frac{2\times N_T}{N} \times \left(\frac{N_C}{(N-1)}\right)$.[25]

---

[22] This is only for cases when there are sibling pairs (e.g., twins); sets of three siblings (triplets) uses a different approach.

[23] No school with this sibling approach had applicants with multiple entries.

[24] The situation becomes more complicated when more than one set of siblings applies, because the probability that one set of siblings is admitted depends on what happens with the other sets. Specifically, having one winning sibling in a pair means there is one fewer slot for other siblings because the winning siblings brings in their sibling.

[25] For triplets, the probability of one sibling of the triplet winning-by-sibling is $3 \times \frac{N_T}{N} \times \frac{N_T-1}{N-1} \times \frac{N_C}{N-2}$ and the probability of two siblings of the triplet winning-by sibling is $3 \times \frac{N_T}{N} \times \frac{N_C}{N-1} \times \frac{N_C-1}{N-2}$. (Multiply by 3, because there are three different ways this can occur.)

- For non-siblings, the estimated slots (ES) taken by these siblings is: (# sibling pairs) × P_{WBS}.[26] Thus p_i: $\frac{(N_T - ES)}{(N-ES)}$

- For siblings, the estimated slots (ES) taken by these siblings is: (# sibling pairs – 1) × P_{WBS}. Thus p_i: $\frac{(N_T - ES)}{(N-ES)}$ [27]

e. Sites

A "site" exists when experimental schools (i.e., applicants applied to two or more experimental schools); a site includes all experimental schools that shared applicants.[28]

For applicants that only apply to one experimental component school in the site, site p_i = school p_i.

For applicants that apply to two experimental component schools at the site, (X & Y), $p_i = 1 - \left((1 - px_i) \times (1 - py_i)\right)$, px_i and py_i are the probabilities of receiving admission to School X and Y respectively.[29]

For applicants that apply to three experimental component schools at the site, (X, Y & Z), $p_i = 1 - \left((1 - px_i) \times (1 - py_i) \times (1 - py_i)\right)$, px_i, py_i, and pz_i are the probabilities of receiving admission to Schools X, Y and Z, respectively.[30]

For example, if one pair of siblings applied to School X and School Y and School X had sibling preference #1 and School Y had sibling preference #2, $p_i = 1 - \left((1 - px_i) \times (1 - p_{iY})\right)$, px_i and py_i are the probabilities of receiving admission to School X and Y respectively.[31]

f. Normalizing Weights.

To make the weights more intuitive (more representative of sample sizes) and appropriate for calculating standard deviations, they were rescaled ("normalize") so that ∑w=N for each site. To normalize, calculate the normalization factor: $\frac{N}{\sum_{i=1}^{N} weight_i}$and multiply each weight by this factor.

---

[26] When there is one set of triplets plus other sibling pairs, ES is: [$P_{WBS(2\ \text{siblings accepted})}$×2]+[$P_{WBS(1\ \text{sibling accepted})}$] + [$P_{WBS}$×(# sibling pairs)].

[27] When there is one set of triplets plus other sibling pairs, the ES for triplets is: (# sibling pairs) × P_{WBS}. For sibling pairs, the ES is: [$P_{WBS(2\ \text{siblings accepted})}$×2]+[$P_{WBS(1\ \text{sibling accepted})}$] + [$P_{WBS}$×(# sibling pairs-1)].

[28] To be included in a site, lotteries must have at least one T and one C student with data. A lottery is the randomization unit, either a stratum, a school, or in some cases multiple schools.

[29] This is literally estimating the probability of losing both lotteries and subtracting that from one (this difference equals the probability of winning at least one lottery).

[30] This is literally estimating the probability of losing all three lotteries and subtracting that from one (this difference equals the probability of winning at least one lottery).

[31] School-level weights were calculated first. Lotteries were then pooled and records for multiple-school students were collapsed. Weights were then recalculated for these multiple-school students using this formula.

APPENDIX C

VALIDATION OF IMPACT ESTIMATION APPROACH

I. Introduction

One of the central goals of this study is to rigorously assess the impacts of CMOs on student academic outcomes using non-experimental methods. This appendix describes how we examined whether a non-experimental, panel-based (NXP) research design using a propensity-score matching (PSM) approach can produce impact estimates that correspond to those produced by randomized experimental methods—the "gold standard" method for causal inference.

This validation effort, conducted for a subset of CMO schools in which experimental analyses could also be conducted, finds that PSM produces impact estimates that are very similar to the benchmark experimental impact estimates in middle and high schools.[32] In both reading and math, PSM estimates differ from experimental estimates by 0.01 to 0.03 standard deviation units. Site-specific estimates across seven lottery sites indicate that PSM results correlate with experimental results at levels of 0.9 or higher. The differences between the PSM and experimental estimates are not statistically significant. As a sensitivity check, two alternate NXP approaches, exact matching (EM) and ordinary least-squares (OLS) regression, also produced impact estimates similar to the experimental method. These results provide evidence supporting the rigor of nationwide NXP impact estimates for the larger set of CMO schools where experimental methods are not feasible.

Elementary schools were not included in this validation, because most students start at an elementary school in kindergarten or first grade, meaning there are no pre-entry baseline test scores that can be included in the NXP analysis. Pre-entry test scores are likely to be critical in the replication effort (Cook et al., 2008). The typical state testing pattern also means that we have no achievement outcomes to measure until three or four years after kindergarten entry lotteries (as standardized tests are not administered until spring of second or third grade); we were able to identify only three schools with lottery records far enough in the past to allow us to observe these test scores, and the sample size was too small to be reliable (N=118). Given these obstacles, the validation focuses on middle and high schools.

This appendix begins by explaining the need for the validation, and how it allows the larger study to take advantage of the different strengths of both experimental and NXP methods. We then describe the data used in the analysis followed by three methodology sections that describe the validation methods, the experimental methods, and the NXP/PSM methods. Finally, we present detailed results of the validation analysis and conclusions.

II. Why Validate?

Properly designed and implemented randomized experiments produce impact estimates that support stronger causal conclusions than any other method, by ensuring that the treatment and

---

[32] This validation included high schools because we plan to estimate high school impacts on test scores; these impacts will be reported for a broader sample of high schools in a future report.

control groups are similar on observed and unobserved characteristics prior to receiving an intervention (Shadish et al., 2002; Murnane and Willet, 2011). Thus, any statistically significant difference between group outcomes can be attributed to the impact of the intervention.

In the charter-school context, randomized experiments can be conducted using the admission lotteries that oversubscribed schools conduct (Gleason, et al., 2010; Dobbie and Fryer, 2011; Angrist et al., 2010). However, not all charter schools are oversubscribed; not all oversubscribed schools use lotteries; and not all schools using lotteries keep good records of winners and losers. Researchers have found that admissions lotteries can be used for experimental analysis in only a small proportion of charter schools nationwide.[33] Of the 161 CMO middle and high schools in the target sample for the overall study (that is, schools with entry grades of 4-9), adequate data for a rigorous experimental analysis was available for only 12 schools and only select grades and cohorts of those schools. Moreover, it would not be surprising if the (oversubscribed, well-organized) schools where experimental analysis is possible are systematically different from the schools where it is not possible (see Abdulkadiroglu et al., 2009 for suggestive evidence of this). In sum, admissions lotteries can be used to conduct experiments producing strong causal inferences about a small subset of CMO schools, but they cannot be used to examine the impacts of most CMO schools across the country.

In contrast, student records data are available to conduct NXP analyses for a large proportion of all CMO schools nationwide. NXP methods for estimating CMO impacts rely on longitudinally-linked data on individual students before and after they enter CMO schools. Our preferred NXP approach, propensity-score matching,[34] involves comparing the achievement of CMO students to non-CMO students who have a similar estimated likelihood of enrolling in a CMO school (that is, similar baseline achievement and other characteristics). Although NXP methods lack the strong causal validity of randomized experiments, because matched students may differ on unobserved characteristics, they allow assessment of the impacts of many more schools and CMOs.

Thus, in this study experimental and NXP methods involve a tradeoff between internal and external validity. An experimental analysis on a small number of schools and CMOs included would not be representative of typical CMOs and schools. A NXP analysis would not enable strong causal inference, as students who attend CMO schools could be different in unobserved ways from the PSM-identified comparison students. By conducting experimental and PSM analyses in a subset of schools, we can assess whether PSM produces results that match the experimental impact estimates for the same schools and the same students. If the PSM approach successfully replicates experimental impact estimates, the replication provides greater confidence that the PSM approach will produce valid impact estimates when applied to the nationwide population of CMO middle and high schools.

---

[33] One national study found that less than 15 percent of charter middle schools could be included in a lottery-based analysis (Tuttle, Gleason, and Clark, 2011).

[34] Three common NXP approaches were considered: PSM, EM and OLS regression. PSM was preferred over EM because some schools districts were too small to provide exact matches for many CMO students. PSM was preferred over OLS, because PSM creates a plausible counterfactual and thus can meet WWC standards. We describe the PSM approach in depth in Appendix C.

III. Data

a. Description

We use student-level administrative data provided by state departments of education, school districts, and CMOs. Our validation sample includes data from four jurisdictions. The treatment schools included 12 oversubscribed middle and high schools across seven CMOs in 2006-2007, 2007-2008, or 2009-2010 academic years. Three of the schools had oversubscribed lotteries in multiple years.

We focused on two outcomes: reading/English language arts and math scores[35] on state achievement tests one year after students enrolled in school following participation in a lottery (year 1). Where data were available, we standardized test scores using state-level means and standard deviations for each grade and cohort. Otherwise, we used district-level means and standard deviations for test score standardization.

Student characteristics available for both the experimental and NXP analyses were: baseline reading and math test scores (including missing test score indicators), sex, race/ethnicity (African-American, Hispanic, white/other), baseline free- or reduced-price lunch (FRPL) eligibility status,[36] English language learner[37] (ELL) status, special education status (IEP), and an indicator of whether a student attended a charter school in the baseline year. The analyses also included indicator controls for the school students applied to, with groups of schools that share applicants combined into experimental "sites." Each site has a common jurisdiction, lottery year, and grade level, and thus the site controls for these factors.

b. Diversity of Validation CMOs

The seven CMOs included in the validation exercise were quite diverse. The sample CMOs had schools in three of the four U.S. Census regions: Northeast, South, and West. The sample CMOs had 80 schools in total, including 25 middle schools eligible for this study's primary (NXP) impact analysis.

Table C.1 provides baseline (pre-enrollment) student characteristics for middle schools in six validation CMOs (one of the validation CMOs only had high schools) and baseline characteristics for middle schools of the other 16 CMOs in the middle school analysis. The validation CMOs are diverse. Prior to enrolling in the validation CMO middle schools, standardized student test scores were as low as -.11 and 0.08 in reading and math respectively, and as high as .63 and .53. The validation CMOs had as little as 5 percent of special education students with an individualized education plan and as much as 14 percent. The percentage of students that were English language learners varies from 4 percent to 33 percent in validation CMOs. Finally, the validation CMOs were

---

[35] In some states, the specific math tests taken in the 8th and 9th grade vary. Because the test taken could depend on whether the student enrolled in a CMO school, in these states we were not able to include students entering high school in the math analysis.

[36] One district did not have reliable information on students' free- or reduced-price lunch status.

[37] One district did not include information on students' English language learner status.

diverse in terms of race/ethnicity: the percentage of African-American students ranged from 11 percent to 81 percent and the percentage of Hispanic students ranged from 17 percent to 75 percent.

Table C.1. Baseline Statistics for Validation and Non-Validation CMO Middle Schools

CMO	Reading Score	Math Score	Percentage Black	Percentage Hispanic
1	-0.08	-0.11	Medium	Medium
2	-0.07	0.04	Low	High
3	-0.07	0.16	Low	High
4	0.02	-0.04	Medium	Low
5	0.05	-0.03	High	Low
6	0.63	0.53	Low	Low
7	-0.62	-0.36	Low	High
8	-0.46	-0.46	Medium	Medium
9	-0.07	-0.07	Low	Low
10	-0.04	0.10	Low	High
11	0.00	-0.04	Low	High
12	0.01	-0.04	Medium	Medium
13	0.01	0.02	High	Low
14	0.03	-0.01	High	Low
15	0.04	-0.16	High	Low
16	0.05	-0.09	High	Low
17	0.12	0.19	Low	High
18	0.20	0.22	Low	High
19	0.24	0.11	Low	High
20	0.27	0.27	Low	Medium
21	0.35	0.46	Low	Medium
22	0.89	0.93	Low	Medium

(1) Notes: Validation CMOs are shaded gray. Within validation and non-validation groupings, CMOs are ordered first by reading score (lowest to highest) and then math score. If the percentage African-American or Latino students was between 0 and 33 percent the CMO is labeled low, CMOs with percentages between 34 and 67 percent are labeled medium, and CMOs with percentages greater than 67 percent are labeled high.

IV. Replication Procedure

a. Replicating Experimental ITT Estimates

Two common analyses in the charter school literature are estimating the effect of receiving an offer of admission to a charter schools (an intent-to-treat or ITT estimate) or the effect of attending a charter school (an effect of treatment-on-the-treated or TOT estimate). These estimates are not identical, because some lottery winners decline their offers, and some losers at the time of the lottery end up enrolling in the schools as a result of wait-list admissions or other post-lottery changes. Both estimates are of interest as a policy matter: the ITT impact is relevant in assessing the likely effects of making more CMO enrollment spaces available; and the TOT impact is relevant in assessing the effect that can be expected for a student who enrolls. These different analyses estimate impacts for different groups of students and make different assumptions.

We chose to use ITT estimates for two reasons. First, it is not possible to generate EX TOT and NXP TOT estimates for the same students when there is control cross-over. The conventional EX TOT estimate, estimated using instrumental variables, is an impact estimate only for students induced into enrollment by the lottery offer (i.e., a local average treatment effect for "compliers"; see Angrist, Imbens, and Rubin, 1996). However, these complying students cannot be identified among all the admitted students who enroll, and thus we cannot estimate NXP estimates only for these compliers—any NXP estimates will include both compliers and "always-takers" (students who enroll regardless of whether admitted at lottery) who could experience different impacts. Second, experimental ITT estimates make fewer assumptions than experimental TOT estimates and therefore have stronger causal validity. Specifically, the validity of the experimental TOT impact estimates depends on the assumption that crossover students—lottery losers who enroll in the CMO schools—experience the same impacts as lottery winners who would have enrolled in CMO schools regardless of whether they won the lottery.[38] Since the assignment variable for our lotteries is winning a lottery at the time of the lottery (as discussed in more detail in section V), the ITT estimates include substantial crossover of control students into treatment. As crossover students receive offers later, it is possible that their enrollment in the CMO begins after the start of the school year, which could mean that crossover students experience a different treatment and have different impacts than the lottery winners. (Unfortunately, the data do not indicate how often this occurs.) Smaller actual impacts for crossover students would produce TOT estimates that are biased upward. As the goal of the experimental analysis is to provide the most causally valid estimates to use as a benchmark, we report ITT estimates.

An experimental ITT estimate is, in basic form, simply the average outcome for treatment students minus the average outcome for control students. In the absence of crossover, replicating experimental impact estimates is therefore equivalent to replicating the experimental control group (given that the treatment group is defined to be the same in experimental and NXP analyses). When some control group students cross over to attend treatment schools ("contamination"), however, replication is more complicated. It is impossible for an NXP approach to replicate the part of the experimental control group that crosses over, because the NXP approach finds comparison students only among the population that did *not* enroll in treatment schools. Nonetheless, an NXP approach can attempt to replicate the experimental ITT estimate by splitting the estimate into components corresponding to the different groups of students who actually receive treatment. The supplement to this appendix derives the following quasi-experimental estimate as the most equivalent to the experimental ITT:

(1) NXP ITT estimate = $P_T \times B_T - P_C \times B_C$

Where P_T is treatment group's CMO enrollment rate, B_T is the CMO impact on treatment enrollees, P_C is control group's CMO enrollment rate (crossover rate), and B_C is the CMO impact on control enrollees.

[38] In other words, among the students who would attend the CMO schools regardless of whether they won or lost the lottery (that is "always-takers"), impacts experienced by the lottery losers must be equivalent to impacts experienced by the lottery winners.

This equation has the virtue of allowing the effects of CMOs on the control crossover students to differ from the effects on lottery winners who attend the schools. To replicate experimental ITT results non-experimentally, we subtract the estimated NXP effects on the crossover students from the effects on the lottery winners who attend the schools. P_T and P_C are observed in the lottery data, and B_T and B_C are separately estimated using PSM on the experimental treatment and control students who enroll in CMO schools.[39]

b. First-Year Achievement Impacts in Math and Reading

Many possible experimental estimates could be replicated in this particular study since there are test scores in reading and math across 12 schools, six CMOs, seven lottery sites, four jurisdictions, and multiple grades and cohorts. To maximize statistical power and the precision of the estimates, we chose (in advance of estimation) to pool all lotteries across cohorts, grades, schools, and CMOs for the students' first year in the school to permit the inclusion of the most-recent lotteries. This provided two overall impact estimates—one for reading and one for math. In addition to producing primary estimates of impacts across sites, we also examine the correlation between first-year experimental and NXP estimates site-by-site.

V. Experimental Methods

a. Sample and Baseline Equivalence

The experimental sample frame consists of students who applied to an oversubscribed CMO school that used a random lottery to admit students. The treatment group is composed of applicants offered admission to a participating CMO school *at the time of the lottery*.[40] Applicants not offered admission at the time of the lottery form the control group. All students who provided consent, were in the correct application grade at the time of the lottery, were randomized in the lottery, were in the proper grade for the lottery, and had baseline test scores were included in the analysis.

Some students applied to more than one experimental CMO school, meaning they could receive an offer to one CMO school even if they lost lotteries at the other school(s). In these cases, we treated all schools sharing applicants as a single site. Further restrictions were made at the site-level to ensure the validity and power of the impact estimates. Experimental sites had to meet each of the following criteria to be included in the analysis:

[39] Experimental impact estimates are limited not only to the subset of schools in which lottery-based analysis is possible, but also to the subset of cohorts and individual students who were randomly assigned to treatment via the lotteries. The NXP treatment sample only included the enrollees from the experimental analysis, excluding students in other years, grades, and cohorts, along with students who did not participate in the randomized lottery.

[40] The CMO enrollment rates of students admitted at the time of the lottery are substantially higher than students rejected at the time of the lottery, meaning that this measure provides enough random assignment to fairly test CMO impacts. Angrist et al. (2010) used a similar approach. A traditional experimental approach where assignment is based on whether students were ever admitted to a CMO school was not possible for two reasons. First, many schools with oversubscribed lotteries ultimately admitted all students who not admitted at the time of the lottery, meaning there were no randomized students who never received an offer (that is, no control students). Second, many schools did not follow the randomization order when admitting students after the lottery.

1. The overall and differential attrition rates must be lower than the What Works Clearinghouse maximum thresholds (liberal attrition standard);
2. If we did not observe the lottery and consequently were unsure of the randomization validity, any difference between treatment and control average baseline test scores must be less than 0.25 effect size and demographic differences must be less than 25 percentage points;[41]
3. The difference in CMO school enrollment between treatment and control groups must be at least 20 percentage points.

After these exclusions, there were 579 treatment and 809 control students with baseline data who were eligible for the reading analysis, and 331 treatment and 574 control students with baseline data who were eligible for the math analysis.[42] In the reading impact analysis, we excluded 52 treatment and 74 control students because we were unable to obtain outcome data or they were in the wrong grade in the outcome year,[43] leaving a final analysis sample size of 527 treatment and 735 control students. In the math analysis, we excluded 13 treatment and 26 control students for the same reasons, leaving a final analysis sample size of 318 treatment and 548 control students. Overall attrition in the reading sample was 9 percent, with no differential attrition between the treatment and control conditions. Overall attrition in the math sample was 4 percent, with a 1 percentage point difference between attrition in the treatment and control conditions. The low attrition levels in this study are unlikely to significantly bias impact estimates, according to the What Works Clearinghouse attrition standards.

Consistent with minimal bias, baseline statistics of observable characteristics indicate that the final treatment and control groups were very similar for both reading and math impact analyses (see Table C.1) with no statistically significant differences (all p-values>.10). Table C.2 presents baseline statistics for students included in the math and reading analysis samples. Table C.2 does not include imputed values, but values for baseline test scores and demographic characteristics were imputed in the analysis. For baseline and pre-baseline test scores, we included a missing data indicator and set each missing test score to the state or district-level mean, which is zero by design. For students missing demographic variables (race/ethnicity, gender, FRPL, LEP, IEP, baseline charter status), we recoded the missing values for these covariates to the mode across all students in the sample (not an English language learner, no IEP, not attending a charter school at baseline, receiving free/reduced price lunch, female, and Hispanic). We did not impute outcome test scores. Students who were missing either a math or reading test score in the follow-up year were excluded from the analysis when that test score was the outcome variable.

---

[41] The effect size measure for test was Hedge's g. Relatively large baseline differences were allowed because some of the sites were small and could have moderate baseline differences even if the randomization was sound.

[42] The math analysis also excludes 396 students applying to a high school because the math test taken in the ninth grade in some states could be affected by treatment (that is, in some states ninth-graders take a course-specific rather than a grade-specific exam, and the school may affect the course students take).

[43] These students without outcome test scores most likely attended a private school or an independent charter school that did not provide data to their district. The students in the wrong grade in the outcome year either repeated or skipped a grade in the outcome year.

Table C.2. Baseline Statistics for Treatment and Control Groups

Characteristic	T Mean	C Mean	Diff (T-C)	p-value
Reading Analysis (n= 527 T and 735 C)				
Baseline Reading Score	.11	.07	.04	0.497
Baseline Math Score	.08	.02	.06	0.328
Prebaseline Reading	.16	.08	.08	0.183
Prebaseline Math	.13	.12	.01	0.949
% Male	.51	.50	.01	0.653
% Black	.43	.43	.00	0.926
% Hispanic	.49	.51	-.02	0.648
% White/Other	.07	.06	.01	0.425
Baseline % FRPL	.76	.75	.01	0.850
Baseline % LEP	.09	.10	-.01	0.859
Baseline % IEP	.06	.08	-.02	0.313
Baseline % Charter Attendance	.09	.11	-.02	0.465
Math Analysis (n= 318 T and 548 C)				
Baseline Reading Score	.12	.09	.03	0.696
Baseline Math Score	.08	.02	.06	0.313
Prebaseline Reading	.14	.09	.05	0.454
Prebaseline Math	.10	.08	.02	0.817
% Male	.53	.47	.06	0.128
% Black	.61	.62	-.01	0.857
% Hispanic	.37	.36	.01	0.823
% White/Other	.02	.02	.00	0.883
Baseline % FRPL	.83	.82	.01	0.907
Baseline % LEP	.06	.05	.01	0.542
Baseline % IEP	.06	.08	-.02	0.437
Baseline % Charter Attendance	.04	.02	.02	0.208

Notes: These tables include only students who were included in the reading and math analyses. Students with missing reading outcome data are excluded from the reading analysis sample; likewise, students with missing math outcome data are excluded from the math analysis sample. Reading and math test scores are standardized using the state district mean and standard deviation. All statistics are weighted to account for admission probabilities.

b. Experimental ITT Estimation and Weights

To estimate an experimental intent-to-treat impact, we compared outcomes of applicants offered admission at the time of the lottery to those of applicants rejected at the time of the lottery, controlling for students' previous test scores and demographic characteristics.[44] The impact estimation model is:

(2) $$y_i = \propto + X_i\beta + \delta T_i + S_i\theta + \epsilon_i,$$

where y_i is the reading or math test score outcome for student i; α is the intercept; X_i is a vector of achievement and demographic characteristics (see Table IV.1); T_i is a binary variable for treatment status, indicating whether student i was admitted at the admission lottery; S_i is a vector of indicators identifying which site the student applied to; ε is a random error term that reflects the influence of unobserved factors on the outcome; and β, δ, and θ are vectors of parameters or parameters to be estimated. The estimated coefficient on treatment status, δ, represents the impact of admission to a CMO school at the time of the lottery. The model assumes that the treatment indicator and covariates influence the outcomes in the same way across all sites. To examine how CMO impacts varied by site, we also estimated a model that interacted the treatment indicator with each site indicator. The site-specific impact estimation model is:

(3) $$y_i = \propto + X_i\beta + \delta T_i + S_i\theta + (T_i \times S_i)\varphi + \epsilon_i,$$

where all the terms are the same as in Equation 2 except for φ, the estimated coefficient on the interaction terms, which is a vector of estimated coefficients on the interaction terms. Students were weighted to account for admission probabilities.

VI. Propensity-Score Matching Method

In order to replicate the experimental ITT impact estimate, we require two sets of non-experimental impact estimates: a treatment group enrollees' impact estimate and a control group enrollees' (crossover) impact estimate (following Equation 1). The treatment group enrollees are lottery winners who attended an experimental CMO school. The control-group enrollees are lottery losers who attended an experimental CMO school. For both groups of enrollees, comparison groups are identified among students who did not attend an experimental CMO school.

a. Estimating Propensity Scores

The first step is to estimate a propensity score for each student in the sample. To determine the appropriate propensity score model for each of the two enrollee groups, we use a forward model selection procedure for the logistic regression. Because baseline math and reading test scores are some of the strongest predictors of later outcomes, we specify that the model-building procedure begins with the model containing the two baseline test scores and corresponding missing test score indicators. At each subsequent step, the forward procedure adds a term from a specified set of

[44] As student admission to CMO schools was randomly determined, we could simply compare the mean outcomes of the treatment and control groups. However, to obtain more precise impact estimates, we adjust for baseline student characteristics in a regression model.

potential covariates to optimize model fit to the data. The procedure could select from a list of 52 potential covariates: the 11 observed baseline covariates, 39 two-way interactions of these covariates, and 2 interactions of test scores with themselves (i.e., quadratic terms). Table C.3 shows baseline covariates and interaction terms included in the final propensity models. These models fit the data well as indicated by the Hosmer and Lemeshow Goodness-of-Fit test *p*-values. The propensity-score method is described in more detail in Appendix D.

Table C.3. Covariates Included in the Final Propensity Score Models

	Treatment-Group Enrollees	Control-Group Enrollees (Crossovers)
Baseline Covariates And Interaction Terms	Baseline Math Test Score (MATH) And Corresponding Missing Indicator (missmath)	Baseline Math Test Score (math) And Corresponding Missing Indicator (missmath)
	Baseline Reading Test Score (read) And Corresponding Missing Indicator (missread)	Baseline Reading Test Score (read) And Corresponding Missing Indicator (missread)
	Sex	Sex
	Free/Reduced Price Lunch (FRPL)	Free/Reduced Price Lunch (FRPL)
	Special Education (iep)	English Language Learner (ELL)
	Jurisdiction	Jurisdiction
	Grade	Grade
	missmath*sex	math*read
	missmath*frpl	math*ell
	read*frpl	missmath*ell
		sex*ell
		math*Frpl
		math*site
		read*site
		sex*site
Hosmer & Lemeshow Goodness-Of-Fit Test p-value	0.45	0.78

b. Select Closely-Matched Comparison Students

After estimating the propensity scores, we identified comparison students whose estimated propensity scores were similar to those of each treatment student (that is, comparison students who had similar probabilities of enrolling in CMO schools). The selection used caliper matching, whereby a given treatment student was matched to all comparison students with estimated propensity scores within a specified range (or caliper), rather than merely selecting a specified number of nearest neighbors. The sampling occurred with replacement. The matching procedure was implemented separately for each jurisdiction. To improve statistical precision, we selected multiple comparison students for each treatment student.

For math outcome samples, the matched comparison students on average had similar math and reading test scores as the treatment students (Table C.4 and C.5). They also had similar distributions on all demographic covariates, with the exception of race/ethnicity and baseline charter school attendance.[45] The results were similar for reading outcome samples (not shown).

Table C.4. Baseline Statistics for Treatment-Group Enrollees (Math Outcome)

Prior Student Achievement or Student Characteristic	Treatment (n= 200) Mean/Percentage	Comparison (n= 5,905) Mean/Percentage	Difference
Baseline Math Score	0.10	0.07	0.02
Baseline Reading Score	0.10	0.06	0.04
Race/Ethnicity			
African American	0.58	0.30	0.28**
Hispanic	0.40	0.48	-0.08**
White/other	0.02	0.22	-0.20**
Male	0.60	0.59	0.01
Free/Reduced Price Lunch	0.84	0.87	-0.03
Special Education	0.05	0.05	0.00
English Language Learner	0.06	0.11	-0.05*
Attended Charter School at Baseline	0.05	0.02	0.03**

*Significantly different from zero at the .10 level, two-tailed test.
**Significantly different from zero at the .05 level, two-tailed test.

Table C.5. Baseline Statistics for Control-Group Enrollees (Math Outcome)

Prior Student Achievement or Student Characteristic	Treatment (n= 124) Mean/Percentage	Comparison (n=3,695) Mean/Percentage	Difference
Baseline Math Score	-0.01	0.03	-0.03
Baseline Reading Score	0.05	0.08	-0.03
Race/Ethnicity			
African American	0.56	0.32	0.25**
Hispanic	0.42	0.46	-0.04**
White/other	0.02	0.22	-0.20**
Male	0.40	0.40	0.00
Free/Reduced Price Lunch	0.89	0.84	0.05
Special Education	0.08	0.10	-0.02
English Language Learner	0.02	0.02	0.00
Attended Charter School at Baseline	0.01	0.02	-0.01

*Significantly different from zero at the .10 level.
**Significantly different from zero at the .05 level.

[45] There were only a few students who were not African-American or Hispanic in the treatment group. While the race/ethnicity variable was selected by the model selection procedure, the associated coefficients had large standard errors. As a result, race/ethnicity were excluded from the propensity score matching model, resulting in an imbalance between treatment and matched comparison students. However, race/ethnicity was a covariate in the impact estimation model. In addition, the PS estimates are almost identical to those from the exact matching approach which included race/ethnicity as a matching characteristic. This suggests that our approach was robust to the exclusion of this variable from the propensity model. Similar problems occurred with baseline charter school attendance.

c. Impact Model

Following the creation of matched samples, we estimated impacts using an OLS regression model; covariates were included to improve statistical precision and to control for any remaining differences in baseline characteristics. The regression model was identical to the model used in ITT experimental analysis—Equations 2 and 3. Here, however, the treatment indicator, T, corresponds to each of the two enrollee groups (treatment-group enrollees and control-group enrollees) defined at the beginning of this section. The parameter of interest in Equation (4) is β, which corresponds to the impact estimate.

In estimating impacts, enrolled students were weighted to account for the probability of winning a lottery admission offer. The matched comparison students are assigned the analysis weight for the enrolled students to whom they are matched. The experimental weights were rescaled so that a given site has the same weight in both the experimental and the PSM approaches. This weighting ensures that any potential differences between experimental and PSM estimated impacts can be attributed to the approaches themselves, rather than differences in weights.

d. Alternative NXP Approaches: Exact Matching and OLS Regression

As noted, two alternate NXP approaches were used to test the sensitivity of results to the particular NXP approach. These alternate approaches are exact matching (EM) and ordinary least squares regression (OLS) without matching.

Exact matching uses comparison group students who exactly match treatment students on a set of demographic characteristics and have very similar baseline test scores (CREDO, 2009). To be selected, the comparison students had to exactly match the treatment students on the following categorical characteristics: baseline charter school attendance, sex, race/ethnicity, FRPL eligibility status, English language learner status, individualized education program status (special education), grade in outcome year, cohort, and jurisdiction. Exact matching on continuous characteristics—such as baseline math and reading test scores—would rarely identify matches, so we define a comparison student to be an exact match if his or her test score falls within 0.10 standard deviations of the treatment student's baseline test score in the same subject. Following the creation of the matched comparison group, impacts were estimated using the same regression model used in the experimental and PSM analyses.

The OLS-only approach does not attempt to create a matched comparison group of students. Instead, the approach uses the entire population of non-CMO students in the local jurisdiction as comparisons, relying entirely on covariates to adjust for baseline differences between treatment students and other students. The same OLS regression model used to estimate impacts in all of the other approaches was used.

VII. Results and Conclusions

The PSM approach successfully replicates average experimental ITT impact estimates for the 12 CMO schools in the replication sample. ITT impact estimates for the two approaches are reported in Table C.6. PSM produces ITT impact estimates within 0.01 of the experimental ITT estimate in math and 0.03 of the experimental ITT estimate in reading. Neither of the differences is statistically significant, and they differ in opposite directions (that is they do not consistently under- or over-estimate impacts). In short, we find no evidence that PSM produces biased impact estimates.[46] The last two rows of Table C.6 show that exact matching and OLS, like PSM, produce impact estimates that are very close to experimental impact estimates.

Table C.6. Experimental and NXP Impact Estimates in 12 CMO Schools

	Math		Reading	
	Impact	SE	Impact	SE
Experimental ITT	0.09	0.04	0.00	0.04
PSM ITT (baseline NXP approach)	0.08	0.07	0.03	0.06
Alternate NXP Estimates				
ITT Based on Exact Matching (EM)	0.08	0.07	0.02	0.06
ITT Based on OLS	0.07	0.07	0.02	0.06

Moreover, NXP impact estimates at the site level are very similar to experimental site estimates, as the high correlations in Table C.7 indicate. At the site level, PSM ITT impact estimates correlate with experimental ITT impact estimates at 0.97 in math and 0.90 in reading. Site-level results are also very similar to experimental impacts for exact matching and OLS approaches.

Table C.7. Experimental and NXP Impact Estimates at the Site Level

	Reading				Math			
Lottery Site	EXP	PSM	EM	OLS	EXP	PSM	EM	OLS
A	0.31	0.36	0.38	0.34	0.05	0.12	0.08	0.10
B	n/a	n/a	n/a	n/a	0.14	0.17	0.16	0.13
C	-0.11	-0.08	-0.11	-0.13	-0.15	-0.21	-0.25	-0.18
D	0.05	0.03	0.06	0.03	-0.10	-0.04	-0.01	-0.03
E	0.19	0.14	0.11	0.14	0.11	0.11	0.11	0.08
F	-0.03	-0.08	-0.07	-0.06	-0.06	-0.11	-0.08	-0.09
G	n/a	n/a	n/a	n/a	-0.10	-0.01	-0.04	-0.002
Correlation with EXP	n/a	0.97	0.96	0.99	n/a	0.90	0.90	0.88

[46] Standard errors for NXP estimates are rough. They assume zero covariance among site-specific impact estimators and between impacts on treatment and control enrollees.

In sum, this validation analysis suggests that propensity-score matching with baseline test scores is capable of producing impact estimates that are not only unbiased but that also replicate experimental impact estimates with a high degree of precision.

Supplement: NXP Derivation of Experimental ITT Estimate

Notation:
μ_k^A= mean outcome of always-takers in experimental group k, where $k \in \{T, C\}$
μ_k^{CP}= mean outcome of compliers in experimental group k, where $k \in \{T, C\}$
μ_k^N= mean outcome of never-takers in experimental group k, where $k \in \{T, C\}$
p^A= proportion of lottery participants who are always-takers (is equivalent between T and C)
p^{CP}= proportion of lottery participants who are compliers (is equivalent between T and C)
p^N= proportion of lottery participants who are never-takers (is equivalent between T and C)
p_k^E= proportion of experimental group k that enrolled in CMOs
Note that $p_T^E = p^A + p^{CP}$ and $p_C^E = p^A$.

Not all of the μ parameters can be estimated. In fact, only μ_T^N and μ_C^A are observable. However, we can always estimate the *p* parameters because the percentage of always-takers, compliers, and never-takers in each treatment group is the same due to randomization, and p^A and p^N are observed enabling the calculation of p^C.

Experimental ITT Impacts:

(1) $ITT^{EXP} = (p^A\mu_T^A + p^{CP}\mu_T^{CP} + p^N\mu_T^N) - (p^A\mu_C^A + p^{CP}\mu_C^{CP} + p^N\mu_C^N)$
$= p^A(\mu_T^A - \mu_C^A) + p^{CP}(\mu_T^{CP} - \mu_C^{CP}) + p^N(\mu_T^N - \mu_C^N)$

If we assume that the exclusion restriction—that is assignment only affects outcomes through attendance—holds for never-takers who do not attend CMO schools, but not for always-takers who may receive different treatment at CMO schools depending on assignment, then $\mu_T^N - \mu_C^N = 0$, so

(1') $ITT^{EXP} = p^A(\mu_T^A - \mu_C^A) + p^{CP}(\mu_T^{CP} - \mu_C^{CP})$

NXP Impacts:

In the NXP, every CMO enrollee from the experimental treatment group is matched with an observationally similar student from a district school.

CMO enrollees from the experimental treatment group have the following mean outcome, which we denote by μ_T^E:

$$\mu_T^E = \frac{p^A}{p^A+p^{CP}}\mu_T^A + \frac{p^{CP}}{p^A+p^{CP}}\mu_T^{CP},$$

which is just a weighted average of the always-taker mean and the complier mean in the experimental treatment group, with weights equal to their proportional representation among CMO enrollees in the experimental treatment group. μ_T^E is observed but μ_T^A: and μ_T^{CP}: are not observed because always-takers and compliers cannot be distinguished.

Consider the matched comparison sample in the NXP analysis. Let μ_D denote the mean outcome in the matched comparison sample, with the subscript "D" connoting "district".

The TOT impact in the NXP is:

(2) $TOT^{QED} = \mu_T^E - \mu_D = \left(\frac{p^A}{p^A+p^{CP}}\mu_T^A + \frac{p^{CP}}{p^A+p^{CP}}\mu_T^{CP}\right) - \mu_D.$

What is μ_D? The matched comparison sample consists of: (1) district students who are matched to treatment group always-takers, and (2) district students who are matched to treatment group compliers. Since we cannot distinguish always-takers and compliers in the treatment group, we also cannot distinguish the two subgroups of the matched comparison sample. Nevertheless, μ_D can be expressed as a weighted average of the unobservable mean outcomes of the two aforementioned subgroups:

(3) $\mu_D = \left(\frac{p^A}{p^A+p^{CP}}\mu_D^{MA} + \frac{p^{CP}}{p^A+p^{CP}}\mu_D^{MC}\right)$

Where μ_D^{MA} is the mean outcome for district students matched to treatment group always-takers, and μ_D^{MC} is mean outcome for district students matched to treatment group compliers. Substituting (3) into (2) and rearranging terms gives:

(4) $TOT^{QED} \frac{p^A}{p^A+p^{CP}}(\mu_T^A - \mu_D^{MA}) + \frac{p^{CP}}{p^A+p^{CP}}(\mu_T^{CP} - \mu_D^{MCP}).$

If CMO schools have an impact, then $\mu_D^{MA} \neq \mu_C^A$ because the outcome of the CMO enrollees in the experimental control group will reflect that impact but the matched district students' outcome will not. However, if the district students who are matched to treatment group compliers are an excellent proxy for the experimental control group compliers (more reasonable since experimental control group compliers do not attend CMO schools), then $\mu_D^{MCP} = \mu_C^{CP}$, which means:

(4') $TOT^{QED} \frac{p^A}{p^A+p^{CP}}(\mu_T^A - \mu_D^{MA}) + \frac{p^{CP}}{p^A+p^{CP}}(\mu_T^{CP} - \mu_C^{CP}).$

Since the NXP goal is to get as close to the experimental ITT as possible, multiply (4') by (p^A+p^{CP}) to eliminate the denominators on the right hand side of equation (4'). Moreover, as noted, $p_T^E = p^A + p^{CP}$. Therefore,

(5) $p_T^E \times TOT^{QED} = p^A(\mu_T^A - \mu_D^{MA}) + p^{CP}(\mu_T^{CP} - \mu_C^{CP}).$

This is very close to the experimental ITT expression in equation (1'). The only difference is that equation (5) has $(\mu_T^A - \mu_D^{MA})$ instead of $(\mu_T^A - \mu_C^A)$. As noted, this is a substantial difference because the mean outcome of always-takers in the experimental control group, μ_C^A, reflects the contribution of CMOs while the mean outcome of district students who are matched to treatment group always-takers, μ_D^{MA}, does not reflect any contribution of CMOs since these students did not enroll in CMOs.

However, let us conduct another matching exercise in which CMO enrollees in the experimental control group (that is, the control group always-takers) are matched with district students who are similar on observed characteristics. Let TOT^{EC} denote the difference in outcomes between CMO enrollees in the experimental control group and their matched counterparts in the district. In the best-case scenario, the district students who are matched to CMO enrollees in the experimental control group have the *same mean outcome* as district students who are matched to always-takers in the experimental treatment group. This assumption is reasonable because both groups of district students are being matched to always-takers from the lottery and randomization means the always-takers should be equivalent at baseline. Under this assumption:

(6) $TOT^{EC} = \mu_C^A - \mu_D^{MA}$, leads to:

(7) $p_C^E \times TOT^{EC} = p_C^E(\mu_C^A - \mu_D^{MA}) = p^A(\mu_C^A - \mu_D^{MA})$.

Finally, subtracting equation (7) from equation (5), gives:

(8) $p_T^E \times TOT^{QED} - p_C^E \times TOT^{EC} = p^A(\mu_T^A - \mu_D^{MA}) + p^{CP}(\mu_T^{CP} - \mu_C^{CP}) - p^A(\mu_C^A - \mu_D^{MA})$
$= p^A(\mu_T^A - \mu_C^A) + p^{CP}(\mu_T^{CP} - \mu_C^{CP})$
$= ITT^{EXP}$

Therefore, $p_T^E \times (TOT^{QED}) - p_C^E \times TOT^{EC}$ should be similar to the experimental ITT impacts. Note that the better the matches created by the NXP matching process, the more likely the equality in equation 8.

APPENDIX D

METHODOLOGY FOR ESTIMATING CMO AND SCHOOL-LEVEL IMPACTS ON ACHIEVEMENT IN MIDDLE-SCHOOLS

To obtain a matched comparison group, we used a propensity score matching (PSM) procedure (Rosenbaum and Rubin 1983). While we used a uniform approach across CMOs to estimate impacts, our main impact analyses were conducted separately for each CMO. This appendix describes our approach to estimating propensity scores, matching treatment and comparison students, estimating CMO-specific impacts and school-level, and conducting sensitivity analyses.

1. Estimating Propensity Score

The first step for the PSM approach is to estimate a propensity score for each student in the sample. The potential comparison students, however, are more diverse in some of their baseline characteristics than the treatment students. To improve the fit of the propensity score model, we conducted a common support check that excluded from the sample any potential comparison students who were outside of the range of the treatment students' baseline characteristics prior to selecting the propensity model. For example, if baseline math test scores for treatment students in a given CMO were between -1.7 and 1.9, only potential comparison students with test scores within this range were kept. Similarly, if all treatment students in a given CMO were classified as non-English language learner (ELL) students, the potential comparison group sample was restricted to non-ELL students as well. Covariates included in the common support check were baseline math and reading test scores, indicators for missing baseline test scores, sex, race/ethnicity, free- or reduced-priced lunch (FRPL) status, disability (IEP) status, English language learner (ELL) status, whether a student attended a charter school in the baseline year, baseline grade, cohort, and district.

A propensity score model was then developed using an automated stepwise model selection procedure for the logistic regression in SAS 9.1. Because baseline math and reading test scores are some of the strongest predictors of later test scores (Glazerman, Levy, and Myers 2003; Cook, Shadish, and Wong 2008), we specified that the model-building procedure begin with the model containing baseline math and reading test scores (measured in the grade prior to the CMO middle school's entry grade) and corresponding missing test score indicators. At each subsequent step, the stepwise procedure then either adds or subtracts a term from a specified set of potential covariates to optimize model fit to the data. The list of potential covariates included sex, race/ethnicity, FRPL status, IEP status, ELL status, whether a student attended a charter school in the baseline year, baseline grade, cohort, district,[47] any two-way interactions of these covariates (including those with baseline test scores and corresponding missing test score indicators), and two interactions of test scores with themselves (i.e., quadratic terms).

---

[47] We did not use pre-baseline (two grades prior to CMO entry) test scores as potential covariates because they were missing for a large proportion of students in our sample. However, we used them as control variables in the impact model.

Propensity model selection was often an iterative process. The Hosmer and Lemeshow (H-L) Goodness-of-Fit test[48] was used, among other indicators, to determine whether the model selected by the model selection procedure fit the data well and whether to proceed with selecting a matched sample. If there were indications of bad model fit (e.g., small H-L p-value or large standard errors for some of the parameters included in the model), we diagnosed and corrected the problems. Most of the issues were resolved by detecting and collapsing sparse cells. For example, if the treatment group had very few Asian students resulting in high standard errors for the Asian parameter in the propensity model, we collapsed Asian with another race/ethnicity category, such as White/Other. We then re-ran the model selection procedure with the collapsed categories and again checked the model fit. We used the final propensity model to estimate the propensity scores for each treatment and comparison student in our sample for that CMO.

2. Selecting a Matched Comparison Group

After estimating the propensity scores, the next step involved selecting a matched comparison group whose estimated propensity scores were similar to those of treatment group students. To improve statistical precision, we selected multiple matches for each treatment student. To ensure the quality of the matches and reduce bias, we matched with replacement (allowing each comparison student to match more than one CMO student) and implemented caliper matching, whereby a given treatment student is matched to all comparison students with estimated propensity scores within a specified range (caliper), rather than merely selecting a specified number of nearest neighbors. We specified five calipers, ranging from 10^{-5} to 10^{-3}. Starting with the smallest caliper, we checked for matches. If a treatment student had between 2 and 30 potential matches, all of these comparison students were identified as the matches for the given treatment student. If the number of potential matches exceeded 30, we identified the 30 comparison students with the closest propensity score (that is, the best-matched students) as the matches to this treatment student. If we did not find at least two matches, we increased the caliper to the next level and tried again. The matching procedure was implemented separately for each grade, cohort, and district combination (we refer to these combinations as the matching strata).

To ensure that we included as many students as possible and to increase the likelihood of a successful match, a given treatment (or comparison) student did not have to have test scores for all outcomes of interest to be included in the sample to be matched. However, this could have resulted in unbalanced treatment and comparison group samples for some outcomes due to differences in the samples of students across outcomes. To ensure that our treatment and comparison samples were balanced for all outcomes of interest, we created outcome-specific analysis weights. In particular, for a given outcome, each treatment student with a valid outcome of interest and at least one matched comparison student was assigned a weight of 1. The matched comparison students (with a valid outcome of interest) were assigned the analysis weight for the treatment students to whom they were matched. In other words, a treatment student and his or her matched comparison(s) had to have the same weighted representation within the treatment and comparison samples, respectively. A given treatment student could potentially have many matched comparison

[48] The H-L Goodness-of-Fit statistic is constructed by first dividing the observations into deciles based on their predicted probabilities and then calculating the chi-square statistic testing whether the distributions of predicted and actual frequencies across deciles are the same. Smaller *p*-values indicate worse model fits.

students; when this happens, the treatment student's weight is divided into even-weight shares among the matched comparison students, so that collectively the matched comparison students have the same weight as the treatment student. Since all matches are done with replacement, a given comparison student could also be matched to many treatment students. In these instances, the comparison student's analysis weight equals the sum of the weights (or weight shares) for all treatment students to whom he or she was matched. Any treatment student with a missing value for the outcome or with no matched comparison students with an observed outcome was assigned a weight of zero. Similarly, any comparison student who was matched to a treatment student with a missing outcome was assigned a zero weight as well. Since we matched students separately within each matching stratum, this procedure for calculating weights also ensured that treatment and comparison students were weighted proportionally within a given grade, cohort, and district combination. Finally, we rescaled the weights to add up to the total number of effective students in the analysis sample (i.e. the number of treatment and matched comparison students with a valid outcome and a positive weight) for this outcome.

Using this procedure, we achieved equivalence between treatment and comparison groups on baseline test scores and most of the observed key covariates (see Appendix E for more detail). This level of baseline equivalence meets the internal validity standards of the U.S. Department of Education's What Works Clearinghouse. Furthermore, for each CMO, we were able to match between 64 percent and 100 percent of CMO enrollees in the primary intake grade who had a valid outcome, with a match rate of at least 90 percent for most of the test scores that we rely on as key outcomes, enhancing the external validity of the analysis.[49]

3. Estimating CMO-Specific Impacts

Following the creation of matched samples and analyses weights, we employed a regression model to improve statistical precision and to control for any remaining differences in baseline characteristics. Our CMO-specific impact regression model is:

$$y_{is} = \propto + X_i\beta + \delta_s T_i + V_i\theta + \epsilon_i,$$

where y_{is} is the subject *s* test score outcome one, two or three years after initial school enrollment for student *i*; α is the intercept; X_i is a vector of achievement (pre-baseline and baseline test scores in reading and math, and corresponding missing test score indicators) and demographic (race/ethnicity, FRPL status, IEP status, and ELL status[50]) characteristics; T_i is a binary variable for treatment status, indicating whether student *i* enrolled in one of the study CMO schools; V_i is a vector of indicators identifying which grade, cohort and district the student belonged to; ε_i is a random error term that reflects the influence of unobserved factors on the outcome; and α, β, δ_s, and θ are vectors of parameters or parameters to be estimated. The estimated coefficient on treatment status, δ_s, represents the one, two or three-year impact in subject *s* of attending a study CMO school. We used robust standard errors that adjust for clustering of students within schools and also implemented the analyses weights described in the preceding section.

[49] Match rates were 90 percent or better for 17 of the 22 CMOs on two-year reading and math outcomes, for 8 of the 11 CMOs on the three-year science outcome, and for 6 of the 9 CMOs on the three-year social studies outcome.

[50] Not all student variables are available in all jurisdictions.

4. Sensitivity Analyses for CMO-specific Impacts

As mentioned before, if all covariates that are related to a student's probability of enrolling in a CMO school and that are also related to the outcome of interest are observed and appropriately accounted for, then the PSM approach could, in theory, result in an unbiased estimator of the impact of attending a CMO school. As described previously, since a large proportion of students were missing pre-baseline test scores, we did not use pre-baseline scores in the propensity model (though we did use them in the impact regression). However, pre-baseline test scores could be related to either a student's probability of enrolling in a CMO school or the outcome of interest in a way that is not fully accounted for by baseline test scores or any of the other matching covariates selected by the stepwise model selection procedure. Furthermore, students who are residing in areas far away from CMO schools may be different from students who reside closer to CMO schools. For example, these students could have attended better (or worse) elementary schools or they could be residing in an area with more (or less) resources (e.g., after-hours educational programs or better libraries), which would make them better (or less) prepared to succeed in middle school. If so, not accounting for the area of residence or the type of school that a student attended prior to enrolling in a CMO school (or a comparison middle school) could introduce bias in the estimates.

Therefore, we explored these possibilities by conducting sensitivity analyses in four CMOs. Under this alternative approach, we re-matched students following the procedure outlined above with two modifications. First, we restricted our potential comparison group to students who at baseline attended one of the schools that the treatment students attended before enrolling in the CMO schools (i.e. the feeder schools). Second, along with baseline test scores, we included pre-baseline math and reading test scores and corresponding missing test score indicators in the stepwise propensity model selection procedure as required covariates. As expected, due to these restrictions the potential pool of comparison students was much smaller and we were able to successfully match fewer treatment students. In particular, for one of the CMOs, we were only able to match 123 (or 30%) of the original 409 successfully matched treatment students. Even so, the CMO-specific impacts estimates were generally quite similar to our benchmark estimates: on average the impact estimates changed by -0.015 and the change never exceeded 0.05 (Table D.1). We concluded that any bias from omitting these variables from the match is negligible, and that our primary method is preferred because it allows us to include many more students.

Table D.1. Main CMO-Specific Impact Estimates Compared to Results from the Sensitivity Analyses

	Main Results					Results of the Sensitivity Analyses				
CMO	N_T	N_C	Impact	SE	p-value	N_T	N_C	Impact	SE	p-value
2-Year Math										
D	837/142	10,177/963	-0.12	0.05	0.01	581/104	5,723/407	-0.11	0.04	<0.01
E	269/27	5,331/223	0.05	0.11	0.66	211/26	1,820/103	0.10	0.10	0.30
J	628/20	11,191/221	0.09	0.03	<0.01	566/20	7,379/97	0.08	0.03	<0.01
L	409/8	2,801/309	-0.02	0.02	0.35	123/4	285/52	0.01	0.08	0.89
2-Year Reading										
D	853/145	10,374/970	-0.10	0.04	0.01	595/106	5,842/414	-0.10	0.04	<0.01
E	269/27	5,327/223	-0.13	0.04	<0.01	212/26	1,837/103	-0.12	0.03	<0.01
J	627/20	11,164/221	0.18	0.02	<0.01	565/20	7,388/98	0.19	0.02	<0.01
L	409/8	2,830/310	-0.10	0.02	<0.01	123/4	289/52	-0.08	0.05	0.10

Notes: (1) The sensitivity analyses restricted the potential comparison group to students attending feeder schools and controlled for pre-baselines test scores in the propensity model. (2) N_T=sample sizes for treatment group. N_C=sample sizes for matched comparison group. The first number indicates the number of students, while the second number indicates the number of schools these students attended in year 2. (3) Standard errors are adjusted for clustering at the school level.

5. Estimating School-Specific Impacts and the Intra-cluster Correlation Coefficient

Finally, to examine whether impacts were consistent across the middle schools that a given CMO operated, we estimated two-year school-specific impacts in each CMO. To investigate this, we used the matched sample from our main analyses and compared the achievement of students who attended a given CMO middle school to the achievement of the overall matched comparison group for this CMO using the following regression model:

$$y_{is} = \propto^* + X_i\beta^* + \sum_k \gamma_{sk} T_{ik} + V_i\theta^* + \epsilon_i,$$

where y_{is} is the subject *s* test score outcome two years after initial school enrollment for student *i*; α is the intercept; X_i is a vector of achievement and demographic characteristics (discussed in section 3 of this appendix); T_{ik} are *k* binary variables for treatment status, indicating whether student *i* enrolled in study CMO middle school *k* or not; V_i is a vector of indicators identifying which grade, cohort, and district the student belonged to; ε_i is a random error term that reflects the influence of unobserved factors on the outcome; and α^*, β^*, γ_{sk}, and θ^* are parameters or vectors of parameters to be estimated. The estimated coefficients of the treatment status, γ_{sk}, represents the two-year impact in subject *s* of attending a CMO school *k*. We used the analyses weights described in the section 2 of this appendix.

To quantify how consistent the impacts within the CMOs are, we estimated the intra-cluster correlation coefficients (ICCs) for year-2 math and reading impacts. The ICC examines the amount of variability in impacts *between* CMOs as compared to the *total* variation in school-specific impacts. In general, ICC ranges between zero and one; a value close to one indicates that the school-specific impacts *within* a CMO are relatively similar with most of the variation in school impacts being due to differences *between* CMOs; a value close to zero indicates that the school-specific impacts within CMOs vary widely relative to the variation in impacts between CMOs. To calculate ICC, we used the following formula:

$$ICC_s = \frac{var(\hat{\delta}_{ps})}{var(\hat{\gamma}_{qs})} = \frac{(1/22)\sum_p(\hat{\delta}_{ps} - \bar{\hat{\delta}}_s)^2}{(1/67)\sum_q(\hat{\gamma}_{qs} - \bar{\hat{\gamma}}_s)^2},$$

where $\hat{\delta}_{ps}$ is the estimated impact of CMO *p* on subject *s*, $\hat{\gamma}_{qs}$ is the estimated impact of CMO middle-school *q* on subject *s*, and $\bar{\hat{\delta}}_s$ and $\bar{\hat{\gamma}}_s$ are the averages of CMO-specific and school-specific impacts on subject *s*, respectively.

APPENDIX E

BASELINE EQUIVALENCE

As presented in Chapter Four, the study's impact analyses used propensity score matching to identify groups of comparison students that are similar to CMO students. Tables E.1 (for second year math impacts) and E.2 (for second year reading impacts) present the average baseline characteristics of students in each CMO and its associated comparison group. The tables also report the percentage of eligible CMO students who were matched successfully—any CMO students who could not be matched were not included in the impact sample.

For the study's second-year math impacts, the matching procedure identified a group of comparison students with very similar characteristics to CMO students. Out of the 22 CMOs, baseline math scores were not significantly different in 20 cases and baseline reading scores were not significantly different in 21 cases (Table E.1). For the two CMOs that did have significant differences in baseline scores, the difference in mean z-scores was below .15 in both subjects. None of the matched comparison groups had significant differences in gender, free and reduced price lunch (FRPL) status, or limited English proficiency (LEP) status. Out of 21 CMOs with data on individualized education plans (IEP), there were no significant differences in 20 cases. For student racial categories (not shown), there were no significant differences in 21 out of 22 cases. Match rates for all 22 CMOs were above 70 percent, and for 20 CMOs the study matched more than 80 percent of all eligible CMO students.

The baseline equivalence results are very similar for the second-year reading impacts (Table E.2). Baseline scores in the matched comparison group were statistically indistinguishable from the CMO group in 20 cases for math and 21 cases for reading; for the two CMOs with significant differences, the difference in mean baseline z-scores was less than .15 in both subjects. None of the matched comparison groups showed significant differences in gender, FRPL status, or LEP status. For IEP status, 20 of the 21 CMOs with data did not have a significant difference compared to the matched student group. For student racial categories (not shown), there were no significant differences in 21 out of 22 cases. Also, the match rates for students at the 22 CMOs were above 70 percent in all cases, and in 20 cases the study matched more than 80 percent of all eligible CMO students.

Tables E.3 and E.4 present baseline equivalence figures for the samples used to estimate high school graduation and postsecondary enrollment impacts. For both outcomes, the CMO group is very similar to the matched comparison group and the proportion of matched CMO high school students is high. For high school graduation, which included 6 high school CMOs, there were no significant differences between the CMO group and the matched comparison group for baseline test scores, gender, FRPL status, LEP status, or IEP status. Match rates were also above 80 percent in all cases (and above 90 percent for 5 of the 6 CMOs). For the 4 CMOs with data on postsecondary enrollment, we also had no significant differences for baseline test scores, gender, FRPL status, LEP status, or IEP status, and match rates were above 90 percent in all cases.

The baseline equivalence results and match rates associated with the other outcome years and test subjects presented in this report were similar to the figures shown below. Full baseline equivalence and match-rate tables for these other outcomes are available upon request to the study authors.

Table E.1. Baseline Equivalence of Student Characteristics, Second Year Math Sample

	Math		Reading		Female		FRPL Status		IEP Status		LEP Status		Match
CMO	CMO	Comp.	CMO	Comp.	CMO	Comp.	CMO	Comp.	CMO	Comp.	CMO	Comp.	Rate %
N	-0.33	-0.27	-0.50	-0.45	0.52	0.48	N/A	N/A	0.03	0.02	0.65	0.66	72
C	0.00	-0.01	0.01	-0.01	0.47	0.47	1.00	1.00	0.10	0.09	0.00	0.00	78
H	0.44	0.36	0.52	0.46	0.52	0.47	N/A	N/A	0.09	0.08	0.10	0.11	81
E	0.43	0.46	0.34	0.34	0.40	0.45	0.37	0.37	0.02	0.02	N/A	N/A	81
D	-0.43	-0.37	-0.25	-0.22	0.50	0.53	0.83	0.83	0.03	0.02	N/A	N/A	88
Q	-0.24	-0.22	-0.17	-0.17	0.46	0.50	0.55	0.57	0.13	0.10	N/A	N/A	90
T	-0.03	0.01	-0.14	-0.09	0.50	0.51	0.91	0.90	0.10	0.08	0.37	0.38	91
L	0.11	0.23*	0.26	0.38*	0.41	0.44	N/A	N/A	0.08	0.09	N/A	N/A	93
P	-0.02	0.01	0.05	0.05	0.42	0.41	0.81	0.81	0.08	0.08	N/A	N/A	96
M	-0.02	0.00	0.02	0.03	0.49	0.48	N/A	N/A	0.11	0.09*	0.09	0.09	96
F	-0.09	-0.08	0.01	0.03	0.50	0.50	N/A	N/A	0.11	0.09	0.03	0.03	98
K	0.17	0.14	0.00	-0.03	0.53	0.53	N/A	N/A	0.19	0.20	0.42	0.40	98
J	0.29	0.30	0.26	0.25	0.47	0.48	0.82	0.81	0.03	0.03	N/A	N/A	99
G	-0.11	-0.14	0.09	0.08	0.45	0.45	0.66	0.67	0.07	0.06	0.00	0.00	99
O	-0.05	-0.06	0.07	0.06	0.41	0.42	N/A	N/A	0.15	0.16	0.02	0.03	99
U	-0.11	-0.06	-0.18	-0.12	0.49	0.50	0.90	0.91	0.06	0.06	0.30	0.29	>99
A	0.86	0.89	0.87	0.93	0.42	0.43	N/A	N/A	0.04	0.05	N/A	N/A	>99
R	0.29	0.33	0.33	0.33	0.43	0.43	0.81	0.83	N/A	N/A	N/A	N/A	>99
I	-0.09	-0.02*	-0.04	0.00	0.54	0.55	0.88	0.87	0.02	0.02	N/A	N/A	>99
B	-0.01	-0.02	-0.01	-0.01	0.48	0.50	0.87	0.88	0.14	0.13	0.33	0.33	>99
V	-0.05	-0.01	0.01	0.08	0.54	0.49	0.83	0.85	0.14	0.14	0.06	0.07	>99
S	0.24	0.25	0.21	0.23	0.46	0.47	0.91	0.90	0.01	0.01	N/A	N/A	>99

* Difference between the CMO and comparison student group is statistically significant at the five percent level.

Table E.2. Baseline Equivalence of Student Characteristics, Second Year Reading Sample

	Math		Reading		Female		FRPL Status		IEP Status		LEP Status		Match
CMO	CMO	Comp.	CMO	Comp.	CMO	Comp.	CMO	Comp.	CMO	Comp.	CMO	Comp.	Rate %
N	-0.34	-0.24	-0.50	-0.44	0.52	0.47	N/A	N/A	0.03	0.02	0.65	0.66	72
C	0.00	-0.01	0.01	-0.01	0.47	0.47	1.00	1.00	0.10	0.09	0.00	0.00	78
E	0.43	0.46	0.35	0.35	0.40	0.45	0.37	0.37	0.01	0.02	N/A	N/A	81
H	0.45	0.43	0.53	0.50	0.51	0.49	N/A	N/A	0.09	0.08	0.10	0.10	82
D	-0.45	-0.41	-0.27	-0.26	0.50	0.53	0.83	0.83	0.03	0.02	N/A	N/A	88
Q	-0.23	-0.21	-0.17	-0.16	0.46	0.50	0.55	0.57	0.13	0.10	N/A	N/A	90
T	-0.02	0.02	-0.12	-0.08	0.50	0.51	0.91	0.89	0.10	0.08	0.37	0.38	91
L	0.11	0.22*	0.26	0.38*	0.42	0.44	N/A	N/A	0.08	0.09	N/A	N/A	93
P	-0.01	0.01	0.04	0.05	0.42	0.41	0.81	0.81	0.08	0.08	N/A	N/A	96
M	-0.02	0.00	0.02	0.03	0.49	0.48	N/A	N/A	0.11	0.09*	0.09	0.09	96
F	-0.10	-0.08	0.01	0.03	0.50	0.50	N/A	N/A	0.11	0.09	0.03	0.03	98
K	0.17	0.14	0.00	-0.03	0.53	0.53	N/A	N/A	0.19	0.20	0.42	0.40	98
J	0.30	0.30	0.26	0.25	0.47	0.48	0.82	0.81	0.03	0.03	N/A	N/A	99
G	-0.11	-0.11	0.09	0.11	0.46	0.45	0.66	0.67	0.08	0.06	0.00	0.00	99
O	-0.06	-0.06	0.07	0.06	0.42	0.42	N/A	N/A	0.15	0.16	0.02	0.03	99
U	0.04	0.07	-0.08	-0.03	0.52	0.50	0.91	0.91	0.05	0.05	0.28	0.28	>99
A	0.86	0.89	0.87	0.93	0.42	0.43	N/A	N/A	0.04	0.05	N/A	N/A	>99
R	0.29	0.33	0.33	0.33	0.43	0.44	0.81	0.83	N/A	N/A	N/A	N/A	>99
I	-0.09	-0.03*	-0.05	-0.02	0.54	0.55	0.88	0.87	0.03	0.02	N/A	N/A	>99
B	-0.01	0.02	0.00	0.03	0.49	0.50	0.87	0.88	0.14	0.12	0.33	0.32	>99
V	-0.05	-0.01	0.01	0.07	0.54	0.49	0.83	0.85	0.14	0.14	0.06	0.07	>99
S	0.24	0.25	0.21	0.23	0.46	0.47	0.91	0.90	0.02	0.01	N/A	N/A	>99

*Difference between the CMO and comparison student group is statistically significant at the five percent level

Table E.3. Baseline Equivalence of Student Characteristics, High School Graduation Sample

	Math		Reading		Female		FRPL Status		IEP Status		LEP Status		Match
CMO	CMO	Comp.	CMO	Comp.	CMO	Comp.	CMO	Comp.	CMO	Comp.	CMO	Comp.	Rate %
I	0.62	0.62	0.38	0.43	0.49	0.43	0.95	0.95	N/A	N/A	N/A	N/A	82
II	N/A	N/A	0.58	0.60	0.48	0.44	N/A	N/A	0.04	0.04	0.06	0.06	93
III	-0.22	-0.20	-0.08	-0.08	0.45	0.45	0.79	0.77	N/A	N/A	N/A	N/A	95
IV	-0.10	-0.07	-0.14	-0.11	0.46	0.45	N/A	N/A	0.22	0.20	0.10	0.08	98
V	0.14	0.15	0.17	0.17	0.44	0.45	N/A	N/A	0.20	0.20	0.04	0.04	98
VI	-0.52	-0.49	-0.38	-0.28	0.48	0.48	N/A	N/A	0.19	0.16	N/A	N/A	>99

*Difference between the CMO and comparison student group is statistically significant at the five percent level

Table E.4. Baseline Equivalence of Student Characteristics, Postsecondary Enrollment Sample

	Math		Reading		Female		FRPL Status		IEP Status		LEP Status		Match
CMO	CMO	Comp.	CMO	Comp.	CMO	Comp.	CMO	Comp.	CMO	Comp.	CMO	Comp.	Rate %
II	N/A	N/A	0.58	0.60	0.48	0.44	N/A	N/A	0.04	0.04	0.06	0.06	93
IV	-0.10	-0.07	-0.14	-0.11	0.46	0.45	N/A	N/A	0.22	0.20	0.10	0.08	98
V	0.14	0.15	0.17	0.17	0.44	0.45	N/A	N/A	0.20	0.20	0.04	0.04	98
VI	-0.52	-0.49	-0.38	-0.28	0.48	0.48	N/A	N/A	0.19	0.16	N/A	N/A	>99

*Difference between the CMO and comparison student group is statistically significant at the five percent level

APPENDIX F

METHOD FOR DEALING WITH GRADE REPETITION

Grade repetition—students retained in a given grade for an additional year, generally because they have not accumulated sufficient knowledge to progress to the next grade—complicates the analysis of academic achievement presented in Chapter Four. Retained students no longer take the same test at the same time as the other students in their cohorts. Since most state assessments are not vertically scaled, the scores of students who are retained cannot be directly compared with the scores of others in their original cohort who have progressed to the next grade, essentially resulting in a missing data problem. One approach to handling this missing data problem would be to drop the grade repeaters from the analysis. However, if CMO students repeat grades at different rates than comparison students, this may bias our impact estimates. Another option would be to include the grade repeaters' scores as they are, a year later than the comparison students, although this would ignore the fact that grade repeaters receive two years to learn material taught to the rest of their cohort in a single year.

Instead, following Tuttle et al (2010), we use information from the past performance of grade repeaters to keep them in the analysis. For each grade repeater, in the first year of repetition and subsequent years, we impute a score on the cohort-appropriate (rather than grade-appropriate) assessment that is equal to the student's last standardized score prior to the initial instance of grade repetition. In other words, we assume that each retained student does neither better nor worse (relative to his/her original cohort) than before retention. If the CMO has a positive impact on the achievement of grade repeaters and a higher rate of grade repetition than comparison students, this will generally produce a conservative estimate of the CMO's impact. If the CMO has a negative impact on the achievement of grade repeaters and a lower rate of grade repetition than comparison students, this would cause us to overestimate the CMO's true impact. In either case, the adjustment is arguably conservative, in the sense that our impact estimates will be biased toward zero.

Of the 22 CMOs for which we are estimating middle school achievement impacts, 16 have rates of grade repetition that differ from rates in their local districts by three percentage points or less; we therefore would not expect the assumption about the scores of grade repeaters to make much of a difference for most of the CMOs. Five CMOs have notably higher rates of grade repetition than their districts, and all five of these have positive math and reading two-year impact estimates. The one CMO with a substantially lower rate of grade repetition than its local district has negative math and reading two-year impact estimates. If CMOs' effects on grade repeaters are similar to their effects on non-repeaters, the impact estimates of all six of these CMOs may be slightly biased toward zero as a result of assuming that their students' scores are unchanged after grade repetition. (Alternatively if the CMOs with positive impacts on non-repeaters have negative impacts on repeaters, this method could inflate our estimates of impacts). Appendix Table F.1 shows the rates of grade repetition for students in each of 22 CMOs with middle school grades and the corresponding rates for students in their local districts.

Table F.1. Mean Rates of Grade Repetition in Middle School Grades, by Treatment Status

CMO	CMO Mean Rate of Grade Repetition	District Mean Rate of Grade Repetition	Difference in Rates (CMO – District)
K	.03	.16	-.12**
O	.02	.05	-.03**
P	.03	.04	-.01*
J	.02	.02	-.01
E	.01	.02	-.01
L	.00	.01	.00*
T	.00	.00	.00**
A	.01	.01	.00
H	.01	.01	.00
G	.01	.00	.00
U	.01	.00	.01*
Q	.02	.01	.01
B	.02	.00	.02**
N	.02	.00	.02**
D	.04	.02	.03**
F	.06	.03	.03**
R	.05	.02	.03**
C	.13	.05	.07**
M	.12	.03	.09**
I	.14	.03	.11**
S	.14	.02	.12**
V	.17	.03	.14**

Source: Administrative data.

*Significantly different from zero at the .05 level, two-tailed test

**Significantly different from zero at the .01 level, two-tailed test.

APPENDIX G

METHODOLOGY AND RESULTS FOR CMO IMPACTS ON HIGH SCHOOL ACHIEVEMENT AND ATTAINMENT

To obtain CMO-specific impacts on high school achievement and attainment, we used a propensity score matching procedure (Rosenbaum and Rubin 1983), in which students who entered a target CMO high school in the entry grade were matched to similar students who were in the same grade, cohort, and district but attended a traditional public or an independent charter high school. The procedure for estimating the propensity score, selecting a matched comparison group, and creating outcome-specific weights was the same as the one used for estimating CMO-specific impacts on middle school achievement outcomes and is described in more detail in Appendix D. This appendix describes our approach to estimating the CMO-specific impacts on high school outcomes, the results of these analyses, and the alternative analyses that were limited to students observed to attend a charter middle or high school. All analyses were done separately for each CMO.

1. Estimating CMO-Specific Impacts

Following the creation of matched samples and analysis weights, we employed a regression model to improve statistical precision of a CMO-specific impact and to control for any remaining differences in baseline characteristics of the treatment and the matched comparison groups. As in the analysis of middle school achievement, we used a linear regression model for the high school achievement outcomes. However, the high school attainment outcomes—high school graduation and enrollment in a 2-year or 4-year college—are binary. Therefore, to estimate the impacts on high school attainment we used a logistic regression model. Both sets of analyses used robust standard errors that adjust for clustering of students within schools and analysis weights described in Appendix D.

Our CMO-specific impact regression model for high school achievement outcomes was:

(G.1) $$y_{is} = \propto + X_i\beta + \delta_s T_i + V_i\theta + \epsilon_i,$$

where y_{is} is the subject *s* test score outcome one year after initial school enrollment for student *i*; X_i is a vector of achievement (pre-baseline and baseline test scores in reading and math, and corresponding missing test score indicators) and demographic (race/ethnicity, FRPL status, IEP status, and ELL status[51]) characteristics; T_i is a binary variable for treatment status, indicating whether student *i* enrolled in one of the study CMO high schools; V_i is a vector of indicators identifying the grade, cohort, and district to which the student belonged; ε_i is a random error term that reflects the influence of unobserved factors on the outcome; and α, β, δ_s, and θ are parameters or vectors of parameters to be estimated. The estimated coefficient on treatment status, δ_s, represents the one-year impact on subject *s* tests score of attending a study CMO school.

[51] Not all student variables are available in all jurisdictions.

The CMO-specific model we used for high school attainment outcomes was logistic regression:

(G.2) $$log \frac{\Pr(Y_{is}=1)}{1-\Pr(Y_{is}=1)} = \propto + X_i\beta + \delta_s T_i + V_i\theta,$$

where Y_{is} is a binary attainment outcome *s* indicating whether student *i* graduated from high school (or enrolled in college) or not; X_i is a vector of achievement (pre-baseline and baseline test scores in reading and math, and corresponding missing test score indicators) and demographic (race/ethnicity, FRPL status, IEP status, and ELL status[52]) characteristics; T_i is a binary variable for treatment status, indicating whether student *i* enrolled in one of the study CMO high schools; V_i is a vector of indicators identifying the grade, cohort, and district to which the student belonged; and α, β, δ_s, and θ are parameters or vectors of parameters to be estimated. The estimated coefficient on treatment status, δ_s, represents the impact on attainment outcome *s* of attending a study CMO high school on a log-odds scale.

To make the impacts on high school attainment more easily interpretable, we converted the CMO-specific impacts into expected impacts in percentage points, i.e. the difference between the chance that an average student in the sample would have graduated from high school (or enrolled in college) if they attended a study CMO high school and the chance of this outcome occurring if the same student did not attend a study CMO high school. To do this, we used the following formula:

(G.3) $$\left(\frac{\exp(\bar{\eta}_s+\hat{\delta}_s)}{1+\exp(\bar{\eta}_s+\hat{\delta}_s)} - \frac{\exp(\bar{\eta}_s)}{1+\exp(\bar{\eta}_s)}\right) * 100\%,$$

where $\bar{\eta}_s = \hat{\propto} + \bar{X}\hat{\beta} + \bar{V}\hat{\theta}$; $\hat{\propto}$, $\hat{\beta}$, $\hat{\delta}_s$, and $\hat{\theta}$ are the estimates from the logistic regression model (Equation G.2) above; and $\bar{X}$ and $\bar{V}$ are the weighted averages of the vector of students' characteristics and vector of indicators indentifying students' grade, cohort, and district, respectively.

2. CMO-Specific Impacts on High School Achievement and Attainment

For the 6 CMOs with available data on attainment outcomes, we also explored whether it is feasible to estimate impacts on test scores (i.e., achievement) during high school. Analyses of high school achievement effects pose several challenges related to the limited number of standardized tests during high school,[53] high rates of grade repetition, and (in many places) tests that are course-specific rather than grade-specific. Only 3 CMOs with attainment data were located in jurisdictions that offered grade-specific exams during high school in at least one subject. Due to high rates of grade repetition, we further restricted the achievement analysis to tests that can be observed in the first year after enrollment, in grade 9. Table G.1 summarizes the sample sizes for the high school achievement analysis, which includes 3 CMOs with results in reading and 2 CMOs with results in math.

---

[52] Not all student variables are available in all jurisdictions.

[53] No Child Left Behind (NCLB) Act of 2001 requires that students academic performance in math and reading is measured annually in grades 3 through 8 and at least once in high school using standardized tests.

Table G.1. Sample Sizes for High School Achievement Analysis After 1 Year of Treatment

Measure	Math	Reading
Number of CMOs	2	3
Number of CMO students	468	984
Number of comparison students	6,649	10,153

Source: State, district, and CMO school records.

Table G.2 presents our CMO-specific impact estimates for reading, math, high school graduation, and enrollment in a two-year or four-year college. The impacts on graduation and college enrollment (included in the table) are described in Chapter IV. This appendix includes test-score effects of the three CMOs for which test-score effects could be estimated. For the three CMOs with results in reading, we find that one CMO has a significant negative effect, another CMO has a significant positive impact, and in a third case there was no significant effect. In math, where we had sufficient data to estimate impacts for two CMOs, we found a significant positive effect in one CMO and an insignificant effect in the second CMO.

These achievement impacts do not perfectly correspond to attainment impacts. In one case (CMO V), the CMO had strongly positive impacts on both reading and math but an insignificant impact on high school graduation. In another case (CMO II), a CMO had strongly positive impacts on both graduation and college entry, despite having an insignificant impact on reading scores. Of the three CMOs with data, there was only one case (CMO VI), where the CMO's achievement impacts appear to match attainment effects. This CMO's negative reading impact and negative (but insignificant) math impact correspond well to the CMO's negative impact on high school graduation. These initial findings suggest that achievement effects alone may not provide a reliable proxy for charter school impacts on longer-term outcomes, such as graduation or college enrollment.

While there appears to be little correspondence between achievement effects and attainment effects, high school graduation impacts are often similar to college enrollment impacts for the CMOs in our sample. The sign and significance of the high school graduation impact matches the sign and significance of the college entry impact for 3 of the 4 CMOs with data on both outcomes. In one case (CMO III), however, the CMO had a positive impact on high school graduation but an insignificant impact (still in the positive direction) on college enrollment.

Table G.2. High School Achievement and Attainment Impacts, by CMO

	Achievement Effect Size		Attainment Impact (in Percentage Points)	
CMO	1-Year Reading	1-Year Math	High School Graduation	College Enrollment
CMO #1	N/A	N/A	23%** (N=977)	23%** (N=977)
CM0 #2	0.06 (N=508, SE=0.04)	N/A	17%** (N=532)	21%** (N=532)
CMO #3	N/A	N/A	12%* (N=189)	3% (N=189)
CMO #4	N/A	N/A	8% (N=452)	4% (N=452)
CM0 #5	0.19** (N=169, SE=0.05)	0.13** (N=169, SE=0.03)	3% (N=182)	N/A
CMO #6	-0.08* (N=307, SE=0.03)	-0.06 (N=299, SE=0.09)	-22%** (N=327)	N/A
Average Impact	**0.06 (SE=0.08)**	**0.03 (SE=0.10)**	**7% (SE=6)**	**13% (SE=5)**

Source: State, District and CMO School Records.

Notes: To account for sample differences across outcomes and our propensity score matching approach which allowed for multiple matches, student-level observations were weighted such that, collectively, the treatment group has the same weight as the matched comparison group. Since we matched students separately within each matching strata (district/state-cohort-grade), treatment and matched comparison students were also weighted proportionally within each strata. The sample sizes (N) in the table report the number of unique treatment students for each CMO. Additionally, we calculated robust standard errors (SE) that were clustered at the school level to determine statistical significance. Refer to Appendix G for more details on the impact estimation procedure. Achievement effects are reported in standard deviation units, whereas attainment impacts are reported as percentage-point effects, derived from each CMO's estimated logistic regression model. The average impacts and corresponding SEs were calculated using by regressing CMO-specific impacts on an intercept only.

*Significantly different from zero at the 0.05 level, two-tailed test.
**Significantly different from zero at the 0.01 level, two-tailed test.

3. Alternative Analyses of CMO Impacts on High School Attainment

Students who have ever enrolled in a charter school may be more similar to one another than to students who have never chosen to enroll in a charter school. We, therefore, conducted a sensitivity analysis of attainment impacts using an alternate method that addresses one potential source of selection bias by comparing CMO students only to other students who chose to enroll in charter schools during middle or high school grades (following the example of Booker et al. 2011). This method implicitly controls for unobserved student and family characteristics that are associated with choosing to enroll in a charter school. It dramatically shrinks the pool of potential students in the comparison group, however, which means it cannot be feasibly used in conjunction with matching on the other variables used in our primary impact analysis.

As with our main analyses, we first conducted a common support check that excluded from the sample any potential comparison students who were outside of the range of the treatment students' baseline characteristics prior to selecting the propensity model. Covariates included in the common support check were baseline math and reading test scores, indicators for missing baseline test scores, sex, race/ethnicity, FRPL status, IEP status, ELL status, whether a student attended a charter school in the baseline year, baseline grade, cohort, and district (See Appendix D for more detail). We then used the logistic regression model (Equation G.2) to control for any remaining differences between treatment and comparison students.

While the overall average impacts on high school graduation and college entry are statistically insignificant regardless of which comparison group is used, the results of the alternate analysis differ from the main analyses for specific CMOs (Table G.3). For high school graduation, the main results indicate that three CMOs (CMOs II, I, and III) have positive statistically significant impacts and one CMO (CMO VI) has a negative statistically significant impact. The alternative analyses, in contrast, indicate three positive statistically significant impacts for a different group of CMOs (CMOs V, II, and I) and no negative CMO impacts. Furthermore, the alternative analyses resulted in five out of six CMOs changing impact ranking by 2 or more out of the 6 possible slots.[54] Finally, the range of impacts on graduation decreased (from between -22 and +23 percentage points with the main analyses to -4 and +24 percentage points with the comparison group of charter choosers). The change in the results for college enrollment was not as dramatic. While one CMO (CMO III) with a statistically insignificant impact in the main analyses ended up with a negative statistically significant impact when limited to the charter chooser sample, no CMOs changed impact rankings by more than 1 slot (the two highest-ranked CMOs switched places). However, the range of impacts increased (from between 3 and 23 percentage points with the main analyses to between -12 and +28 percentage points with the alternate analyses).

To conclude, the overall average impacts of CMOs on high school graduation and college entry did not change with the alternate analysis. However, the sensitivity of the CMO-specific impacts to the choice of comparison group (without a strong theoretical justification for preferring one approach over the other), combined with small number of CMOs that had appropriate data to be included in the analyses, suggest caution when interpreting the CMO-specific attainment results.

[54] Since a change in ranking of one slot is likely to be due to random noise and there are only 6 ranking slots possible for high school graduation and only 4 possible for college enrollment impacts, we decided to use a change by 2 or more slots as our criteria of meaningfully significant change in ranking.

Table G.3. Comparing Attainment Impacts from Main and Alternative Analyses, by CMO, in Percentage Points

	High School Graduation		College Enrollment	
CMO	Main Results	Charter Choosers	Main Results	Charter Choosers
CMO #1	23%** (R=1)	13%** (R=3)	23%** (R=1)	14%** (R=2)
CMO #2	17%** (R=2)	24%** (R=1)	21%** (R=2)	28%** (R=1)
CMO #3	12%* (R=3)	-2% (R=5)	3% (R=4)	-12%** (R=4)
CMO #4	8% (R=4)	-4% (R=6)	4%(R=3)	-8% (R=3)
CMO #5	3% (R=5)	18% **(R=2)	N/A	N/A
CMO #6	-22% **(R=6)	4% (R=4)	N/A	N/A
Average Impact	**7% (SE=6)**	**9% (SE=5)**	**13% (SE=5)**	**5% (SE=9)**

Source: State, district and CMO school records.

Notes: The main analyses used propensity score matching approach to account for selection, while the alternative analyses used a sample of students who attended a charter school at any time during middle or high school grades. Refer to Appendix G for more details on the impact estimation procedure. Both sets of analyses utilized logistic regression model to control for any remaining differences in student characteristics, analyses weights, and robust standard errors. The average impacts and their standard errors (SE) were calculated using by regressing CMO-specific impacts on an intercept only. Smaller ranking (R) indicates higher impacts.

*Significantly different from zero at the 0.05 level, two-tailed test.
**Significantly different from zero at the 0.01 level, two-tailed test.

APPENDIX H

IMPACTS ON MIDDLE SCHOOL TEST SCORES BY CMO, YEAR, AND SUBJECT

Table H.1. Math Test-Score Impacts in Middle Schools, by CMO and Number of Years after Enrollment (Impact Units are Standard Deviation Effect Sizes)

CMO	1-Year	2-Year	3-Year
A	-0.06 (N=395)	-0.13** (N=179)	N/A
B	0.16* (N=1,492)	0.36** (N=1,042)	N/A
C	0.47** (N=710)	0.63** (N=500)	0.74** (N=315)
D	-0.12** (N=1,078)	-0.12* (N=837)	-0.15* (N=223)
E	0.03 (N=502)	0.05 (N=269)	-0.13** (N=67)
F	-0.02 (N=1,157)	0.28** (N=826)	0.34** (N=482)
G	0.16** (N=858)	0.31* (N=534)	N/A
H	-0.35** (N=658)	-0.30** (N=499)	N/A
I	-0.14** (N=1,276)	0.12** (N=961)	0.22** (N=637)
J	0.00 (N=916)	0.09** (N=628)	0.11* (N=357)
K	-0.27** (N=517)	-0.21** (N=403)	-0.23** (N=279)
L	0.03 (N=544)	-0.02 (N=409)	-0.09** (N=291)
M	0.35** (N=1,417)	0.50** (N=1,125)	0.48** (N=858)
N	-0.11** (N=323)	-0.27** (N=207)	N/A
O	-0.04** (N=569)	-0.09** (N=422)	-0.11** (N=188)
P	0.12** (N=792)	0.17** (N=746)	N/A
Q	-0.19** (N=489)	-0.26** (N=342)	-0.05 (N=208)
R	-0.13** (N=545)	-0.05 (N=428)	-0.04 (N=371)
S	0.39** (N=2,309)	0.40** (N=1,766)	0.27** (N=1,245)
T	0.58** (N=773)	0.42** (N=519)	N/A
U	0.10** (N=805)	0.02 (N=449)	N/A
V	0.30** (N=481)	0.55** (N=343)	0.72** (N=226)

Table H.1 *(continued)*

CMO	1-Year	2-Year	3-Year
Number of CMOs	**22**	**22**	**14**

Source: State, district and CMO school records.

Notes: To account for sample differences across outcomes and our propensity score matching approach which allowed for multiple matches, student-level observations were weighted such that collectively, the treatment group has the same weight as the matched comparison group. Since we matched students separately within each matching strata (district/state-cohort-grade), treatment and matched comparison students were also weighted proportionally within each strata. The sample sizes in the table report the number of unique treatment students for each CMO. Additionally, we calculated robust standard errors that were clustered at the school level. Refer to Appendix D for more details on the impact estimation procedure.

*Significantly different from zero at the .05 level, two-tailed test.
**Significantly different from zero at the .01 level, two-tailed test.

Table H.2. Reading Test-Score Impacts in Middle Schools, by CMO and Number of Years after Enrollment (Impact Units are Standard Deviation Effect Sizes)

CMO	1-Year	2-Year	3-Year
A	-0.01 (N=395)	-0.09 (N=179)	N/A
B	0.09 (N=1,493)	0.18** (N=1,052)	0.13 (N=763)
C	0.14* (N=710)	0.22** (N=500)	0.31** (N=315)
D	-0.14** (N=1,082)	-0.10* (N=853)	-0.01 (N=226)
E	-0.10** (N=503)	-0.13** (N=269)	-0.11 (N=71)
F	-0.10** (N=1,161)	-0.05 (N=824)	0.00 (N=478)
G	0.17** (N=860)	0.20** (N=548)	0.28** (N=304)
H	-0.14** (N=660)	-0.15** (N=509)	-0.30** (N=376)
I	-0.05 (N=1,289)	0.13** (N=970)	0.13** (N=641)
J	0.07** (N=916)	0.18** (N=627)	0.10** (N=358)
K	-0.20** (N=523)	-0.17** (N=404)	-0.21** (N=280)
L	-0.15** (N=544)	-0.10** (N=409)	-0.25** (N=291)
M	0.07 (N=1,416)	0.22** (N=1,126)	0.31** (N=857)
N	-0.14** (N=325)	-0.22** (N=208)	0.00 (N=131)
O	-0.05 (N=570)	-0.07 (N=423)	-0.03 (N=189)
P	0.12** (N=791)	0.16** (N=748)	N/A
Q	-0.10 (N=489)	-0.13 (N=343)	-0.06 (N=208)
R	-0.07** (N=545)	0.01 (N=426)	0.03 (N=374)
S	0.10** (N=2,318)	0.08** (N=1,770)	0.03 (N=1,254)
T	0.15** (N=772)	0.24** (N=522)	0.30* (N=389)
U	0.03** (N=925)	0.06 (N=621)	0.16* (N=401)
V	0.06 (N=482)	0.23** (N=343)	0.22** (N=225)
Number of CMOs	**22**	**22**	**20**

Table H.2 *(continued)*

Source: State, district and CMO school records.

Notes: To account for sample differences across outcomes and our propensity score matching approach which allowed for multiple matches, student-level observations were weighted such that collectively, the treatment group has the same weight as the matched comparison group. Since we matched students separately within each matching strata (district/state-cohort-grade), treatment and matched comparison students were also weighted proportionally within each strata. The sample sizes in the table report the number of unique treatment students for each CMO. Additionally, we calculated robust standard errors that were clustered at the school level. Refer to Appendix D for more details on the impact estimation procedure.

*Significantly different from zero at the .05 level, two-tailed test.
**Significantly different from zero at the .01 level, two-tailed test.

Table H.3. Science Test-Score Impacts in Middle Schools, by CMO and Number of Years after Enrollment (Impact Units are Standard Deviation Effect Sizes)

CMO	3-Year
A	N/A
B	0.21 (N=744)
C	N/A
D	N/A
E	-0.17** (N=67)
F	N/A
G	0.61** (N=301)
H	-0.49** (N=367)
I	0.03 (N=104)
J	0.31** (N=352)
K	N/A
L	-0.11** (N=188)
M	N/A
N	-0.11 (N=125)
O	N/A
P	N/A
Q	N/A
R	0.06 (N=350)
S	0.32** (N=1,004)
T	N/A
U	0.01 (N=201)
V	N/A
Number of CMOs	**11**

Source: State, district and CMO school records.

Notes: To account for sample differences across outcomes and our propensity score matching approach which allowed for multiple matches, student-level observations were weighted such that collectively, the treatment group has the same weight as the matched comparison group. Since we matched students separately within each matching strata (district/state-cohort-grade), treatment and matched comparison students were also weighted proportionally within each strata. The sample sizes in the table report the number of unique treatment students for each CMO. Additionally, we calculated robust standard errors that were clustered at the school level. Refer to Appendix D for more details on the impact estimation procedure.

*Significantly different from zero at the .05 level, two-tailed test.
**Significantly different from zero at the .01 level, two-tailed test.

Table H.4. Social Studies Test-Score Impacts in Middle Schools, by CMO and Number of Years after Enrollment (Impact Units are in Standard Deviation Effect Sizes)

CMO	3-Year
A	N/A
B	0.12 (N=747)
C	N/A
D	N/A
E	-0.02 (N=68)
F	N/A
G	0.22** (N=307)
H	-0.48** (N=371)
I	N/A
J	0.40** (N=351)
K	N/A
L	N/A
M	N/A
N	-0.03 (N=128)
O	N/A
P	N/A
Q	N/A
R	0.15* (N=350)
S	0.19** (N=1,004)
T	N/A
U	0.31** (N=203)
V	N/A
Number of CMOs	**9**

Source: State, district and CMO school records.

Notes: To account for sample differences across outcomes and our propensity score matching approach which allowed for multiple matches, student-level observations were weighted such that collectively, the treatment group has the same weight as the matched comparison group. Since we matched students separately within each matching strata (district/state-cohort-grade), treatment and matched comparison students were also weighted proportionally within each strata. The sample sizes in the table report the number of unique treatment students for each CMO. Additionally, we calculated robust standard errors that were clustered at the school level. Refer to Appendix D for more details on the impact estimation procedure.

*Significantly different from zero at the .05 level, two-tailed test.
**Significantly different from zero at the .01 level, two-tailed test.

Table H.5. Second Year Math and Reading Impacts in Middle Schools, by CMO School

CMO School	2- Year Math Impacts	2-Year Reading Impacts
A-1	-0.20	-0.23*
	(0.13)	(0.10)
A-2	-0.16**	-0.11
	(0.05)	(0.07)
A-3	-0.05	0.03
	(0.06)	(0.08)
B-1	0.25**	0.05
	(0.03)	(0.03)
B-2	0.29**	0.18**
	(0.03)	(0.03)
B-3	0.40**	-0.01
	(0.04)	(0.04)
B-4	0.46**	0.36**
	(0.04)	(0.03)
B-5	0.68**	0.44**
	(0.06)	(0.04)
C-1	0.53**	0.27**
	(0.05)	(0.06)
C-2	0.61**	0.20**
	(0.05)	(0.05)
C-3	0.73**	0.20**
	(0.05)	(0.05)
D-1	-0.20**	-0.13**
	(0.04)	(0.04)
D-2	-0.16	-0.26**
	(0.11)	(0.09)
D-3	0.03	0.02
	(0.06)	(0.06)
E-1	-0.14	-0.22*
	(0.08)	(0.08)
E-2	-0.10	0.06
	(0.11)	(0.09)
E-3	0.45**	-0.09
	(0.08)	(0.06)
F-1	0.12*	0.13*
	(0.05)	(0.06)
F-2	0.23**	-0.13**
	(0.04)	(0.04)
F-3	0.31**	-0.08**
	(0.03)	(0.03)
F-4	0.36**	0.06
	(0.06)	(0.08)
F-5	0.38**	-0.08
	(0.04)	(0.04)
G-1	-0.03	0.20**
	(0.06)	(0.05)
G-2	0.09	0.18**
	(0.06)	(0.05)
G-3	0.32**	0.11*
	(0.06)	(0.04)
G-4	0.55**	0.24**
	(0.04)	(0.03)

Table H.5 *(continued)*

CMO School	2- Year Math Impacts	2-Year Reading Impacts
H-1	-0.37**	-0.12**
	(0.04)	(0.03)
H-2	-0.24**	-0.18**
	(0.04)	(0.03)
I-1	-0.02	0.02
	(0.05)	(0.06)
I-2	0.00	0.08
	(0.06)	(0.06)
I-3	0.03	0.13*
	(0.05)	(0.06)
I-4	0.04	0.15*
	(0.06)	(0.07)
I-5	0.28**	0.15**
	(0.04)	(0.04)
I-6	0.36**	0.27**
	(0.08)	(0.07)
J-1	-0.04	0.08
	(0.06)	(0.06)
J-2	0.07	0.10
	(0.05)	(0.06)
J-3	0.17**	0.28**
	(0.04)	(0.04)
K-1	-0.20**	-0.16**
	(0.03)	(0.03)
L-1	-0.02	-0.10**
	(0.03)	(0.03)
M-1	0.17**	0.05
	(0.03)	(0.04)
M-2	0.51**	0.19**
	(0.03)	(0.04)
M-3	0.52**	0.15**
	(0.03)	(0.03)
M-4	0.75**	0.49**
	(0.04)	(0.05)
N-1	-0.27**	-0.22**
	(0.04)	(0.04)
O-1	-0.17**	-0.24**
	(0.03)	(0.04)
O-2	-0.03	0.06
	(0.03)	(0.03)
P-1	0.17**	0.16**
	(0.03)	(0.02)
Q-1	-0.46**	-0.19**
	(0.07)	(0.06)
Q-2	-0.38**	-0.19**
	(0.07)	(0.07)
Q-3	-0.24**	-0.25**
	(0.08)	(0.07)
Q-4	-0.01	0.10
	(0.08)	(0.07)
R-1	-0.08*	0.05
	(0.04)	(0.04)
R-2	-0.01	-0.02
	(0.05)	(0.04)

Table H.5 *(continued)*

CMO School	2- Year Math Impacts	2-Year Reading Impacts
S-1	0.28**	0.04
	(0.03)	(0.03)
S-2	0.36**	0.05
	(0.05)	(0.05)
S-3	0.37**	0.06*
	(0.03)	(0.03)
S-4	0.47**	0.09*
	(0.05)	(0.04)
S-5	0.59**	0.16**
	(0.05)	(0.04)
T-1	0.01	-0.03
	(0.25)	(0.14)
T-2	0.02	0.13
	(0.12)	(0.10)
T-3	0.28**	-0.04
	(0.07)	(0.06)
T-4	0.31**	0.03
	(0.06)	(0.05)
T-5	0.67**	0.59**
	(0.06)	(0.07)
U-1	-0.06	0.12**
	(0.04)	(0.02)
U-2	0.11**	-0.04
	(0.04)	(0.04)
V-1	0.54**	0.16**
	(0.06)	(0.06)
V-2	0.56**	0.26**
	(0.04)	(0.06)
Number of CMOs	**22**	**22**
CMO Schools	**67**	**67**

Source: State, district and CMO school records.

Notes: The CMO School column of this table shows the letter-code of the CMO followed by a number-code for each school. Impact estimates compare the CMO treatment group of each school to the relevant CMO's matched comparison group, using robust standard errors clustered at the school level. Sample sizes are not reported in the table, to preserve the anonymity of CMO schools—the sample size of unique CMO treatment students ranged from 16 to 748, and the average sample size was 201 in 2-year reading and 204 in 2-year math. Refer to Appendix D for more details on the impact estimation procedure for CMO schools.

*Significantly different from zero at the .05 level, two-tailed test.
**Significantly different from zero at the .01 level, two-tailed test

APPENDIX I

COMPARING CMO AND INDEPENDENT CHARTER IMPACTS

Many CMOs began as attempts to scale-up educational approaches that appeared effective at individual charter schools. Reporting nationwide CMO impacts raises questions about how those impacts differ from impacts of charter schools not part of CMOs (here labeled *independent charters*). Are more effective charter approaches the ones that are scaled by CMOs? Do CMOs' scale provide them with advantages or disadvantages over independent charters? To provide a comparison, we estimated impacts for CMOs and independent charters in four large school districts. Clearly, any differences between CMO and independent charter impacts in these districts cannot be reliably attributed to CMO management; there are many potential differences between sample CMO schools and independent charter schools that are not inherently due to CMO management (for example, CMO schools may be older than independent charter schools or have different intake grades).[55] Instead, the comparison provides a non-causal reference for CMO impacts. The diversity of CMO impacts suggests that any comparison with independent charters will likely vary, and impacts across four districts indicate that is the case—some CMO had higher impacts than independent charters in the same district and some CMOs had lower impacts.

1. Data and Methods

a. Sample and Data

The comparison reported in this appendix involves CMO and independent charter schools in four districts. These districts were selected for having the largest number of independent charter middle schools from among all the states and districts who provided student records data that indicated whether students attended a charter school. Limiting to jurisdictions with clear and clean charter school indicators ensured that we could reliably identify all charters schools in the district. The four districts include 3 of the largest 20 school districts in the United States and are located in three of the four national census regions.

As with the overall impact analysis, this comparison is limited to middle schools. Each of the four districts included three or more independent charter middle schools.[56] District A had three CMOs with middle schools, Districts B and C had middle schools from two CMOs, and District D had middle schools from one CMO.

b. Methods

Impacts were estimated using OLS regression with statistical controls for baseline achievement and demographic characteristics. The validation exercise (Appendix C) indicated that OLS regression estimated impacts very close to experimental impact estimates. Specifically, the OLS estimates were usually within .03 of the experimental estimates, and the correlation with the experimental estimates was .99 for math and .88 for reading. Given that OLS provides rigorous

---

[55] In addition, these results cannot be generalized to all CMOs or independent charters since this sample of districts was not representative, but a convenience sample chosen based on data availability.

[56] The exact number of independent charters in each district is not revealed to prevent identification of districts.

impact estimates with more efficient estimation than the propensity score matching used to estimate primary study impacts (OLS does not require a matching process), we chose to use OLS.[57]

For each district, impacts were estimated for all independent charter middle schools (pooled estimate) as well as each CMO's middle schools. To estimate impacts, we compared reading and math achievement outcomes of CMO/independent charter enrollees to those of CMO/independent charter non-enrollees[58] in the same district and grade during the same year, controlling for students' previous test scores and demographic characteristics. The impact model is:

$$(1) \quad y_i = \propto + X_i\beta + \delta T_i + \epsilon_i,$$

where y_i is the reading or math test score outcome for student i two years after the most common enrollment grade (for example, if the most common enrollment grade at a school was fifth grade, this would be the sixth grade);[59] α is the intercept; X_i is a vector of previous achievement and demographic characteristics;[60] T_i is a binary variable for treatment status, indicating whether student i was enrolled in a CMO/independent charter during intake grade; ε is a random error term that reflects the influence of unobserved factors on the outcome; and β and δ are vectors of parameters or parameters to be estimated. The estimated coefficient on treatment status, δ, is the two-year impact estimate.

We addressed school attrition, grade retention, and missing data in the same manner as the primary impact analysis (see Chapter IV and Appendices D and F). Since CMO and district independent charter estimates includes multiple cohorts, it is possible that one cohort of comparison students but a different cohort of charter students could drive the results (the number of charter students generally increases with the length of time the school has been operating). Like the primary impact analysis, we used a weighting approach to make the contribution of each cohort to the impact estimate proportional to charter student (treatment group) sample size. Specifically, each CMO/independent charter student was assigned a weight of 1 and each comparison student was assigned a weight of $\frac{N_{CHARTER}}{N_{DISTRICT}}$, where $N_{CHARTER}$ and $N_{DISTRICT}$ are charter and district sample sizes by grade (if multiple intake grades), cohort, and outcome (reading or math). Weights were rescaled such that the sum of weights equaled the total number of students in the analysis.

[57] For each of the eight CMOs, the propensity-score impact estimates from the primary analysis and OLS impact estimates were always within .06 standard deviations except for the ELA estimate for CMO 8. For this CMO, the propensity estimates included a school outside the district and that may partially explain the difference.

[58] Independent charter schools were eligible comparison students for CMO and CMO students were eligible comparison students for independent charter students. In both cases, charter students were a small portion of the comparison group.

[59] As with the primary impact analysis in this study, two year impacts were chosen because they estimate more relevant longer-term impacts; three-year or four-year impacts were not possible because some students took course-specific math tests in their third and fourth years and the course taken could be affected by treatment.

[60] The characteristics were: pre-baseline (two grades prior to entry) and baseline test scores in reading and math with corresponding missing test score indicators; race/ethnicity; poverty status (as measured by eligibility for free or reduced-price lunches); special education status, ELL status, and whether the student was enrolled in a charter school at baseline. If the charter schools in the analysis had multiple entry grades, a covariate for entry grade was also included in the analysis.

2. Results

Results are shown in Table I.1. In the one district with three CMOs, CMO middle schools generally had more positive impacts than independent charters, and in the other three districts CMOs generally had more negative impacts than independent charters. Although the trend slightly favored independent charters, that may simply result from the districts sampled, a convenience sample not randomly selected.[61] Two CMOs had larger positive impacts than the average independent charter in their district. Five CMOs had negative impacts in districts where the average independent charter impact was positive. As noted, any observed differences between CMO and independent charter impacts may not be caused by CMO management, but may result from other differences between CMOs and independent charters in the sample districts.

Table I.1. CMO and Independent Charter Impacts in Four Districts

	Math	Reading
District A		
All Independent Charters	.21**	.02
CMO 1	.55**	.27**
CMO 2	.26**	-.06**
CMO 3	.57**	.23**
District B		
All Independent Charters	.17**	-.10*
CMO 4	-.22**	-.16**
CMO 5	-.08**	-.06*
District C		
All Independent Charters	.09**	.01
CMO 6	-.27**	-.12**
CMO 7	-.21**	-.16**
District D		
All Independent Charters	.10*	.24**
CMO 8	-.28**	-.29**

*Significantly different from zero at the .05 level, two tailed test.
**Significantly different from zero at the .01 level, two tailed test.

[61] For example, five of the eight CMOs in this sample (about 63 percent) had statistically significant negative impacts in math, but in the primary impact analysis only 7 of 22 did (approximately 32 percent).

APPENDIX J

SUBGROUP IMPACTS

We examined whether CMO schools have a differential impact on two-year math and reading test scores for five subgroups of students. The five subgroups are: males vs. females, African American vs. non-African American students, Hispanic vs. non-Hispanic students, students eligible for free- or reduced-price lunch (FRPL) vs. students who are not eligible for FRPL, and those below vs. above average on baseline test scores (where the average is based on the baseline scores of all students in either the district or state in the same subject).

The subgroup analyses were performed on students who were in our main impact analysis sample by augmenting our CMO-specific impact model with an interaction between treatment group status and the subgroup indicator of interest. See Appendix D for details on our CMO-specific impact model. The standard errors were adjusted for clustering of students within schools.

Due to their exploratory nature, we did not adjust subgroup impact results for multiple comparisons. Furthermore, some subgroup analyses were performed in a limited number of CMOs due to insufficient data (for example, missing FRPL status) or homogeneous CMO student populations (defined as less than 15% or greater than 85% of the treatment group sample belonged to a specific subgroup). Thus, these results should be interpreted with caution.

Overall, we found little evidence that the CMO-specific impacts consistently vary by subgroup. However, as noted in chapter IV, for math, Hispanics had significantly larger impacts than non-Hispanics in five out of the nine CMOs for which we could estimate impacts for Hispanics; Hispanics had significantly smaller impacts in only one of these CMOs. And in reading, Hispanics had significantly larger impacts in four out of nine of these CMOs and there were no CMOs where Hispanics had significantly smaller impacts.

Tables J.1 and J.2 below present detailed results of all five subgroup differences in impacts on two-year math and reading test scores, respectively.

Table J.1. Subgroup Differences in Impacts on Students' Two-Year Math Test Score, by CMO

CMO	Male (vs. Female)	African American (vs. non-African American)	Hispanic (vs. non-Hispanic)	FRPL-eligible (vs. non FRPL-eligible)	Below Average Baseline Achievement[a] (vs. Above Average Baseline Achievement)
A	-0.01 (0.07)	N/A	N/A	N/A	N/A[b]
B	-0.05 (0.04)	N/A	N/A	N/A	-0.07 (0.05)
C	-0.01 (0.07)	N/A	N/A	N/A	0.01 (0.06)
D	0.08 (0.07)	0.16* (0.07)	-0.17* (0.07)	-0.18* (0.09)	0.02 (0.09)
E	0.11 (0.09)	-0.13 (0.13)	0.52** (0.17)	0.45** (0.11)	0.12 (0.15)
F	-0.06 (0.04)	N/A	N/A	N/A	0.09 (0.06)
G	0.06 (0.03)	N/A	N/A	0.01 (0.07)	-0.06 (0.10)
H	-0.01 (0.07)	N/A	0.02 (0.05)	N/A	0.09* (0.04)
I	0.04 (0.05)	-0.18** (0.04)	0.18** (0.04)	N/A	-0.10* (0.05)
J	-0.06 (0.03)	N/A	N/A	-0.01 (0.05)	-0.08 (0.06)
K	-0.05 (0.03)	N/A	0.15** (0.03)	N/A	0.13** (0.03)
L	0.09* (0.04)	N/A	N/A	N/A	0.05 (0.04)
M	0.02 (0.03)	-0.17 (0.11)	0.31* (0.12)	N/A	0.06** (0.02)
N	0.00 (0.05)	N/A	N/A	N/A	0.13 (0.07)
O	0.01 (0.04)	-0.12* (0.05)	0.12* (0.05)	N/A	0.13** (0.02)
P	-0.02 (0.05)	N/A	N/A	N/A	0.01 (0.05)
Q	-0.24** (0.09)	0.08 (0.09)	N/A	0.05 (0.13)	-0.12 (0.13)
R	0.10 (0.05)	0.03 (0.08)	0.01 (0.09)	-0.11 (0.11)	0.04 (0.09)
S	0.01 (0.02)	N/A	N/A	N/A	-0.05 (0.08)
T	-0.04 (0.06)	N/A	N/A	0.40* (0.19)	-0.05 (0.06)
U	0.03 (0.07)	N/A	N/A	N/A	0.14 (0.15)
V	-0.06* (0.02)	0.02 (0.06)	0.11 (0.07)	N/A	0.07 (0.06)

Table J.1 *(continued)*

Source: State, district and CMO school records.

Note: The differences in impacts of being enrolled in a CMO school by subgroups were estimated by including an appropriate interaction between treatment status and subgroup indicator in the impact regression model. Standard errors, presented in parenthesis, account for clustering of students within schools. CMOs that had less than 15% or greater than 85% of students in a given subgroup in the analysis sample were excluded from these analyses.

[a]Below Average Baseline Achievement students were defined as students who performed below the mean for their district or state on the baseline math test.

[b]Less than 15% of both treatment and matched comparison students were low achieving students.

*Significantly different from zero at the 0.05 level, two-tailed test.
**Significantly different from zero at the 0.01 level, two-tailed test.

Table J.2. Subgroup Differences In Impacts on Students' Two-Year Reading Test Score, by CMO

CMO	Male (vs. Female)	African American (vs. non-African American)	Hispanic (vs. non-Hispanic)	FRPL-eligible (vs. non FRPL-eligible)	Below Average Baseline Achievement[a] (vs. Above Average Baseline Achievement)
A	-0.04 (0.09)	N/A	N/A	N/A	N/A[b]
B	-0.01 (0.03)	N/A	N/A	N/A	0.04 (0.04)
C	0.06 (0.07)	N/A	N/A	N/A	0.02 (0.07)
D	0.12* (0.05)	-0.02 (0.07)	0.09 (0.06)	-0.07 (0.13)	-0.07 (0.06)
E	-0.03 (0.07)	0.34** (0.07)	0.04 (0.06)	0.15* (0.06)	0.16* (0.06)
F	-0.03 (0.05)	N/A	N/A	N/A	-0.07 (0.05)
G	0.00 (0.03)	N/A	N/A	0.08 (0.05)	0.02 (0.06)
H	-0.10 (0.09)	N/A	0.14** (0.04)	N/A	0.09 (0.06)
I	-0.01 (0.03)	0.01 (0.05)	0.00 (0.05)	N/A	-0.11* (0.05)
J	-0.10** (0.04)	N/A	N/A	-0.09 (0.05)	0.10** (0.03)
K	0.01 (0.06)	N/A	N/A	0.36** (0.08)	0.39 (0.26)
L	0.19** (0.04)	N/A	N/A	N/A	0.02 (0.04)
M	0.01 (0.02)	-0.17 (0.09)	0.25** (0.09)	N/A	0.03 (0.05)
N	-0.05 (0.06)	N/A	N/A	N/A	0.01 (0.06)
O	0.08 (0.06)	-0.19** (0.06)	0.19** (0.06)	N/A	-0.02 (0.03)
P	0.03 (0.04)	N/A	N/A	N/A	0.08* (0.04)
Q	-0.13 (0.09)	0.00 (0.13)	N/A	0.00 (0.12)	-0.20* (0.09)
R	0.01 (0.05)	0.18** (0.07)	-0.02 (0.08)	0.13 (0.10)	-0.08 (0.08)
S	0.00 (0.03)	N/A	N/A	N/A	0.01 (0.03)
T	0.00 (0.03)	N/A	0.03 (0.03)	N/A	0.04 (0.03)
U	0.03* (0.01)	N/A	N/A	N/A	0.15** (0.02)
V	-0.01 (0.06)	-0.07* (0.03)	0.14** (0.03)	N/A	-0.02 (0.05)

Table J.2 *(continued)*

Source: State, district and CMO school records.

Note: The differences in impacts of being enrolled in a CMO school by subgroups were estimated by including an appropriate interaction between treatment status and subgroup indicator in the impact regression model. Standard errors, presented in parenthesis, account for clustering of students within schools. CMOs that had less than 15% or greater than 85% of students in a given category in the analyses sample were excluded from these analyses.

[a]Below Average Baseline Achievement students were defined as students who performed below the mean for their district or state on the baseline reading test

[b]Less than 15% of both treatment and matched comparison students were low achieving students

*Significantly different from zero at the 0.05 level, two-tailed test.
**Significantly different from zero at the 0.01 level, two-tailed test.

APPENDIX K

MULTIPLE COMPARISON ADJUSTMENTS FOR IMPACT ANALYSES

Following the procedures and standards of the What Works Clearinghouse we used the Benjamini-Hochberg method to adjust for multiple comparisons for our math and reading impact estimates. The method protects against Type 1 errors and specifically against falsely detecting an impact when multiple impact hypotheses are tested. Table K.1 below presents the p-values of our two-year math and reading impact estimates and the corresponding critical p-values (cutoff) for each statistically significant impact estimate based on our pre-specified level of statistical significance of 5 percent. We observe that all impacts that were statistically significant at the 5 percent level remain statistically significant after applying the Benjamini-Hochberg correction.

Table K.1. P-Values of Two-Year Math and Reading CMO Impacts and Corresponding Critical P-Values after Benjamini-Hochberg Correction for Multiple Comparisons

	Math		Reading	
CMO	*p*-value	Critical *p*	*p*-value	Critical *p*
A	**0.00****	0.03	0.07	N/A
B	**0.00****	0.02	**0.01****	0.03
C	**0.00****	0.00	**0.00****	0.02
D	**0.01***	0.04	**0.01***	0.04
E	0.66	N/A	**0.00****	0.03
F	**0.00****	0.01	0.07	N/A
G	**0.01***	0.04	**0.00****	0.01
H	**0.00****	0.02	**0.00****	0.01
I	**0.01****	0.04	**0.00****	0.02
J	**0.00****	0.03	**0.00****	0.00
K	**0.00****	0.00	**0.00****	0.00
L	0.35	N/A	**0.00****	0.01
M	**0.00****	0.01	**0.00****	0.03
N	**0.00****	0.02	**0.00****	0.02
O	**0.00****	0.03	0.15	N/A
P	**0.00****	0.03	**0.00****	0.02
Q	**0.00****	0.03	0.06	N/A
R	0.22	N/A	0.70	N/A
S	**0.00****	0.01	**0.00****	0.03
T	**0.00****	0.02	**0.01****	0.03
U	0.65	N/A	0.14	N/A
V	**0.00****	0.01	**0.00****	0.01

Notes: P-values that are in bold indicate impacts that remain statistically significant after applying the Bejamini-Hochberg correction for multiple comparisons. A critical p is only calculated for impacts that were statistically significant at the 5 percent level. Results insignificant before the correction remain insignificant.

*Significantly different from zero at the .05 level, two-tailed test.
**Significantly different from zero at the .01 level, two-tailed test.

APPENDIX L

METHODS FOR CORRELATING IMPACTS AND CMO CHARACTERISTICS

Chapter V presents our analysis of the relationship between CMO impacts on student achievement and various CMO practices, structures, and contextual factors. These analyses include both (1) bivariate associations of impacts against individual CMO characteristics, and (2) multivariate analyses where impacts are regressed on several CMO characteristics.

The bivariate analyses make use of bivariate ordinary least squares (OLS) models to gauge the correlation between student impacts and CMO practices. The outcome variables are the estimated two-year middle school impacts of the CMO on students' achievement in reading and math. We treat each estimated impact as if it were the result of a mini-study, using robust standard errors to account for the fact that our outcome variable is an estimated parameter, rather than an observed value (Lewis and Linzer 2005). We present the results of two-tailed tests throughout. Each of our primary hypotheses and many of our secondary hypotheses were specified with reference to a single direction of association. For these hypotheses, one could argue that one-tailed tests would be appropriate. Nonetheless, to be conservative, we present the results of two-tailed tests. (If one believes a one-tailed test is appropriate, one could divide the p values and significance thresholds in half).

The multivariate analyses employ multivariate OLS regressions. Again we use robust standard errors and present the results of two-tailed tests.

The independent variables in the correlational analyses are CMO practices, structures, and contextual factors. Most of these variables are constructed from responses to the Principal Survey. For these analyses, we have data for 19 CMOs and 219 original CMO principal responses (or 294 responses after imputation); these are the CMOs for which we have both estimates of impacts and principal survey data. For some secondary hypotheses and intermediate outcomes, we also make use of measures constructed from responses to the CMO Central Office Staff Survey (for which we have 17 CMOs that also have estimated impacts) and the Teacher Survey (for which we have data for 12 CMOs from 384 teachers).

The association between CMOs' impacts and their practices is meaningful only if there is sufficient variation in impacts. A Q-test conducted on the 19 CMOs with impact data and data from the Principal Survey strongly rejected the null hypothesis of homogeneity of impacts ($p<0.001$ for both subjects), so it is appropriate to attempt to explain variation in student impacts with variation in CMOs' practices.

Variable Construction

Measures drawn from the Teacher Survey and the CMO Central Office Staff Survey are constructed in the manner described in Chapter III. Each coefficient can be interpreted as the expected change in impacts for an increase of one standard deviation in the CMO (or CMO-average) report of that practice.

For measures based on the Principal Survey, we are able to use responses from principals of both CMO schools and matched district comparison schools to generate a measure of the divergence between the practices of CMO schools and nearby district schools. This is comparable to the conceptualization of student impacts, which estimate the difference between a student's achievement in CMO schools as opposed to if she had attended a district school. We make use of the paired schools and create a school-level difference between the measure for the CMO principal and the measure for the district principal. These differences are then aggregated up to the CMO level.

Imputation

All data from the Principal Survey are multiply imputed to account for nonresponse. Multiple imputation is particularly valuable for measures drawn from the Principal Survey because they are constructed at the school-pair level. By multiply imputing, we are able to retain observations from school pairs in which one of the principals responded to the survey and the other did not. Multiple imputation also allows us to take into account uncertainty concerning the responses of nonresponding principals. If we constructed a CMO-level average measure based only on the responses that we have, we would overstate our degree of certainty in the CMO average. In our multiply imputed dataset, this uncertainty will be reflected by somewhat different CMO-level averages in each imputation.

We use five imputations and the ICE procedure in Stata. The imputation model is at the school-pair level. In other words, paired schools appear as a single observation in the dataset, with responses from the CMO school principal and the district school principal treated as separate variables within the observation. The imputation models include:

- School traits: district of the comparison school (when possible), state, and school level (elementary, middle, high).
- Demographic features of the CMO school and the comparison district school: pupil-teacher ratio, percent racial minority in the student body, and percent of the student body that is eligible for free or reduced-price lunch (FRPL).
- Two-year middle school impacts in math and reading.62
- Survey measures: survey response to the same item from the other school in the pair, survey response to other items in the same index by both schools in the pair (where appropriate).

[62] Including the outcome variable in the imputation model may seem circular, but is the preferred approach for reducing bias in estimated coefficients based on imputed data (Moons et al. 2006).

APPENDIX M

CORRELATIONAL ANALYSIS RESULTS

Table M.1. CMO Impacts with Rankings on Baseline Scores, Size, and Practices

CMO	Math Impact	Reading Impact	Baseline Math	Baseline Reading	CMO Size	Behavior Policy	Formative Assessment	Educational Approach	Teacher Coaching	Performance-Based Comp.	Instruct. Time
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
C	0.63**	0.22**	Medium	Medium	Small	High	Medium	Low	High	Low	High
V	0.55**	0.23**	Medium	Medium	Large	High	Low	Medium	High	High	High
M	0.50**	0.22**	Medium	Medium	Small	N/A	N/A	N/A	N/A	N/A	N/A
T	0.42**	0.24**	Medium	Low	Large	Medium	High	High	High	High	Medium
S	0.40**	0.08**	High	High	Small	High	Medium	Low	High	Medium	High
B	0.36**	0.18**	Medium	Low	Large	Medium	Medium	Medium	Medium	Low	Medium
G	0.31*	0.20**	Low	Medium	Large	High	Medium	High	Medium	Medium	Low
F	0.28**	-0.05	Low	Medium	Large	High	High	Medium	Medium	Medium	High
P	0.17**	0.16**	Medium	Medium	Large	Medium	High	High	High	High	Medium
I	0.12**	0.13**	Low	Low	Large	Medium	Low	Low	Medium	High	High
J	0.09**	0.18**	High	High	Large	Medium	Medium	High	High	High	High
E	0.05	-0.13**	High	High	Small	N/A	N/A	N/A	N/A	N/A	N/A
U	0.02	0.06	Low	Low	Large	Medium	Medium	Medium	Low	Medium	Medium
L	-0.02	-0.10**	High	High	Small	High	Low	Low	Medium	Low	Low
R	-0.05	0.01	High	High	Large	Low	Medium	High	Low	High	Medium
O	-0.09**	-0.07	Medium	Medium	Small	Low	High	Medium	Medium	Low	Low
D	-0.12*	-0.10*	Low	Low	Small	Medium	High	N/A	Medium	Medium	Medium
A	-0.13**	-0.09	High	High	Small	Low	Low	High	Low	Low	Low
K	-0.21**	-0.17**	High	Low	Small	Low	Low	Medium	Low	Medium	Low
Q	-0.26**	-0.13	Low	Low	Large	N/A	N/A	N/A	N/A	N/A	N/A
N	-0.27**	-0.22**	Low	Low	Small	Low	High	Low	Low	Medium	Medium
H	-0.30**	-0.15**	High	High	Large	Low	Low	Low	Low	Low	Low

Table M.1 *(continued)*

Note: Columns 1 and 2 report year-two math and reading impacts. The rankings in Columns 3-4 and 6-11 correspond to the percentile rank of each CMO relative to the other CMOs in the sample with data: CMO's in the bottom third receive a 'low' ranking, those in the middle third receive a 'medium' ranking, and those in the top third receive a 'high' ranking. Column 5 reports the size of the CMO, as measured by the total number of schools operated in the fall of 2009. CMOs operating more than 8 schools are classified as "large," while those operating 8 or fewer schools are classified as "small." Column 10 shows rankings on Performance Based Compensation, and column 11 shows rankings on Instructional Time.

*Significantly different from zero at the 0.05 level, two-tailed test.
**Significantly different from zero at the 0.01 level, two-tailed test.

Table M.2. Correlations between Common Instructional Framework Components and Impacts

	Math	Reading
Number of districts where CMO has schools[a]	0.01 (0.02)	0.01 (0.01)
Ratio of central office staff to teachers[a]	-0.05^ (0.02)	-0.04** (0.01)
Ratio of HR and operations staff to teachers[a]	-0.24 (0.15)	-0.17* (0.07)
Ratio of educational support staff to teachers[a]	-0.09^ (0.05)	-0.07** (0.03)
Ratio of finance-related staff to teachers[a]	-0.15^ (0.07)	-0.12** (0.04)
Ratio of other central office staff to teachers[a]	-2.16 (3.17)	-1.09 (1.74)
CMO specifies student behavior policies	-0.06 (0.10)	-0.05 (0.05)
CMO provides system of assessments	0.10 (0.08)	0.05 (0.04)
CMO sets teacher salaries and evaluation	-0.01 (0.10)	-0.02 (0.05)
Frequency of CMO visits to school	0.01 (0.11)	0.01 (0.05)
CMO provides professional development support	0.19* (0.08)	0.12* (0.05)
CMO provides assistance in areas with weak test scores	0.12 (0.11)	0.05 (0.05)
Principal previously worked in CMO	-0.01 (0.13)	-0.05 (0.06)
Importance of sample teaching performance for teacher hiring	-0.00 (0.11)	0.03 (0.05)
Importance of commitment to school mission for teacher hiring	0.04 (0.08)	0.00 (0.04)
Fraction of teachers hired from TFA / Teaching Fellows[a,b]	0.97** (0.27)	0.28 (0.19)
Teachers can earn tenure	0.08 (0.09)	0.04 (0.04)
Principal can get bonus for student achievement results	-0.06 (0.06)	0.01 (0.04)
Principal salary	0.07 (0.08)	0.07 (0.04)
Teacher looping over grades	0.04 (0.09)	0.04 (0.04)
Students instructed in math/ reading with students of similar ability	-0.01 (0.11)	-0.02 (0.04)
Importance placed on students exceeding state academic standards	0.16^ (0.07)	0.08^ (0.03)

Table M.2 *(continued)*

	Math	Reading
Enrollment of school	-0.00 (0.08)	-0.02 (0.04)
Teacher-student ratio in math / reading classes	-0.03 (0.13)	0.04 (0.05)
New students required to attend summer session	0.14* (0.05)	0.07* (0.03)
Teachers are covered by collective bargaining agreement	-0.09^ (0.05)	-0.01 (0.04)
Frequency of teacher observations by master teacher / teaching coach[a,c]	0.29** (0.06)	0.11^ (0.06)
Frequency of teacher observations by monitor of teacher performance[a,c]	0.21^ (0.09)	0.04 (0.07)
Frequency of teachers receiving feedback in formal evaluation[a,c]	0.17 (0.14)	0.03 (0.09)
Frequency of teachers receiving feedback outside of formal evaluation[a,c]	0.26** (0.08)	0.09 (0.06)
Frequency of CMO staff observing classrooms[a,d]	0.04 (0.04)	-0.01 (0.02)
Frequency of CMO staff meeting with teachers one-on-one[a,d]	0.00 (0.05)	0.00 (0.03)
Frequency of CMO staff meeting with principals one-on-one[a,d]	0.27** (0.08)	0.15** (0.04)
Frequency of CMO staff performing school walkthroughs[a,d]	0.15^ (0.08)	0.07 (0.05)
Frequency of CMO staff analyzing or explaining data in school[a,d]	0.15^ (0.08)	0.08^ (0.04)

Source: State, district, and CMO school records, Principal Survey, Teacher Survey, and CMO Central Office Staff Survey.

[a]Measure is on a natural scale (not standardized).

[b]Scale is 0-1.

[c]Scale is 1-5.

[d]Scale is 0-5.

^Significantly different from zero at the .10 level, two-tailed test.
*Significantly different from zero at the .05 level, two-tailed test.
**Significantly different from zero at the .01 level, two-tailed test.

Table M.3. Correlations between States' Autonomy Afforded to Charters and Impacts

	Math	Reading
Low (TX, IN, OH, NV, MD) (reference group)	---	---
Medium (NY, IL, CO, CT, GA, LA, NJ)	0.08 (0.18)	0.01 (0.09)
High (DC, FL, AZ, PA, OR, CA, MO)	-0.01 (0.14)	0.03 (0.08)
Constant	0.13 (0.09)	0.02 (0.05)

Source: State, district, and CMO school records, and data from the National Alliance for Public Charter Schools.

^Significantly different from zero at the .10 level, two-tailed test.
*Significantly different from zero at the .05 level, two-tailed test.
**Significantly different from zero at the .01 level, two-tailed test.

Table M.4. Correlations between Instructional Coherence, Organizational Health, and Impacts

	Correlation with Math Achievement Impacts	Correlation with Reading Achievement Impacts
Instructional Coherence		
Use of a common instructional framework	0.13^ (0.07)	0.04 (0.05)
Cooperative and supportive school environment	0.08 (0.07)	0.02 (0.05)
Organizational Health		
High teacher job satisfaction	0.01 (0.09)	-0.01 (0.06)
Stable administrative staff	-0.02 (0.13)	-0.05 (0.06)
Number of applications for teaching positions	-0.01 (0.08)	-0.01 (0.04)
Rate of student turnover	0.00 (0.10)	0.01 (0.06)
Minimal legal, financial, and other administrative challenges	0.03 (0.14)	-0.00 (0.07)

Source: State, district, and CMO school records, Principal Survey, and Teacher Survey.

^Significantly different from zero at the .10 level, two-tailed test.
*Significantly different from zero at the .05 level, two-tailed test.
**Significantly different from zero at the .01 level, two-tailed test.

Table M.5. School-Level Correlations Between Six Primary CMO Practices and Impacts

	Math	Reading
Consistent educational approach	-0.02 (0.05)	-0.02 (0.03)
Comprehensive behavior policy	0.08* (0.03)	0.05* (0.02)
Emphasize formative assessment	0.04 (0.05)	0.02 (0.02)
Emphasize intensive teacher coaching	0.04 (0.03)	0.02 (0.02)
Emphasize performance-based compensation	-0.06 (0.05)	-0.02 (0.03)
Instructional hours per year	0.02 (0.04)	0.01 (0.02)

Source: State, district, and CMO school records, and Principal Survey

^Significantly different from zero at the .10 level, two-tailed test.
*Significantly different from zero at the .05 level, two-tailed test.
**Significantly different from zero at the .01 level, two-tailed test.

Table M.6. Multivariate Associations between Primary CMO Practices and Impacts

	Math	Reading
Comprehensive behavior policy	0.15** (0.05)	0.06^ (0.04)
Emphasize intensive teacher coaching	0.12 (0.07)	0.05 (0.05)
Instructional hours per year	0.04 (0.08)	0.02 (0.06)
Constant	-0.00 (0.06)	-0.01 (0.05)
R^2	0.62	0.38

Source: State, district, and CMO school records, Principal Survey.

^Significantly different from zero at the .10 level, two-tailed test.
*Significantly different from zero at the .05 level, two-tailed test.
**Significantly different from zero at the .01 level, two-tailed test.

Table M.7. Correlations between Behavior Policy Components and Impacts

	Math	Reading
Consistent school-wide behavioral standards and disciplinary policy	0.15^ (0.08)	0.06 (0.05)
Zero tolerance policies for potentially dangerous behaviors	0.21^ (0.09)	0.10^ (0.05)
Schools have behavior code with student rewards	0.14^ (0.07)	0.05 (0.04)
Schools have behavior code with student sanctions	0.17* (0.06)	0.07^ (0.03)
Parent or student signs responsibility agreement	0.12 (0.09)	0.06 (0.05)

Source: State, district, and CMO school records, Principal Survey.

^Significantly different from zero at the .10 level, two-tailed test.
*Significantly different from zero at the .05 level, two-tailed test.
**Significantly different from zero at the .01 level, two-tailed test.

Table M.8. Correlations between Teacher Coaching Components and Impacts

	Math	Reading
Frequency with which new teachers observed by coaches	0.12^ (0.06)	0.05 (0.03)
Frequency with which new teachers observed by principals / administrators	0.11 (0.06)	0.03 (0.04)
Frequency with which new teachers receive feedback from observers	0.10 (0.08)	0.04 (0.04)
Frequency with which new teachers must submit lesson plans for review	0.18 (0.10)	0.08 (0.04)

Source: State, district, and CMO school records, Principal Survey.

^Significantly different from zero at the .10 level, two-tailed test.
*Significantly different from zero at the .05 level, two-tailed test.
**Significantly different from zero at the .01 level, two-tailed test.

Table M.9. Correlations between Common Instructional Framework Components and Impacts

	Math	Reading
Curriculum focuses on helping students meet learning standards	0.31 (0.18)	0.10 (0.12)
Consistency among same-grade teachers in material taught	0.57 (0.34)	0.22 (0.20)
Coordination across grades in material taught	-0.01 (0.24)	-0.07 (0.14)
Teachers follow school-wide instructional calendar / pacing plan	0.27 (0.21)	0.04 (0.15)
Continuity across school instructional initiatives	0.26 (0.21)	0.05 (0.13)
School has well-defined learning standards for all students	0.33^ (0.17)	0.12 (0.11)
Consistent school-wide grading and passing standards	0.31 (0.20)	0.09 (0.14)
Curriculum has clear path for accelerating under-achieving students	0.36^ (0.19)	0.10 (0.13)
Teachers have access to student assessment information	0.30* (0.11)	0.10 (0.10)
Teachers frequently modify lesson plans based on student test results	0.62** (0.10)	0.28** (0.07)
School day organized to minimize events that reduce instruction time	0.33 (0.23)	0.05 (0.15)

Source: State, district, and CMO school records, Teacher Survey.

Note: Each measure is on a 1-4 scale.

^Significantly different from zero at the .10 level, two-tailed test.
*Significantly different from zero at the .05 level, two-tailed test.
**Significantly different from zero at the .01 level, two-tailed test.

Table M.10. Revising Lesson Plans as Mediator of Associations between Primary Hypotheses and Impacts

	Math	Reading
Teachers frequently modify lesson plans based on student test results	0.62** (0.16)	0.30* (0.11)
Comprehensive behavior policy	0.00 (0.09)	-0.01 (0.06)
Teachers frequently modify lesson plans based on student test results	0.52* (0.15)	0.20^ (0.10)
Emphasize intensive teacher coaching	0.11 (0.11)	0.09 (0.07)
Teachers frequently modify lesson plans based on student test results	0.57** (0.11)	0.24* (0.07)
Instructional hours per year	0.05 (0.06)	0.04 (0.05)

Source: State, district, and CMO school records, Principal Survey, and Teacher Survey.

^Significantly different from zero at the .10 level, two-tailed test.
*Significantly different from zero at the .05 level, two-tailed test.
**Significantly different from zero at the .01 level, two-tailed test.

Table M.11. Interactions between Intensive Teacher Coaching and Other Core Measures

	With Formative Assessment		With Performance- based Compensation	
	Math	Reading	Math	Reading
Intensive teacher coaching	0.15 (0.10)	0.06 (0.06)	0.17 (0.11)	0.07 (0.05)
Emphasize formative assessments	0.06 (0.11)	0.04 (0.07)		
Intensive teacher coaching * Emphasize intensive teacher coaching	-0.04 (0.16)	-0.05 (0.08)		
Emphasize performance-based compensation			-0.04 (0.11)	0.01 (0.06)
Intensive Coaching * Emphasize performance-based compensation			0.06 (0.10)	0.01 (0.05)

Source: State, district, and CMO school records, Principal Survey.

^Significantly different from zero at the .10 level, two-tailed test.
*Significantly different from zero at the .05 level, two-tailed test.
**Significantly different from zero at the .01 level, two-tailed test.

MATHEMATICA Policy Research

center on reinventing public education

UNIVERSITY of WASHINGTON | BOTHELL