WORKING PAPER

Do Charter Schools Improve Student Achievement? Evidence from a National Randomized Study

December 2011

Melissa A. Clark, Philip Gleason,
Christina Clark Tuttle (Mathematica Policy Research);
and Marsha K. Silverberg (U.S. Department
of Education)

MATHEMATICA
Policy Research

Abstract

This paper presents findings from the first national randomized study of the impacts of charter schools on student achievement, which included 36 charter middle schools across 15 states. The paper compares students who applied and were admitted to these schools through randomized admissions lotteries with students who applied and were not admitted. It finds that, on average, charter middle schools in the study were neither more nor less successful than traditional public schools in improving student achievement. However, impacts varied significantly across schools and students, with positive impacts for more disadvantaged schools and students and negative impacts for the more advantaged.

Introduction

Charter schools are a central component of current efforts to reform the public education system in the United States. These schools are publicly financed, but are freed from many of the regulations that govern traditional public schools, such as those involving staffing, curriculum, and budget decisions. As of fall 2010, more than 5,400 charter schools served about 1.7 million students—about 3.5 percent of all public school students—in 40 states and the District of Columbia.[1] These numbers reflect rapid growth in the charter school sector in recent years; for example, there were just 2,800 charter schools serving 0.7 million students as of 2003. The number of charter schools and students is likely to continue to increase in response to the federal Race to the Top program, first introduced in 2009, which gave states incentives to remove caps on charter school growth in order to compete for millions of dollars in federal grants.

Yet despite the increased policy emphasis on charter schools and the growth in their numbers, rigorous evidence of their effectiveness on a broad scale is limited. Previous research includes student fixed effects analyses across several school districts or states (see, for example, Sass 2006; Betts et al. 2006; Bifulco and Ladd 2006; Booker et al. 2007; Hanushek et al. 2007; Ballou et al. 2008; Zimmer et al. 2009)[2] and lottery-based studies that each focused on a single large urban area (Hoxby and Rockoff 2005; Hoxby et al. 2009; Dobbie and Fryer 2009; Abdulkadiroglu et al. 2009; Angrist et al. 2010). The fixed effects studies have typically found impacts that were insignificant or negative, while the lottery-based studies have found impacts that were large and positive.[3] The previous fixed effects analyses potentially provide meaningful external validity through the inclusion of a relatively broad geographic sample of schools, but their internal validity may be compromised if students attending charter schools in a given year differ from those who do not in ways that are not fully captured by the fixed effects models. In contrast, the lottery-based studies potentially provide strong internal validity by comparing lottery applicants who

---

[1] Center for Education Reform. "K-12 Facts." Available at [http://www.edreform.com/Fast_Facts/K12_Facts/]. Accessed November 2009.

[2] Another recent non-experimental study of charter schools was conducted by the Center for Research on Education Outcomes (CREDO 2009). This study did not use the same fixed effects approach used by the other studies cited here, but instead used a matching procedure to compare the year-to-year growth in test scores among a sample of charter school students with the growth in test scores among a comparison sample of students in traditional public schools. The study found that charter schools, on average, had small negative impacts on student achievement in reading and mathematics.

[3] Exceptions to the findings of negative or insignificant effects in the fixed effects literature include Witte et al. (2007), who found positive impacts in Wisconsin, and Ballou et al. (2008), who found positive impacts for charter elementary schools but no statistically significant impacts for charter middle schools in Idaho.

were randomly admitted to charter schools to those who were not. However, because they focus on single, large, urban areas (Boston, New York, and Chicago), their findings are not broadly generalizable to charter schools nationwide, fewer than half of which are located in large, urban areas (Gleason et al. 2010).

Another challenge faced by previous lottery-based studies is that they have relied on schools to hold truly random admissions lotteries and to adequately document the students who participated in the lotteries, any special rules or procedures for conducting the lotteries, and the original randomly ordered list of the lottery results. Complicating factors include the admission of selected students (such as siblings of current students) to the school outside of the lottery process, stratified lottery procedures with unequal chances of admission for students in different strata, special rules for students who apply together, and unexpected problems with the lottery mechanism (McEwan and Olsen 2010; Tuttle et al. forthcoming). Moreover, in some cases schools do not document the distinction between students on the waiting list who participated in the lottery and those who applied after the lottery. The implication of these complications is that the lottery results provided by schools after the fact may not include sufficient information for researchers to appropriately account for each student's true probability of admission to a charter school, and may in some cases be inaccurate.[4] Due to potential problems such as this, one cannot be entirely confident that the indicator of a "charter school treatment" in existing lottery-based studies is truly exogenous and that the studies are free from selection bias.

To address these challenges and provide rigorous evidence of charter schools' effectiveness, this article presents findings from an evaluation of 36 charter middle schools in 15 states.[5] Through the use of these schools' randomized admissions lotteries to determine the student sample, the study is the first to provide rigorous evidence of charter school impacts on student achievement from a large and geographically diverse sample. The study team's careful monitoring of the charter school admissions lotteries helped to ensure that the lottery procedures, and students' resulting admission status, were truly random, so that the study results would be as rigorous as possible.

Consistent with many previous studies that have focused on broad sets of charter schools, we found no evidence that, on average, attending charter schools had a positive impact on student achievement. The estimated impact of attending the average charter school in the study was negative but not statistically significant after adjusting for the multiple hypotheses tested. However, the average impact of attending charter schools in large urban areas or those serving lower achieving or more disadvantaged students was large and positive. In contrast, the average impact of charter schools in non-urban areas or those serving higher achieving or more advantaged students was large and negative. While the study's design does not allow us to determine the causes of this variation in impacts with the same level of rigor as the impact estimation itself, we present exploratory evidence examining various potential hypotheses.

The rest of this paper proceeds as follows: Section I provides background on the charter school movement. Section II describes the evaluation design, Section III describes the

[4] For example, in their lottery-based study of a single Massachusetts school, Angrist et al. (2010) acknowledge that "for some applicants, lottery status was over-written with enrollment status."
[5] The research presented here was part of an evaluation of charter schools conducted by Mathematica Policy Research for the U.S. Department of Education's Institute of Education Sciences (Gleason et al. 2010). Data used in the analyses are available in a restricted use file, which researchers can request from the U.S. Department of Education's National Center for Education Statistics (NCES) through its Electronic Application System, available at http://nces.ed.gov/statprog/instruct.asp. In accordance with NCES publication policy, sample sizes from analyses presented in this paper that were not previously reported in Gleason et al. (2010) are rounded to the nearest 10.

data collection, and Section IV describes the analytic methods. Section V presents the impact estimates, Section VI explores reasons for variation in impacts across sites, and Section VII concludes.

Background on Charter Schools

Charter schools are public schools that are established on the basis of a contract, or charter, that a private board holds with a charter authorizer over some pre-determined number of years. As part of the contract, charter schools are released from many state and district regulations that govern traditional public schools, such as those involving staffing, curriculum, and budget decisions. In exchange for this flexibility, charter schools are expected to be held accountable for the quality of student outcomes and may be closed by their authorizer if they fail to meet expectations. Charter schools are typically open enrollment schools—in most cases, any student within the district or state in which the school is located may attend the school if space is available. Proponents argue that the schools' autonomy allows them to innovate, test new ideas, and bring competitive pressures to improve traditional public school systems. Critics are concerned that these schools draw students and resources away from traditional public schools and that inadequate oversight will lead to many low-quality charter schools.

The charter school movement in the United States is nearly two decades old. The first charter school opened its doors in Minnesota in 1992. The number of states permitting charter schools grew rapidly during the 1990s, as did the number of charter schools and students enrolled. The growth in the number of charter schools and the number of students they enrolled continued to increase into the 2000s, despite the fact that only four new states passed authorizing legislation between 1999 and 2003 and none did so between 2004 and 2009. Charter schools are likely to see another period of significant growth as a result of guidelines drafted in July 2009 for grant applications under the Race to the Top Fund established under the American Recovery and Reinvestment Act of 2009. The criteria for aid receipt include the extent to which a given state has legislation that "does not prohibit or effectively inhibit increasing the number of charter schools … or otherwise restrict student enrollment in charter schools."[6] As of May 2010, four states (Delaware, Illinois, Louisiana, and Tennessee) had enacted new legislation to raise or eliminate existing limits on charter school growth.[7]

Study Design

The study's experimental design relies on the random assignment of students through the lotteries held by oversubscribed charter schools—schools that had a larger number of applicants than they had spaces available. The lottery winners form the treatment group for the evaluation, while the lottery losers form the control group. The randomized lotteries ensure that the only systematic difference between the treatment and control groups is whether they were admitted to a study charter school—on average, there should be no differences in the characteristics, motivation, or expectations of the students or their parents. Therefore, comparing the outcomes of the two groups yields unbiased estimates of the causal effects of being offered admission to the charter schools in the study.

---

[6] See the Notice of Proposed Priorities for the Race to the Top Fund, available at http://www.ed.gov/programs/racetothetop/index.html.

[7] Robelen, Erik W. "State Picture on Charter Caps Still Mixed." Education Week, August 12, 2009 (updated May 2010). Accessed at http://www.edweek.org/ew/articles/2009/08/03/37charter.h28.html?tkn=OOYFD4TYjprEuWKkM2KUf yD6RRL5hvVuQB1Z on August 2, 2011.

The Sample of Charter Middle Schools

Schools were recruited for the study sample over a two-year period from any state with eligible charter schools. To be eligible for the study, a charter school had to meet three criteria. First, its entry grade had to be between grades 4 and 7.[8] Second, it had to have been operating as a charter school for at least two years when it was recruited to minimize the chances that participating schools would still be under development and thus undergo a substantial amount of change during the evaluation period. Third, it had to be sufficiently oversubscribed—that is, to have more applicants than could be offered admission to the school—so that it could accommodate the study's experimental design.

The first cohort of schools were those holding admissions lotteries for the 2005–2006 school year, and the second cohort were those holding lotteries for the 2006–2007 school year. Using national databases, we identified 492 charter middle schools that had been open at least two years when they were recruited and were thus potentially eligible for the study. Although 77 schools both agreed to participate and initially appeared eligible for the study, ultimately 36 charter schools in 32 sites remained sufficiently oversubscribed through the study period (that is, they had at least 10 students who participated in the lottery but remained too far down on the waiting list to be offered admission) and participated in the study in at least one of the two study years (Gleason et al. 2010).[9]

Table 1 compares the charter middle schools in the study sample with all other charter middle schools nationwide, based on a survey conducted by the evaluators.[10] The study sample is statistically similar to the nonstudy charter middle schools along several dimensions, including percentage located in a large urban area, student enrollment, student-teacher ratio, length of school day and year, teacher experience and certification, and revenue per pupil. However, there are some differences that generally suggest the study charter schools serve a somewhat more advantaged student population than the schools that were not included in the study. For instance, the schools in the study sample have a higher percentage of white students, on average, and a lower percentage of black students, than the nonstudy charter middle schools. They also have a lower percentage of students who are eligible for free or reduced-price lunches and a higher percentage of 7th graders who meet their state proficiency standards in both reading and math. Not surprisingly, given that schools had to have been in operation for at least two years to be eligible for inclusion in the study, the average study charter school had been in operation longer than the average nonstudy charter school (7.0 versus 5.9 years).

These differences highlight the notion that the school sample is not nationally representative, and that impacts for the selected sample may differ from those of all charter middle schools nationwide. Nonetheless, the fact that the study includes a broad set of

[8] This grade range was chosen primarily on the basis of the availability of both baseline and follow-up test score data from school records—by federal law, schools are required to test students in reading and math in grades 3-8. While schools with 4th grade entry were eligible for inclusion in the study sample, the primary analysis sample only includes schools with entry grades ranging from 5 to 7—we refer to these as "middle schools."

[9] In general, each site corresponded to a single charter school. However, five pairs of participating charter schools had common applicants to their lotteries—we refer to these as "dual applicants." We treated these pairs of schools as single, combined sites in the analysis. (If a pair of schools had common applicants in one cohort but not the other, they were treated as a single site in the cohort in which they shared applicants and as individual sites in the other cohort.) Ultimately, the final sample included 32 sites (Gleason et al. 2010).

[10] The "other" charter middle schools—those not participating in the study—include charter schools that did not receive enough applications to hold a lottery, that held a lottery but ended up offering admission to most or all of the lottery losers who ended up on a waiting list, and charter schools that refused to participate in the study.

Table 1. Characteristics of Study Charter Schools and Nonstudy Charter Middle Schools	Charter Middle Schools in Study	All Other Charter Middle Schools	Difference	p-Value
Located in Large Urban Area (Percentage)	36%	41%	-4%	0.602
Enrollment (Means)				
Total enrollment	387	298	90	0.080
Enrollment per grade	111	88	23	0.259
Student-teacher ratio	14.6	16.7	-2.1	0.150
Time in School (Means)				
School day length in hours	7.3	7.0	0.3	0.117
School year length in days	182.4	181.4	1.0	0.968
Staff				
Experience of principal (mean number of years as principal)	6.1	5.7	0.5	0.562
Percentage of schools at which 2/3 of teachers have 5+ years experience	50%	34%	16%	0.060
Midpoint of teacher salary range at school (mean)	$48,168	$44,280	$3,888	0.022*
Percentage of teachers at school with full state certification (mean)	77%	78%	-2%	0.924
Characteristics of Students at School (Means)				
Percentage Hispanic	26%	25%	1%	0.825
Percentage White	53%	38%	15%	0.012*
Percentage Black	16%	29%	-13%	0.024*
Average daily attendance rate	95%	92%	4%	0.067
Percentage of students receiving free or reduced-price lunch	44%	62%	-18%	0.003**
Percentage of students with learning disability and/or IEP	12%	12%	0%	0.705
Percentage of students classified as LEP	3%	9%	-6%	0.069
Academic Achievement of Students at School (Means)				
Percentage of 7th graders meeting state proficiency in math	66%	51%	15%	0.001**
Percentage of 7th graders meeting state proficiency in reading	75%	57%	19%	<0.001**
Autonomy Index (Mean)	4.6	5.2	-0.6	0.083
Charter School Characteristics				
Age of school (mean)	7.0	5.9	1.2	0.015*
Authorized by local school district (percentage)	56%	44%	12%	0.214
Operated by CMO (percentage)	11%	20%	-9%	0.384
Total $ revenues per student, including private funding	$8,030	$8,710	-$679	0.402
Accountability Index (Mean)	2.59	2.45	0.14	0.296
Sample Size: Characteristics Based on Principal Survey or Common Core of Data	36	434		
Sample Size: Characteristics Based on Principal Survey Alone	35	299		
Sample Size: School Test Scores	36	380		

Source: Gleason et al. (2010).

Note: The sources of the information provided in this table include a survey administered in fall 2006 or fall 2007 to the principals of all charter middle schools nationally, the Common Core of Data from the National Center for Education Statistics, and the School Data Direct database maintained by the State Education Data Center of the Council of Chief State School Officers.

** Difference is statistically significant at the 0.01 level, two-tailed test.

* Difference is statistically significant at the 0.05 level, two-tailed test.

charter middle schools across 15 states in both urban and non-urban areas, and includes schools serving both highly disadvantaged populations and more advantaged populations, allows us to provide estimates that may be more reflective of charter school impacts nationwide than estimates from previous lottery-based studies that have focused on a single urban area.

Charter Schools' Admissions Lotteries

Participating charter schools typically held their admissions lotteries in the winter or spring prior to the school year for which students were applying for admission. To maintain the integrity of the study's experimental design, a member of the study team monitored each lottery to ensure that the mechanism for selecting lottery winners and determining the order of the waiting list was truly random. After documenting the lottery outcomes, we confirmed with the school that our record of the lottery results matched the record of the school and, if there were discrepancies, we worked to resolve them. We also documented any special features of the lottery, including exemptions, stratification, or special rules for siblings who applied at the same time. We also documented whether sample members applied to more than one charter school participating in the study.

The information we obtained on schools' lotteries enabled us to create sampling weights, reflecting each student's probability of admission, to ensure that the control group of lottery losers formed an appropriate counterfactual for the treatment group of lottery winners in the analysis The sample weights ensured that both the weighted sample of treatment group students and the control group students were representative of the full set of students who applied to the study schools, consented to participate in the study, and participated in the schools' lotteries. Gleason et al. (2010) provides more details on how these weights were calculated.[11]

After the lotteries were conducted and lottery winners were offered admission, the study charter schools continued to admit applicants from the randomly ordered waiting list as space became available. All students who were admitted in the lottery or were offered admission in proper order from the waiting list (whether or not they opted to attend) were included in the study's treatment group, while all other students who participated in the lottery were included in the control group.[12]

---

[11] Without using sample weights to account for students' probability of admission, particular students may have an undue influence on the treatment or control group. For example, since students who apply to more than one study school would have a higher probability of admission than those who apply to just a single school, all else equal, these "dual applicants" would likely be more heavily represented in the treatment group than in the control group without the weights.

[12] The study sample included only those students whose parents consented for them to participate. In almost all of the school lotteries included in the study, parental consent was obtained prior to the schools' admissions lotteries. Obtaining consent prior to the lottery ensured that there was no systematic relationship between the likelihood of consent for a given student and whether he or she was offered admission to the school (and thus was in the treatment group) or not offered admission (and thus was in the control group). The average consent rate among lottery participants in participating charter schools was 62 percent and was statistically equivalent for treatment and control group students (62 percent and 61 percent, respectively). In only 4 of the 41 school lotteries, parental consent for some applicants was obtained after the lottery. Consent rates in these four sites were similar for treatment (89 percent) and control group (87 percent) students.

Student Sample

The full student sample included 2,904 students—1,744 in the treatment group and 1,160 in the control group—from two study cohorts that were each followed over a two-year follow-up period. For the main analysis, we further restricted the sample to a set of 2,330 students (1,400 treatment and 930 control) for whom we could most reliably estimate charter school impacts, by imposing two additional restrictions (Gleason et al. 2010). First, we included only sample members for whom we obtained baseline data on student achievement. Second, we included only students at the charter school sites at which we successfully obtained data on student outcomes for a sufficiently high number and percentage of students in both the treatment and control groups. Results were not sensitive to these exclusions, as described further below.

We restricted the primary analysis sample to those students for whom we obtained achievement data for the baseline year (the school year before the treatment group enrolled in study charter schools) to minimize differences in the availability of outcome data for treatment and control group students, as these differences could bias the impact estimates. Students with baseline achievement data were likely to have attended a public school in the baseline year and thus also were more likely to have attended a public school and have achievement data in the follow-up years, regardless of whether they won or lost the lottery.[13] For instance, among students without baseline achievement data, 63 percent of the treatment group and 30 percent of the control group had valid first follow-up (Year 1) math scores. Rates of missing follow-up scores and the disparity between the treatment and control groups were considerably lower among the sample with valid baseline data—among this sample, 94 percent of the treatment group and 89 percent of the control group had valid Year 1 math scores. This restriction led us to drop 538 students from the analysis sample. Our use of this restriction is consistent with analyses of charter school impacts reported in most of the other lottery-based studies of charter schools.[14]

The second restriction was imposed to ensure the validity of within-site impact estimates, which were averaged to form the overall impact estimates. For the sample from a given site to be considered valid, it had to meet the following three criteria: (1) the treatment and control groups each had to include at least five students with valid data for the outcome being examined, (2) the overall percentage of sample members with valid data for the outcome had to be at least 50 percent in each group, and (3) the difference in the proportion of treatment and control group students with valid data for that outcome could be no larger than 30 percentage points. For sites meeting these criteria, we considered the lottery-based experimental design to have been completed successfully and we retained the site in the primary analysis sample. If the site failed to meet one or more of those criteria, we consid-

[13] More than half (52 percent) of the students without baseline achievement data attended a private school or were home schooled when they applied to a study charter school, compared with less than one percent of those with baseline achievement data. Among those who attended a private school or were home schooled when they applied to the charter school, 90 percent of treatment group students attended a public school (typically the study charter school) during the first follow-up period, compared with only 34 percent of control group students.

[14] In their lottery-based study of charter schools in Boston, for example, Abdulkadiroglu et al. (2009) used a similar sample restriction. Hoxby et al. (2009) restricted the sample upon which their impact estimates were based to students with some test score availability, although they allowed this to be either in the baseline or follow-up period. The non-experimental fixed effects studies of charter school impacts that compare test scores of students in charter schools with their test scores prior to their entry into a charter school also restrict the sample to those with valid achievement data during a baseline period (for example, Sass 2006; Hanushek et al. 2007; Bifulco and Ladd 2006; Zimmer et al. 2009).

ered the implementation of the study's experimental design to be compromised and dropped the site's students from the primary analysis sample used to estimate impacts for that outcome. Most of the study's sites met all three criteria and were included in the primary analysis sample for all outcomes.[15]

Table 2 displays baseline characteristics of treatment and control group students in the main analysis sample. As expected given that the admission lotteries were random, treatment and control group students exhibited few statistically significant differences in baseline characteristics. Of the 32 characteristics in Table 2, there were statistically significant differences between the treatment and control groups for only two.[16] Treatment group students had higher pre-baseline mathematics scores (scores from two years before the treatment group enrolled in the study schools) than control group students.[17] On the other hand, treatment and control group students had identical mean mathematics scores in the baseline year. Treatment group students were also less likely (47 versus 52 percent) to have family incomes above 30 percent of the poverty line. Two statistically significant differences are approximately what we would expect due to chance when examining differences in 32 characteristics with a 5 percent critical value. This suggests that the treatment and control groups in the main analysis sample were well balanced according to baseline characteristics, providing a strong foundation for the impact evaluation. Comparisons of the baseline characteristics of treatment and control group students among the full sample, including those without baseline test scores, showed that these two groups were also well balanced with respect to baseline characteristics (Appendix Table 1), as did comparisons of the characteristics of treatment and control group students among the main analysis sample with valid Year 2 test score data (Appendix Table 2), which was the sample that contributed to the main impact estimates.

Data

To measure the effects of charter schools on student achievement, the evaluation relied on test score data from state assessments.[18] These data were obtained from schools, districts, or states for the baseline year and the preceding "pre-baseline year" as well as for the two follow-up years. Among members of our analysis sample, in Year 1 we obtained valid math scores for 94 percent of the treatment group and 89 percent of the control group, and valid reading scores for 95 percent of the treatment group and 89 percent of the control group. In Year 2, we obtained valid math scores for 90 percent of the treatment group and 84 percent of the control group, and valid reading scores for 91 percent of the treatment group and 84 percent of the control group.

[15] Of the 32 sites, 3 (containing a total of 64 lottery participants) were excluded from the Year 2 reading impact estimates. Four sites (containing 141 lottery participants) were excluded from the Year 1 math impact estimates, and 4 sites (containing 207 lottery participants) were excluded from the Year 2 test score impact estimates (Gleason et al. 2010).

[16] For consistency with our primary impact estimation model, the means presented in Table 2 are estimated at the site level and averaged across sites, giving equal weight to each site. We weighted estimates to account for differential probabilities of assignment to the treatment and control groups in each site.

[17] These pre-baseline data were missing for a substantial portion of the sample (51 percent for pre-baseline reading and math scores, 27 percent for pre-baseline reading proficiency levels, and 29 percent for pre-baseline math proficiency levels).

[18] As part of the overall evaluation, impacts were also estimated for a range of other student outcomes, including other measures of student achievement, student behavior, student and parent satisfaction with school, and parental involvement. These outcomes were measured based on surveys administered to the students in the sample as well as their parents. See Gleason et al. (2010) for details on these findings.

Table 2. Baseline Characteristics of Treatment and Control Group Students in Main Analysis Sample *(Proportions Unless Otherwise Indicated)*

	Mean, Treatment Group	Mean, Control Group	Difference	p-Value of Difference
Achievement (z-score units)				
Baseline reading score	0.42	0.43	-0.01	0.796
Pre-baseline reading score	0.47	0.38	0.09	0.175
Baseline math score	0.45	0.45	0.00	0.997
Pre-baseline math score	0.47	0.32	0.15	0.030*
Disciplinary Measures				
Number of days absent in baseline school year	6.07	5.62	0.46	0.123
Student suspended in baseline school year	0.04	0.03	0.01	0.539
Demographic Characteristics				
White, Non-Hispanic[a]	0.57	0.55	0.02	0.371
Black, Non-Hispanic[a]	0.10	0.09	0.00	0.877
Other race, Non-Hispanic[a]	0.07	0.08	-0.01	0.412
Hispanic	0.27	0.28	-0.02	0.373
Male	0.46	0.48	-0.01	0.590
Age at start of school year (years)	11.53	11.52	0.01	0.552
Has Individualized Education Plan (IEP)	0.18	0.16	0.02	0.560
Limited English proficiency/ELL	0.10	0.08	0.02	0.095
Family Characteristics (proportions)				
Income to poverty ratio 0 to 100 percent	0.13	0.12	0.01	0.475
Income to poverty ratio 100 to 200 percent	0.21	0.19	0.02	0.362
Income to poverty ratio 200 to 300 percent	0.18	0.16	0.02	0.319
Income to poverty ratio >300 percent	0.49	0.54	-0.05	0.033*
Two-parent family	0.78	0.79	-0.01	0.704
Not two-parent family, but more than one adult	0.05	0.04	0.01	0.260
English main language spoken at home	0.89	0.90	-0.01	0.577
Mother's education: high school or less	0.23	0.24	-0.01	0.755
Mother's education: some college	0.35	0.35	0.00	0.867
Mother's education: college	0.42	0.42	0.00	0.924
Born in U.S.	0.92	0.92	0.00	0.895
Family received TANF or food stamps in past 12 months	0.05	0.05	0.00	0.961
Free or reduced-price lunch-eligible	0.34	0.35	0.00	0.844
School Enrollment (proportions)				
Enrolled in charter school at baseline	0.05	0.06	-0.01	0.267
Enrolled in private school at baseline	0.00	0.01	0.00	0.254
Enrolled in public school at baseline	0.94	0.93	0.01	0.352
Home schooled at baseline	0.01	0.00	0.01	0.162
Baseline school type unknown	0.00	0.00	0.00	0.602
Number of Students[b]	1,400	930		
Number of Sites	29	29		

Source: Gleason et al. (2010).
Note: Sample includes students in main analysis sample (students with nonmissing baseline test score data in the sites included in the main impact analyses). Means are estimated at the site-level and averaged across sites, giving equal weight to each site. Estimates are weighted to account for differential probabilities of assignment to the treatment and control groups in each site.
[a] Race categories are mutually exclusive and may not equal 100 percent due to rounding.
[b] Sample size differs for some of the individual baseline characteristics due to differential rates of missing data for different characteristics.
** Difference significantly different from zero at 0.01 level, two-tailed test.
* Difference significantly different from zero at 0.05 level, two-tailed test.

Because sample members were spread across 15 states, each of which administered a different assessment, test scores had to be converted to a comparable scale for the analysis. We converted all scores to z-scores, defined as the student's raw score on the state assessment minus the mean score on the test among all students in the state who took the test, divided by the standard deviation of the scores for that same group, by grade level.[19] Thus, students' z-scores reflect their performance on the state assessment relative to the typical student in that state and grade.

Additional covariates for the impact analysis were obtained from a baseline survey completed by parents when their children applied to a study charter school. The survey collected demographic and socioeconomic information from parents at the time of application, as well as their reasons for applying to the participating charter school and information on other schools to which they were applying. The overall response rate on the baseline survey among analysis sample members was 91 percent—92 percent among the treatment group and 90 percent among the control group.[20]

Analytic Methods

Estimating the Impact of Charter School Admission

To generate intent-to-treat (ITT) estimates of the impact of study charter school admission on various outcomes, we first estimated the impacts in each study site and then averaged them to obtain an overall impact estimate. To obtain the site-level ITT impact estimates, we used the following regression model:

(1) $y_{ij} = \alpha_j + X_{ij}\beta + \delta_j T_{ij} + \varepsilon_{ij}$,

where y_{ij} is the outcome of interest for student *i* in site *j*; α_j is a site-specific intercept; X_{ij} is a vector of characteristics of student *i* in site *j*, including an indicator for whether the student was in cohort 1 or 2 of the sample; T_{ij} is a binary variable for treatment status, indicating whether student *i* won the admission lottery in site *j*; ε_{ij} is a random error term that reflects the influence of unobserved factors on the outcome; β and δ_j are parameters or vectors of parameters to be estimated. The estimated coefficient on treatment status in site *j*, δ_j, represents the impact of admission to a charter school in site *j*. As noted above, observations were weighted to account for unequal selection probabilities in the charter school lotteries. Covariates included baseline test scores, demographic characteristics, and type of school attended at baseline—the full set of covariates is listed in Appendix Table 3. Missing values of covariates were imputed as the mean value of the covariate by site and sample cohort.[21]

[19] This approach for analyzing state assessment data in educational studies involving multiple states is one of the approaches recommended by a recent report on the use of state tests in education experiments (May et al. 2009). It is also similar to the approach used by two other recent multistate studies of charter school impacts (Zimmer et al. 2009; CREDO 2009).

[20] We also collected information on both the charter and non-charter schools attended by students in the sample, as well as on all other charter middle schools nationwide, through a principal survey. This survey, which was sent by mail with telephone follow-up, was completed by 92 percent of the principals of schools attended by treatment group students, 77 percent of the principals of schools attended by control group students, and 70 percent of the principals of all other charter middle schools nationwide.

[21] Standard errors were not clustered at the site level, reflecting the purposive selection of charter schools for the sample. Because of this purposive sample selection, results do not generalize beyond the study charter schools.

To obtain an overall estimate of the average impact of the study charter schools on the outcome of interest, we averaged the site-specific impact estimates $\hat{\delta}$ over the J sites included in the estimation, taking an equally weighted average as follows:

$$(2)\quad \hat{\delta} = \tfrac{1}{J}\sum_{j=1}^{J} \hat{\delta}_j$$

By equally weighting the estimated impacts from each site, we allowed each impact to have an equal influence on the overall impact estimate, thereby providing unbiased estimates of the impact of the average study charter school. However, we also tested the sensitivity of our results to our approach for calculating the average impact by according more weight to more precisely estimated site-level impacts, as described below.

Estimating the Impact of Charter School Attendance

While most (78 percent) of the treatment group attended the study charter school to which they were admitted in the year following the lottery and a few attended a nonstudy charter school, 19 percent of the treatment group did not attend a charter school. A smaller percentage of the control group attended some charter school, with 6 percent attending a study charter school and 9 percent attending another nearby charter school.[22] To investigate the effects of study charter middle schools on the students who actually attended these schools, we used admission to a study charter school through the lotteries as an instrumental variable for charter school attendance. Results reflect the impact of attending a charter school—either a study charter school or a nearby nonstudy charter school—attended by any of the treatment or control group students. As with the ITT estimates, we estimated the treatment on the treated (TOT) impacts in each site and then averaged these estimates over all sites to produce an overall TOT impact estimate.

Subgroup Estimates

In addition to estimating overall effects of study charter school admission for the full study sample, we estimated the impact of study charter school admission for several population subgroups. To estimate these impacts, we used the following regression model:

$$(3)\quad y_{ij} = a'_j + X_{ij}B' + \delta'_j T_{ij} + \gamma'_j S_{ij} + \zeta'_j T_{ij} S + \varepsilon'_{ij} \; ,$$

where S is an indicator for whether the student is in subgroup S, and all other parameters are as defined in equation (1). The estimated coefficient on treatment status, δ_j, provides an estimate of the impact of study charter school admission for students not in subgroup S in site j, and the estimated coefficient on treatment status interacted with subgroup ζ'_J represents the difference in impacts between students in subgroup S and students not in subgroup S in site j. Summing ζ'_J and ζ'_J thus provides an estimate of the impact for students in subgroup S in site j. We then averaged the impact estimates for each subgroup across all sites to obtain an overall impact estimate for that subgroup (following the same approach used to average impact estimates for the full sample in equation (2)).

[22] Students who initially lost the lottery to a study charter school and were not offered admission to the school from the waiting list through the beginning of the school year were assigned to the control group regardless of whether they later gained admission to the school. Some of these control group students received "late offers" to attend the study charter school during the second semester of that school year. These control group students who received late offers are included among the 6 percent of the full control group who attended a study charter school.

Sensitivity Analyses

To assess the sensitivity of our main estimates to the specific estimation method used, we also estimated impacts using several alternative approaches, including an alternative approach to averaging site-level impact estimates, inclusion of covariates, alternative rules for dropping or retaining sites, and alternative approaches to accounting for missing outcome test score data.

Method of Averaging Impacts Across Sites. To obtain our main impact estimates, we computed an equally weighted average of the site-level impact estimates (equation (2)). Thus, sites with estimated impacts based on relatively small samples received the same weight as sites with impacts based on large samples. To test the sensitivity of our results to this approach for weighting site-level impact estimates, we estimated impacts by using a two-stage generalized least squares procedure described by Hanushek (1974). This approach assigns more weight to more precisely estimated site-level impacts. The GLS approach may be statistically more efficient than the equally weighted average.

Inclusion of Covariates. Our main model controlled for baseline student test scores and other baseline student characteristics. Controlling for baseline characteristics improves the precision of the impact estimates. However, as noted by Freedman (2008), theory suggests that inclusion of baseline covariates may bias impact estimates, although in practice this bias tends to be small (Schochet 2010). To assess the sensitivity of our models to inclusion of baseline covariates, we estimated models that did not include any covariates other than site fixed effects and site-treatment status interactions.

Rules for Dropping or Retaining Sites. Our main impact estimates excluded sites with fewer than five treatment or control group students, an overall response rate lower than 50 percent, or a difference in response rates between treatment and control groups greater than 30 percentage points. Since each site represented a separate and independent experiment, we dropped sites in which we felt that the integrity of the design could be called into question. However, we could have reasonably applied different rules for retaining or dropping sites. To assess the sensitivity of our results to these restrictions, we estimated models that included all sites with any valid data.

Inclusion of Students with Missing Baseline Test Scores. As described above, to minimize the possibility of bias attributable to differential rates of missing test score outcome data between the treatment and control groups, we limited the sample to students with valid baseline test score data. Such students were more likely to have nonmissing follow-up test scores regardless of admission to a study charter school. As an alternative to accounting for missing outcome data, we estimated impacts by using data from all sample members, regardless of whether they had valid baseline test scores, and adjusted for differential rates of missing outcome data by using nonresponse weights.[23] In addition, to assess the possible effects of bias attributable to differential rates of missing data under the most extreme circumstances, we estimated bounds on the impact estimates by following an approach proposed by Lee (2005).[24]

[23] In particular, we adjusted our basic sampling weights, which account for students' likelihood of being in the treatment or control group, so that the overall nonresponse weights also accounted for differences between the characteristics of sample members for whom we have outcome data versus those for whom we do not have outcome data.

[24] This approach identified the excess proportion of lottery losers with missing data relative to lottery winners. Then, given that the two most extreme possible situations for determining the impact estimate were that all the (unobserved) lottery losers with missing data were either in the upper or lower tail of the test score distribution, the approach established bounds on the impact estimate based on the two extremes. "Trimming" the upper tail of the test score distribution among lottery winners provided a lower bound on the impact estimate; trimming the lower tail provided an upper bound.

Multiple Hypothesis Testing

As is well documented, standard hypothesis testing procedures may yield misleading results if impacts are estimated on multiple outcomes or for multiple population subgroups (Schochet 2009). For example, when applying a 5 percent critical value for hypothesis testing, the likelihood of finding an impact that is statistically significant at the 5 percent level for any given outcome or subgroup simply due to chance is greater than 5 percent unless formal adjustments for multiple hypothesis testing are made. Because we were estimating impacts on four main outcomes (reading and math scores in Years 1 and 2) for the full sample, we applied the procedure described by Benjamini and Hochberg (1995) to adjust for multiple hypothesis testing. For the subgroup analyses, we tested whether differences in impacts across subgroups (for instance, urban and non-urban) were statistically significant and then applied the Benjamini-Hochberg procedure to these differences across the four test score outcomes.[25] Impacts (or differences in impacts) that are statistically significant at the 5 percent level prior to this adjustment are denoted with a cross sign, while impacts (or differences in impacts) that are statistically significant after this adjustment are denoted with an asterisk.

The Average Impact of Study Charter Schools

On average, study charter schools did not have statistically significant impacts on student achievement once the adjustment for multiple hypothesis testing was applied. The treatment group students scored lower on state reading and mathematics assessments than did the control group students, and in the case of Year 2 reading scores the estimated difference of 0.07 standard deviations (or 2.6 percentile points) was statistically significant at the 5 percent level before, but not after, adjustment for multiple hypothesis testing (Table 3).[26] As an alternative measure of student achievement, we also examined impacts on the proportion of students achieving proficiency on their state assessments in reading and math in Years 1 and 2. These results, shown in Appendix Table 4, indicate that there was virtually no difference in the proficiency rates of treatment and control group students.

The estimated average impacts of study charter schools on Year 1 reading scores or math scores in Years 1 or 2 were not statistically significant at the 5 percent level either before or after the adjustment for multiple hypothesis testing. Results are similar in the sensitivity tests we conducted (Tables 4 and 5). Estimates for all specifications are similar in magnitude—indicating charter schools impacts that are negative, ranging from -0.03 to -0.08 standard deviations, and sometimes statistically significant before, but not after, the adjustment for multiple hypothesis

[25] Applying the framework recommended by Schochet (2009), prior to conducting the analysis we designated the impact estimates for the full sample as the study's sole "confirmatory analysis" and the subgroup estimates as "exploratory." Because the subgroup analyses are considered exploratory, we did not adjust for multiple hypothesis testing across all the subgroups examined (only across the four test score outcomes for the differences between two sets of subgroups, such as male versus female). Thus estimates for subgroups are not as rigorous as estimates for the full sample, and are more likely to be spurious.

[26] Ideally we would translate this effect size into test score gains relative to the typical test score gains of the control group sample over the course of the school year. However, this calculation is not possible, as most of the assessments from the study sample were not vertically aligned from year to year. As an alternative, we relied on estimates from Hill et al. (2007), who found that the average annual test score gains across a sample of seven nationally normed tests in grades 5 through 8 were, on average, 0.26 standard deviations in reading and 0.31 standard deviations in math. While these estimates may not be directly relevant to the particular students and assessments in our study, they suggest that the estimated effect on Year 2 reading of -0.07 standard deviations—which is cumulative over the two-year follow-up period—is equal to approximately one-quarter-year less instruction for students in charter schools than what they would have received had they not been admitted.

Table 3. Charter School Impacts on Student Achievement

	Impact of Admission Offer (ITT)				Impact of Attendance (TOT)	
Outcome (z-scores)	Mean, Treatment Group	Mean, Control Group	Difference (Impact Estimate)	p-Value	Adjusted Impact Estimate	p-Value
Reading Achievement						
Year 1	0.40	0.44	-0.04	0.214	-0.06	0.231
Year 2	0.31	0.38	-0.07	0.032†	-0.08	0.117
Math Achievement						
Year 1	0.34	0.39	-0.06	0.061	-0.09	0.072
Year 2	0.32	0.38	-0.06	0.136	-0.08	0.202
Number of Students	1,328	822	2,150		2,141	
Number of Sites			29		29	

Source: Gleason et al. (2010).

Note: Means, impact estimates, and effect sizes are estimated at the site level and averaged across sites. Means for lottery losers are not regression adjusted; means for lottery winners are computed as the unadjusted mean for lottery losers plus the regression-adjusted impact estimate. Test scores were standardized across states by converting to z-scores (raw scores minus the state mean score for that subject and grade, divided by the standard deviation of scores for that subject and grade), and impact estimates represent charter schools' effects on student scores expressed in terms of statewide standard deviations of scores for the student's grade. The Benjamini-Hochberg procedure was used to adjust for multiple hypothesis testing. The sample sizes represent the number of students or sites with nonmissing data for at least one of the outcomes. Sample sizes vary for individual outcomes.

ITT = Intent to treat; TOT = Treatment on treated.

†† Difference between lottery winners and losers is statistically significant at the 0.01 level, two-tailed test.

† Difference between lottery winners and losers is statistically significant at the 0.05 level, two-tailed test.

** Difference between lottery winners and losers is statistically significant at the 0.01 level after adjusting for multiple hypothesis testing, two-tailed test.

* Difference between lottery winners and losers is statistically significant at the 0.05 level after adjusting for multiple hypothesis testing, two-tailed test.

Table 4. Sensitivity of Intent-to-Treat Impact Estimates

			Alternative Model 1		Alternative Model 2		Alternative Model 3	
	Primary Impact Model		No Covariates		GLS Weighting of Site-level Impact Estimates		Include All Sites with Valid Data	
Outcome	Impact Estimate	p-Value	Impact Estimate	p-Value	Impact Estimate	p-Value	Impact Estimate	p-Value
Reading Achievement								
Year 1	-0.04	0.214	-0.04	0.426	-0.04	0.329	-0.05	0.098
Year 2	-0.07	0.032†	-0.05	0.327	-0.07	0.114	-0.08	0.032†
Math Achievement								
Year 1	-0.06	0.061	-0.03	0.585	-0.06	0.092	-0.07	0.025†
Year 2	-0.06	0.136	-0.03	0.570	-0.06	0.380	-0.08	0.049†
Number of Students	2,150		2,150		2,150		2,179	
Number of Sites	29		29		29		31	

Source: Gleason et al. (2010).

†† Difference between lottery winners and losers is statistically significant at the 0.01 level, two-tailed test.

† Difference between lottery winners and losers is statistically significant at the 0.05 level, two-tailed test.

** Difference between lottery winners and losers is statistically significant at the 0.01 level after adjusting for multiple hypothesis testing, two-tailed test.

* Difference between lottery winners and losers is statistically significant at the 0.05 level after adjusting for multiple hypothesis testing, two-tailed test.

Table 5. Sensitivity of Imapct Estimates to Approach for Accounting for Missing Data

	Primary Impact Model		Alternative Model 1: Full Sample with Nonresponse Weights		Alternative Model 2: Bound Potential Impact Estimates			
	ITT Impact Estimate	p-Value	ITT Impact Estimate	p-Value	ITT Estimate Lower Bound	p-Value	ITT Estimate Upper Bound	p-Value
Reading Achievement								
Year 1	-0.04	0.214	-0.05	0.145	-0.15	<0.001††**	0.10	0.001††**
Year 2	-0.07	0.032†	-0.07	0.033†	-0.19	<0.001††**	0.04	0.198
Math Achievement								
Year 1	-0.06	0.061	-0.07	0.025†	-0.16	<0.001††**	0.05	0.086
Year 2	-0.06	0.136	-0.03	0.520	-0.21	<0.001††**	0.07	0.055
Number of Students	2,150		2,069		2,176		2,181	
Number of Sites	29		27		29		29	

Source: Gleason et al. (2010).
Note: The Benjamini-Hochberg procedure was used to adjust for multiple comparisons within this domain. Sample sizes vary for individual outcomes.
ITT = Intent to treat.
†† Difference between lottery winners and losers is statistically significant at the 0.01 level, two-tailed test.
† Difference between lottery winners and losers is statistically significant at the 0.05 level, two-tailed test.
** Difference between lottery winners and losers is statistically significant at the 0.01 level after adjusting for multiple comparisons, two-tailed test.
* Difference between lottery winners and losers is statistically significant at the 0.05 level after adjusting for multiple comparisons, two-tailed test.

testing. The bounds on estimates that include the full student sample with no adjustment for nonresponse indicate that impacts range from large, negative, and statistically significant for all four tests to positive but not statistically significant for all tests except Year 1 reading, on which the upper bound impact estimate of 0.10 was statistically significant (Table 5).

We also estimated impacts for subgroups of students (Table 6).[27] There are no statistically significant differences (or clear pattern of differences) across subgroups defined by students' race (white non-Hispanic vs. non-white or Hispanic) or gender. However, estimated impacts were positive for more disadvantaged students as measured by certification for free or reduced-price lunch, and large and negative for more advantaged students, for reading in Year 1 and math in Years 1 and 2, and these differences were statistically significant after the adjustment for multiple hypothesis testing.[28] These same patterns persisted for Year 2 test scores for subgroups defined by students' baseline achievement in reading or math (defined by whether the student scored above or below the sample median on the respective test), although differences were not statistically significant after adjustment for multiple hypothesis testing. The more positive impacts for more disadvantaged students could reflect the fact that the study charter schools tended to be more effective for more disadvantaged students, or could reflect the fact that, within each site, the alternative educational opportunities available to the more disadvantaged control group students were less effective than those available to the more advantaged control group students, an issue we explore further below.

[27] Impacts for each subgroup include only sites that met the sample size and response rate criteria described above for that particular subgroup. Thus differences between subgroups may reflect differences in impacts for particular types of students or differences in impacts of sites that serve particular types of students.
[28] This pattern of impacts is not simply a function of the particular charter schools attended by large numbers of disadvantaged students in the sample, as subgroup estimates were computed in each site and the overall estimate for each subgroup was computed as an equally weighted average of the site-level estimates. Thus they suggest that, on average, the charter schools in the study had more positive impacts for more disadvantaged students than for more advantaged students.

Table 6. Impacts for Subgroups of Students

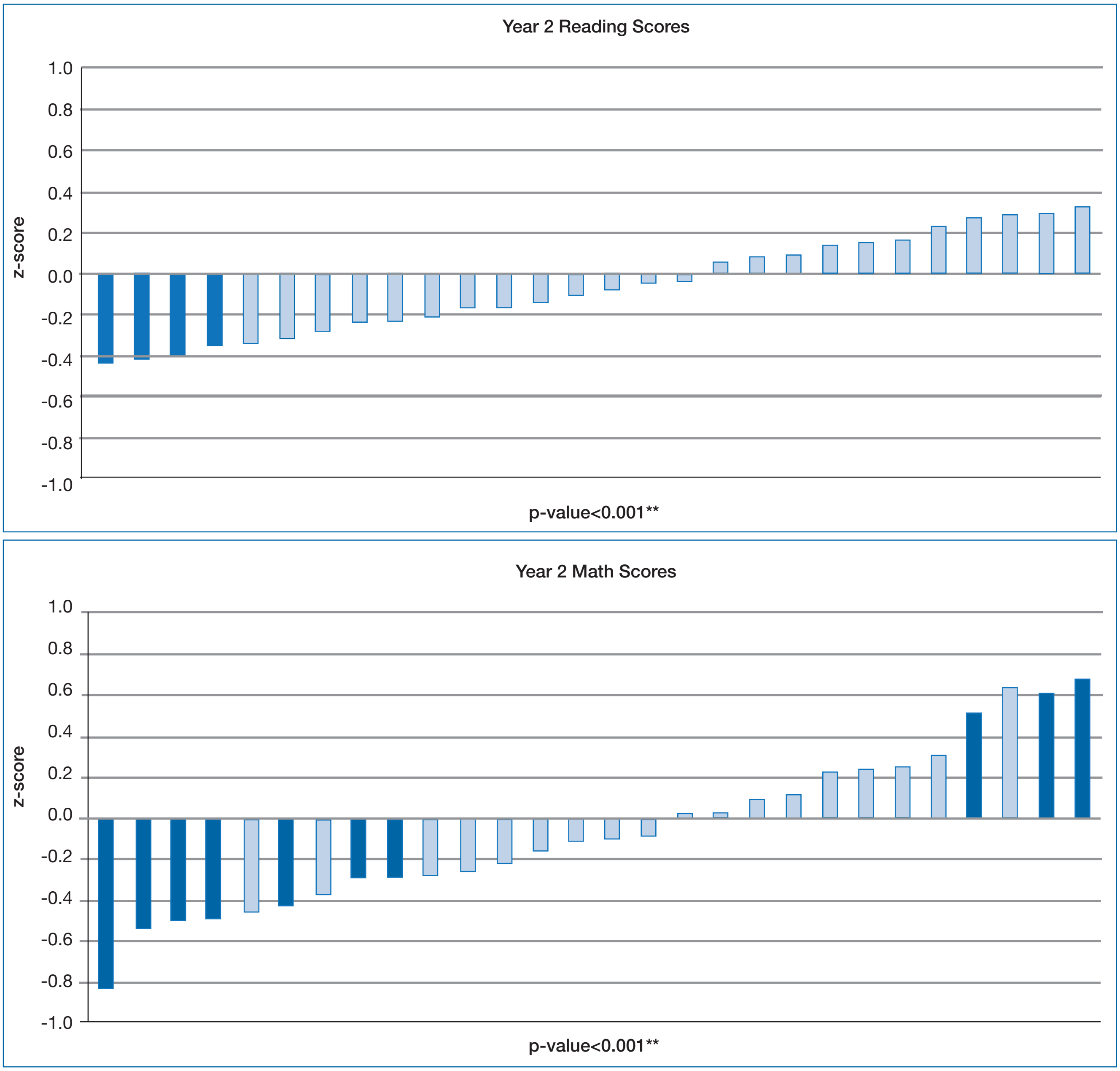
	ITT Impact Estimate	p-Value	ITT Impact Estimate	p-Value	Difference in Impact Estimates	p-Value
	Nonwhite and/or Hispanic		**White, non-Hispanic**		**Difference Between Subgroups**	
Reading Achievement						
Year 1	-0.03	0.632	0.02	0.672	0.05	0.525
Year 2	-0.08	0.220	-0.07	0.150	0.02	0.837
Math Achievement						
Year 1	0.01	0.890	-0.09	0.033†	-0.09	0.171
Year 2	-0.03	0.706	-0.09	0.147	-0.05	0.608
Number of Students	994		1,106			
Number of Sites	22		23			
	Female		**Male**		**Difference Between Subgroups**	
Reading Achievement						
Year 1	0.01	0.808	-0.02	0.642	-0.03	0.614
Year 2	-0.08	0.055	0.03	0.548	0.11	0.086
Math Achievement						
Year 1	-0.04	0.285	-0.03	0.480	0.01	0.841
Year 2	-0.09	0.066	0.02	0.760	0.11	0.151
Number of Students	1,098		1,003			
Number of Sites	28		27			
	Certified for Free or Reduced-Price Lunch		**Not Certified for Free or Reduced-Price Lunch**		**Difference Between Subgroups**	
Reading Achievement						
Year 1	-0.07	0.272	-0.02	0.565	0.04	0.584
Year 2	0.05	0.416	-0.12	0.002††**	-0.17	0.018†*
Math Achievement						
Year 1	0.06	0.248	-0.14	<0.001††**	-0.20	0.002††**
Year 2	0.17	0.003††*	-0.14	0.013†*	-0.31	<0.001††**
Number of Students	770		1,333			
Number of Sites	19		28			
	Baseline Reading Achievement Below Median		**Baseline Reading Achievement Above Median**		**Difference Between Subgroups**	
Reading Achievement						
Year 1	-0.08	0.089	-0.04	0.434	0.04	0.536
Year 2	-0.02	0.655	-0.13	0.007††*	-0.11	0.117
Math Achievement						
Year 1	-0.03	0.482	-0.03	0.539	0.00	0.983
Year 2	0.05	0.337	-0.11	0.057	-0.16	0.036†
Number of Students	1,077		1,019			
Number of Sites	26		26			
	Baseline Math Achievement Below Median		**Baseline Math Achievement Above Median**		**Difference Between Subgroups**	
Reading Achievement						
Year 1	-0.01	0.874	-0.08	0.075	-0.07	0.253
Year 2	0.01	0.747	-0.10	0.026†	-0.12	0.064
Math Achievement						
Year 1	-0.05	0.252	-0.03	0.478	0.02	0.778
Year 2	0.08	0.124	-0.10	0.063	-0.18	0.016†
Number of Students	983		1,068			
Number of Sites	26		27			

Source: Gleason et al. (2010).
†† Difference between lottery winners and losers is statistically significant at the 0.01 level, two-tailed test.
† Difference between lottery winners and losers is statistically significant at the 0.05 level, two-tailed test.
** Difference between lottery winners and losers is statistically significant at the 0.01 level after adjusting for multiple hypothesis testing, two-tailed test.
* Difference between lottery winners and losers is statistically significant at the 0.05 level after adjusting for multiple hypothesis testing, two-tailed test.

Exploring Variation in Charter School Impacts

While the overall average impacts of the study charter schools were negative and not statistically significant after the adjustment for multiple hypothesis testing, estimates varied across sites. Figure 1 presents the distribution of estimated impacts on Year 2 reading and

Figure 1. Distribution of Site-Level Impact Estimates



Source: Gleason et al. (2010).
Note: p-values are from tests of homogeneity of impacts.
*Variation in impacts is statistically significant at the 0.05 level, two-tailed test.
**Variation in impacts is statistically significant at the 0.01 level, two-tailed test.
Shaded bars are statistically significant impacts at the 0.05 level, two-tailed test.

mathematics scores across study charter school sites, arranged by magnitude of impact. The figure shows substantial variation in the impacts. Impacts on Year 2 reading z-scores ranged from -0.43 to +0.33, with a standard deviation of 0.24. Four estimated impacts were statistically significant and negative, with the remainder not significantly different from zero. Impacts on Year 2 mathematics z-scores ranged from -0.78 to +0.65, with a standard deviation of 0.36. Nine of the site-level estimated impacts were statistically significant, including six negative and three positive impacts, with the remainder not significant. While we would expect some variation in impact estimates across sites due to chance, the observed variation is much larger than would be expected due to chance alone. A Q-test for the homogeneity of impacts (Lipsey and Wilson 2001) strongly rejects the null hypothesis that study charter school impact is constant across sites (p-value<0.001 for both Year 2 test score outcomes).

To further investigate the circumstances under which charter schools are more or less effective relative to nearby public schools, we estimated impacts for several subgroups of charter schools in the sample (Table 7). Consistent with the findings for the student subgroup analysis, these results show that charter schools serving a high proportion of students certified for free or reduced-price lunch have a positive impact on Year 2 math achievement, while charter schools serving a low proportion of these students have negative impacts on math and reading achievement. Differences across these subgroups were statistically significant for math achievement in Years 1 and 2 after adjustment for multiple hypothesis testing. Similarly, schools serving high proportions of students with low baseline achievement have more positive impacts than those serving a lower proportion of these students—these differences were statistically significant for reading and math in Year 2 after adjustment for multiple hypothesis testing. Impacts for schools in urban areas are similar to those in nonurban areas for reading achievement, but positive and significant for schools in urban areas (in Year 2) and negative and significant for schools in non-urban areas for math achievement. These differences were statistically significant after adjustment for multiple hypothesis testing for math achievement in Years 1 and 2.

Given the large variation in impacts across charter schools in the study sample, a better understanding of the specific factors that influence charter school impacts is important for policymakers and practitioners interested in these institutions. To further explore these factors, we focused on the difference in impacts between urban and non-urban schools in our sample, primarily because the finding of large positive impacts in urban schools and insignificant or negative impacts in non-urban schools has been a central finding of the previous literature.

In exploring the variation in impacts across sites, it is important to keep in mind the distinction between *impacts* and *effectiveness*. An impact for a particular study charter school reflects how that school influenced the achievement of its students relative to the counterfactual for those students—the schools they would have attended had the study charter school not been available. In contrast, we use the term effectiveness to reflect how well a particular set of students would perform at one school relative to some other school—for instance, School A is more effective than School B for a particular group of students if this group of students would perform better at School A than School B. Thus, the fact that impacts are greater in one charter school site in our study than another does not necessarily imply that the charter school in the former site is more effective than the charter school in the latter site. Rather, the difference in impacts could have been driven by differences in the counterfactual schools in the two sites, or by differences in the student populations served in the two sites.

Table 7. Impacts for Subgroups of Sites

	ITT Impact Estimate	p-Value	ITT Impact Estimate	p-Value	Difference in Impact Estimates	p-Value
	High Percentage Eligible for Free or Reduced-Price School Meals		Low Percentage Eligible for Free or Reduced-Price School Meals		Difference Between Subgroups	
Reading Achievement						
Year 1	-0.07	0.128	-0.02	0.703	0.05	0.430
Year 2	0.00	0.965	-0.11	0.010†*	-0.11	0.076
Math Achievement						
Year 1	0.03	0.540	-0.11	0.006††*	-0.14	0.019†*
Year 2	0.18	0.002††**	-0.24	<0.001††**	-0.41	<0.001††**
Number of Students	1,141		1,006			
Number of Sites	13		16			
	Average Baseline Reading Achievement in Site Below Median		Average Baseline Reading Achievement in Site Above Median		Difference Between Subgroups	
Reading Achievement						
Year 1	0.00	0.917	-0.08	0.114	-0.07	0.302
Year 2	0.03	0.544	-0.15	0.001††**	-0.17	0.006††*
Math Achievement						
Year 1	0.01	0.890	-0.09	0.019†*	-0.10	0.090
Year 2	0.12	0.033†	-0.21	<0.001††**	-0.33	<0.001††**
Number of Students	1,093		1,057			
Number of Sites	14		15			
	Average Baseline Math Achievement in Site Below Median		Average Baseline Math Achievement in Site Above Median		Difference Between Subgroups	
Reading Achievement						
Year 1	0.01	0.827	-0.09	0.058	-0.10	0.136
Year 2	0.08	0.051	-0.20	<0.001††**	-0.29	<0.001††**
Math Achievement						
Year 1	-0.03	0.548	-0.07	0.090	-0.04	0.456
Year 2	0.16	0.006††*	-0.25	<0.001††**	-0.40	<0.001††**
Number of Students	1,004		1,146			
Number of Sites	14		15			
	Low Percentage White		High Percentage White		Difference Between Subgroups	
Reading Achievement						
Year 1	-0.10	0.021†	0.02	0.737	0.11	0.089
Year 2	-0.08	0.033†	-0.03	0.516	0.05	0.396
Math Achievement						
Year 1	-0.05	0.183	-0.04	0.336	0.01	0.826
Year 2	0.01	0.820	-0.11	0.077	-0.12	0.131
Number of Students	1,309		841			
Number of Sites	13		14			
	Urban		Not Urban		Difference Between Subgroups	
Reading Achievement						
Year 1	-0.04	0.393	-0.04	0.340	0.00	0.944
Year 2	-0.02	0.709	-0.08	0.041†	-0.06	0.366
Math Achievement						
Year 1	0.06	0.265	-0.10	0.004††**	-0.16	0.015†*
Year 2	0.16	0.033†	-0.14	0.003††**	-0.30	0.001††**
Number of Students	678		1,472			
Number of Sites	9		20			

Source: Gleason et al. (2010).
†† Difference between lottery winners and losers is statistically significant at the 0.01 level, two-tailed test.
† Difference between lottery winners and losers is statistically significant at the 0.05 level, two-tailed test.
** Difference between lottery winners and losers is statistically significant at the 0.01 level after adjusting for multiple hypothesis testing, two-tailed test.
* Difference between lottery winners and losers is statistically significant at the 0.05 level after adjusting for multiple hypothesis testing, two-tailed test.

We considered three possible explanations (not necessarily mutually exclusive) for the differences in impacts between the urban and non-urban schools in our sample:

1. *Charter schools have more positive impacts for the types of students served by urban schools than for the types of students served by non-urban schools.* As shown in the student subgroup analysis, the study charter schools, on average, had positive impacts for more disadvantaged students, and negative impacts for more advantaged students, and disadvantaged students comprised a higher percentage of the total student sample in the urban schools (56 percent) than in the non-urban schools (26 percent). Thus the more positive impacts of urban schools might be explained solely by the populations they serve. Charter schools, regardless of their location, may be particularly well equipped to meet the needs of more disadvantaged students, and the fact that urban charter schools serve more disadvantaged populations than non-urban charters could explain why they have more positive impacts.

2. *The urban charter schools in the study are more effective than the non-urban charter schools.* That is, a given population of students would realize greater gains in student achievement at an urban school in our sample than at a non-urban school. (Hypothetically, if all the sample students were randomly assigned between the urban and non-urban charter schools in the study, those assigned to the urban schools would score higher on average at the end of the school year than those assigned to the non-urban schools.) The greater effectiveness could be due to more effective policies and practices at urban schools, more effective teaching, a more positive influence of other students in the school, or some other factor.

3. *The alternative schools available to charter applicants in the urban study sites are less effective than those in the non-urban sites.* Impacts are estimated as the difference in test scores between treatment and control group students in each site. The greater impacts in the urban schools in our sample may reflect the fact that the schools attended by control group students ("the comparison schools") in the urban sites were less effective than the schools attended by control group members in the non-urban sites. That is, a given population of students would fare better in the comparison schools in non-urban sites than in the comparison schools in urban sites. (Or hypothetically, if all the students in our sample were randomly assigned between urban and non-urban comparison schools, those assigned to the non-urban comparison schools would score higher on average at the end of the school year than those assigned to the urban comparison schools.) The lesser effectiveness of the urban comparison schools could be due to less effective policies and practices at urban schools, less effective teaching, a less positive influence of other students in the school, or some other factor.

These three explanations cannot be explored with the rigor of the impact estimation. Students were not randomly assigned to study sites, and thus it is not possible to definitively disentangle whether differences in impacts are due to differences in charter school effectiveness, differences in characteristics of students served, or differences in effectiveness of the comparison schools in each site. Nonetheless, we conducted exploratory analyses to attempt to further investigate these hypotheses.

To investigate whether the differences in impacts across urban and non-urban charter schools might be solely due to differences in the populations served, we estimated impacts in the non-urban and urban sites for particular subgroups of students, including those certified for free and reduced-price lunch, those not certified for free and reduced-price lunch, white non-Hispanic students, black and/or Hispanic students, students with baseline achievement below the sample median, and students with baseline achievement above the sample median (Table 8). Of course, these broad student subgroups are unlikely to fully capture differences in the populations served in each site—for instance, even among students certified for free and reduced-price lunch, students in the urban sites may be more disadvantaged than those in non-urban sites. Nonetheless, this approach can reveal whether impacts are greater in the urban than non-urban sites for specific subgroups of students. Sample sizes for this analysis are small, but the same pattern of more positive (or less negative) impacts in urban sites persists across all subgroups. These results suggest that the differences in impacts across urban and non-urban sites are not solely due to differences in populations served.

It is more difficult to investigate the second and third hypotheses. We cannot directly examine the effectiveness of the schools attended by treatment group students in urban sites versus non-urban sites. Comparing mean test scores or proficiency rates among all students in the urban versus non-urban charter schools in the study would not allow us to disentangle characteristics of the student population served from the quality of the school—a school serving very disadvantaged students may have low baseline scores even if it is highly effective for that population. For the same reason, we cannot determine whether comparison schools are less effective in the study's urban versus non-urban sites. Nonetheless, to explore these two hypotheses we examined the correlation of site-level impact estimates with treatment and control group mean scores. If the more positive impacts in urban sites are driven by effects of the study charter schools alone, one might expect a positive correlation between impacts and mean test scores in treatment schools. If the more positive impacts in urban sites are driven by what was happening in the control schools (that is, the traditional public schools surrounding the study's charter schools), then one would expect a negative correlation between impacts and mean test scores in control schools.

As shown in Table 9, the correlations of impacts and control group mean scores were negative for all four outcomes examined, with correlations for Year 1 and Year 2 reading and Year 2 math control group means and associated impact estimates close to -0.40 and statistically significant prior to the adjustment for multiple hypothesis testing. In contrast, correlations between site-level impact estimates and treatment group mean scores were all positive, but smaller and not statistically significant. Because the student populations, as well as the tests themselves, varied across sites, these results are far from definitive. But they are consistent with the notion that variation in charter school impacts may be driven by the opportunities available to the control group students in that site rather than by the effectiveness of the study charter schools relative to one another.

Taken together, the facts that (1) on average the study charter schools have more positive impacts for more disadvantaged students, (2) that even among subgroups of students, urban charter schools have more positive impacts than non-urban schools, and (3) that impact estimates are negatively and significantly correlated with control group mean scores, but not significantly correlated with treatment group mean scores, suggest that differences in the opportunities available to control group students may play a role in the more positive impacts of the urban charter schools in the study and in the variation in site-level impact estimates more generally. For instance, the schools available to control group students may be less effective in the urban sites than in the non-urban sites. Within sites, more disadvan-

Table 8. Impacts Across Urban and Non-Urban Sites for Subgroups of Students

	Urban		Not Urban		Difference Between Subgroups	
	ITT Impact Estimate	p-Value	ITT Impact Estimate	p-Value	Difference in Impact Estimates	p-Value of Difference
Students Certified for Free or Reduced-Price Lunch						
Reading Achievement						
Year 1	-0.03	0.708	-0.08	0.334	-0.05	0.664
Year 2	0.07	0.407	0.06	0.455	-0.01	0.959
Math Achievement						
Year 1	0.23	0.003††**	-0.03	0.682	-0.26	0.014†
Year 2	0.34	0.000††**	0.11	0.160	-0.23	0.064
Number of Students[a]	380		390			
Number of Sites[a]	10		10			
Students Not Certified for Free or Reduced-Price Lunch						
Reading Achievement						
Year 1	-0.07	0.366	0.00	0.992	0.07	0.441
Year 2	-0.08	0.331	-0.16	0.001††**	-0.08	0.423
Math Achievement						
Year 1	-0.14	0.084	-0.15	0.000††**	-0.01	0.889
Year 2	0.08	0.523	-0.24	0.000††**	-0.31	0.020†
Number of Students[a]	270		1,060			
Number of Sites[a]	10		20			
Students with Low Baseline Achievement in Reading and Math						
Reading Achievement						
Year 1	-0.11	0.081	-0.04	0.440	0.07	0.411
Year 2	-0.08	0.271	`0.02	0.727	0.10	0.301
Math Achievement						
Year 1	0.07	0.335	-0.10	0.024†	-0.17	0.046†
Year 2	0.22	0.025	-0.02	0.832	-0.24	0.059
Number of Students[a]	390		660			
Number of Sites[a]	10		20			
Students with High Baseline Achievement in Reading and Math						
Reading Achievement						
Year 1	0.05	0.559	-0.11	0.054	-0.15	0.119
Year 2	0.07	0.407	-0.25	0.000††**	-0.31	0.001††**
Math Achievement						
Year 1	0.14	0.114	-0.10	0.051	-0.24	0.018†*
Year 2	0.17	0.104	-0.29	0.000††**	-0.47	0.000††**
Number of Students[a]	290		770			
Number of Sites[a]	10		20			

Source: Authors' tabulations based on National Center for Education Statistics. "The Evaluation of Charter School Impacts Restricted Use Data Files." U.S. Department of Education, 2010. (Accessed August 1, 2011). Restricted use data can be requested from the U.S. Department of Education's National Center for Education Statistics through its Electronic Application System, available at http://nces.ed.gov/statprog/instruct.asp.

[a] Reported sample sizes are rounded to the nearest 10 in accordance with NCES publication policy for analyses not previously published in an NCES report.

†† Difference between lottery winners and losers is statistically significant at the 0.01 level, two-tailed test.

† Difference between lottery winners and losers is statistically significant at the 0.05 level, two-tailed test.

** Difference between lottery winners and losers is statistically significant at the 0.01 level after adjusting for multiple hypothesis testing, two-tailed test.

* Difference between lottery winners and losers is statistically significant at the 0.05 level after adjusting for multiple hypothesis testing, two-tailed test.

Table 9. Correlation of Site-Level Impact Estimates With Treatment and Control Group Mean Scores				
	Control Group Mean Scores		Treatment Group Mean Scores	
	Correlation with Impact Estimate	p-Value	Correlation with Impact Estimate	p-Value
Reading Achievement				
Year 1	-0.40	0.032†	0.07	0.715
Year 2	-0.39	0.039†	0.07	0.732
Math Achievement				
Year 1	-0.11	0.584	0.27	0.168
Year 2	-0.41	0.030†	0.18	0.360
Number of Sites[a]	29		29	

Source: Authors' tabulations based on National Center for Education Statistics. "The Evaluation of Charter School Impacts Restricted Use Data Files." U.S. Department of Education, 2010. (Accessed August 1, 2011). Restricted use data can be requested from the U.S. Department of Education's National Center for Education Statistics through its Electronic Application System, available at http://nces.ed.gov/statprog/instruct.asp.

Note: The Benjamini-Hochberg procedure was used to adjust for multiple hypothesis testing. Sample sizes vary for individual outcomes.

[a] The number of sites for which impacts are estimated was reported by Gleason et al. (2010). Thus NCES policy does not require that these sample sizes be rounded to the nearest 10.

†† Correlation is statistically significant at the 0.01 level, two-tailed test.

† Correlation is statistically significant at the 0.05 level, two-tailed test.

** Correlation is statistically significant at the 0.01 level after adjusting for multiple hypothesis testing, two-tailed test.

* Correlation is statistically significant at the 0.05 level after adjusting for multiple hypothesis testing, two-tailed test.

taged control group students may attend less effective schools than more advantaged control group students. Even within the same school, more disadvantaged students may be placed in classes with less effective teachers than more advantaged students.

Conclusions

This paper presents results from the first large-scale, randomized evaluation of charter school impacts, encompassing 36 charter schools in 15 states. We found that, on average, the charter schools in the study had an insignificant or negative impact on student achievement in reading and math. Impacts generally did not vary across subgroups defined by students' race, or gender. However, impacts were insignificant or positive for more disadvantaged students and negative for more advantaged students, and this same pattern persisted across groups defined by baseline test scores. There was also considerable variation in impacts across schools. Those in urban areas or serving more disadvantaged populations had more positive (or less negative) impacts than those in non-urban areas or serving more advantaged populations. These results provide rigorous evidence for the patterns suggested by previous studies, which have estimated negative or insignificant impacts for geographically diverse samples of charter schools, but positive impacts for charter schools in urban areas.

Understanding the reasons why some charter schools in the sample had positive impacts while others had negative impacts is important for those seeking to use charter schools as a tool for improving student achievement. Were the study charter schools more effective for more disadvantaged students? Were the study charter schools in urban areas more effective than those in non-urban areas? Or were the educational opportunities available to control group students weaker in urban areas or for less advantaged students? While it is not possible to definitively investigate these possibilities in our data, our exploratory analyses suggest that the educational opportunities available to control group students may have been weaker for less advantaged students or those in urban areas.

This study is the first lottery-based analysis of charter school impacts to span multiple states and both urban and non-urban areas. Previous lottery-based studies have each focused on charter schools within a limited geographic area, and collectively they cover only the charter schools in a few large, urban areas. Moreover, this study is the first to include careful monitoring of charter school lotteries to ensure that the resulting treatment and control groups were truly randomly determined. Previous lottery-based studies have had to assume that the selection of students to be admitted to charter schools was conducted using a truly random mechanism, and they have had to trust the schools' documentation of the lottery results and subsequent admissions.

It is important to keep in mind that charter schools were not randomly selected for the study, and the resulting sample is thus not nationally representative. The study included only oversubscribed charter schools that held admissions lotteries, and impacts for these schools may differ from impacts of charter schools that are not oversubscribed. Similarly, our finding that the study charter schools in urban areas had more positive (or less negative) impacts than the study charter schools in non-urban areas does not imply that any charter school opened in an urban area will have positive impacts on student achievement—results only apply to the particular set of charter and non-charter schools in our study. Despite these limitations, our findings add significantly to the growing empirical evidence base on this important aspect of educational reform and management.

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: Boston Foundation, January 2009.

Angrist, Joshua D., Susan M. Dynarski, Thomas J. Kane, Parag A. Pathak, and Christopher R. Walters. "Who Benefits from KIPP?" NBER Working Paper Series No. 15740. Cambridge, MA: National Bureau of Economic Research, February 2010.

Ballou, Dale, Bettie Teasley, and Tim Zeidner. "Charter Schools in Idaho." In M. Berends, M.G. Springer, and H.J. Walberg (eds.),. *Charter School Outcomes*. New York: Lawrence Erlbaum Associates, 2008.

Benjamini, Yoav, and Yosef Hochberg. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society Series B* (Methodological), vol. 51, no. 7, 1995, pp. 289-300.

Betts, Julian R., Lorien A. Rice, Andrew C. Zau, Y. Emily Tang, and Cory R. Koedel. "Does School Choice Work? Effects on Student Integration and Achievement." San Francisco: Public Policy Institute of California, August 2006.

Bifulco, Robert, and Helen F. Ladd. "The Impact of Charter Schools on Student Achievement: Evidence from North Carolina." *Education Finance and Policy,* vol. 1, no. 1, 2006, pp. 50-90.

Booker, Toby Kevin, Scott Gilpatric, Timothy Gronberg, and Dennis Jansen. "The Impact of Charter School Attendance on Student Performance." *Journal of Public Economics,* vol. 91, nos. 5-6, June 2007, pp. 849-876.

Center for Research on Education Outcomes (CREDO). "Multiple Choice: Charter School Performance in 16 States." Stanford, CA: Stanford University, June 2009.

Dobbie, Will, and Roland G. Fryer, Jr. "Are High-Quality Schools Enough to Close the Achievement Gap? Evidence from a Bold Social Experiment in Harlem." Unpublished paper. Cambridge, MA: Harvard University, April 2009.

Freedman, David A. "On Regression Adjustments to Experimental Data." *Advances in Applied Mathematics,* vol.40, 2008, pp. 180-193.

Gleason, Philip, Melissa Clark, Christina Clark Tuttle, and Emily Dwoyer. "The Evaluation of Charter School Impacts: Final Report." Princeton, NJ: Mathematica Policy Research, 2010.

Hanushek, Eric A. "Efficient Estimators for Regressing Regression Coefficients." *American Statistician,* vol. 28, no. 2, 1974, pp. 66-67.

Hanushek, Eric A., John F. Kain, Steven G. Rivkin, and Gregory F. Branch. "Charter School Quality and Parental Decision Making with School Choice." *Journal of Public Economics,* vol. 91, 2007, pp. 823-848.

Hill, Carolyn J., Howard S. Bloom, Alison Rebeck Black, and Mark W. Lipsey. "Empirical Benchmarks for Interpreting Effect Sizes in Research." MDRC Working Paper on Research Methodology. New York: MDRC, 2007.

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

Hoxby, Caroline M., and Jonah E. Rockoff. "Findings from the City of Big Shoulders." *Education Next,* vol. 5, no. 4, 2005, pp. 52-58.

Lee, David S. "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects." NBER Working Paper #11721. Cambridge, MA: National Bureau of Economic Research, 2005.

Lipsey, Mark W., and David B. Wilson. *Practical Meta-Analysis.* Thousand Oaks, CA: Sage Publications, 2001.

May, Henry, Irma Perez-Johnson, Joshua Haimson, Samina Sattar, and Phil Gleason. "Using State Tests in Education Experiments: A Discussion of the Issues." Final report. Princeton, NJ: Mathematica Policy Research, November 2009.

McEwan, Patrick, and Robert Olsen. "Admissions Lotteries in Charter Schools." In *Taking Measure of Charter Schools: Better Assessments, Better Policymaking, Better Schools.* Lanham, MD: Rowman, Littlefield Publishing Group, May 2010.

Sass, Tim R. "Charter Schools and Student Achievement in Florida." *Education Finance and Policy,* vol. 1, no. 1, winter 2006, pp. 91-122.

Schochet, Peter Z. "An Approach for Addressing the Multiple Testing Problem in Social Policy Impact Evaluations." *Evaluation Review,* vol. 33, no. 6, 2009, pp. 539-567.

Schochet, Peter Z. "Is Regression Adjustment Supported by the Neyman Model for Causal Inference?" *Journal of Statistical Planning and Inference,* vol. 140, no. 1, January 2010.

Tuttle, Christina Clark, Philip Gleason, and Melissa Clark. "Using Lotteries to Evaluate Schools of Choice: Evidence from a National Study of Charter Schools." *Economics of Education Review,* forthcoming.

Witte, John F., David L. Weimer, Arnold Shober, and Paul Schlomer. "The Performance of Charter Schools in Wisconsin." *Journal of Policy Analysis and Management,* vol. 26, no. 3, 2007, pp. 557-573.

Zimmer, Ron, Brian Gill, and Kevin Booker. "Charter Schools in Eight States: Effects on Achievement, Attainment, Integration, and Competition." Santa Monica, CA: RAND Corporation, 2009.

Appendix

Appendix Table1. Baseline Characteristics of Treatment and Control Group Students in Full Sample *(Proportions Unless Otherwise Indicated)*

	Mean, Treatment Group	Mean, Control Group	Difference	p-Value of Difference
Achievement (z-score units)				
Baseline reading score	0.42	0.43	-0.01	0.796
Pre-baseline reading score	0.46	0.41	0.06	0.379
Baseline math score	0.45	0.45	0.00	0.997
Pre-baseline math score	0.48	0.36	0.12	0.078
Disciplinary Measures				
Number of days absent in baseline school year	5.99	5.80	0.19	0.517
Student suspended in baseline school year	0.03	0.03	0.00	0.815
Demographic Characteristics				
White, Non-Hispanic[a]	0.60	0.57	0.04	0.053
Black, Non-Hispanic[a]	0.11	0.10	0.01	0.584
Other race, Non-Hispanic[a]	0.05	0.07	-0.02	0.085
Hispanic	0.26	0.28	-0.02	0.278
Male	0.47	0.47	0.00	0.959
Age at start of school year (years)	11.54	11.52	0.02	0.323
Has individualized education plan (IEP)	0.17	0.16	0.01	0.567
Limited English proficiency/ELL	0.10	0.08	0.02	0.080
Family Characteristics				
Income to poverty ratio 0 to 100 percent	0.12	0.11	0.01	0.573
Income to poverty ratio 100 to 200 percent	0.20	0.19	0.01	0.467
Income to poverty ratio 200 to 300 percent	0.18	0.16	0.02	0.184
Income to poverty ratio >300 percent	0.50	0.54	-0.05	0.032†
Two-parent family	0.78	0.78	0.01	0.781
Not two-parent family, but more than one adult	0.05	0.04	0.01	0.324
English main language spoken at home	0.89	0.90	-0.01	0.607
Mother's education: high school or less	0.22	0.22	0.00	0.955
Mother's education: some college	0.34	0.34	-0.01	0.792
Mother's education: college	0.44	0.44	0.00	0.822
Born in U.S.	0.93	0.92	0.00	0.899
Family received TANF or food stamps in past 12 months	0.05	0.05	0.00	0.924
Free or reduced-price lunch-eligible	0.33	0.33	0.00	0.980
School Enrollment				
Enrolled in charter school at baseline	0.04	0.06	-0.02	0.129
Enrolled in private school at baseline	0.08	0.10	-0.02	0.115
Enrolled in public school at baseline	0.86	0.83	0.03	0.080
Home schooled at baseline	0.02	0.01	0.01	0.145
Baseline school type unknown	0.01	0.01	0.00	0.859
Number of Students[b]	1,698	1,144		
Number of Sites	29	29		

Source: Gleason et al. (2010).
Note: Sample includes students in full analysis sample (whether or not they have baseline test score data) in the sites included in the main impact analyses.
[a] Race categories are mutually exclusive.
[b] Sample size differs for some of the individual baseline characteristics due to differential rates of missing data for different characteristics.
†† Difference between treatment and control group students significantly different from zero at 0.01 level, two-tailed test.
† Difference between treatment and control group students significantly different from zero at 0.05 level, two-tailed test.

Appendix Table 2. Baseline Characteristics of Treatment and Control Group Students Included in Analysis of Year 2 Test Score Data *(Proportions Unless Otherwise Indicated)*

	Mean, Treatment Group	Mean, Control Group	Difference	p-Value of Difference
Achievement (z-score units)				
Baseline reading score	0.40	0.38	0.02	0.598
Pre-baseline reading score	0.43	0.36	0.08	0.270
Baseline math score	0.41	0.39	0.02	0.645
Pre-baseline math score	0.43	0.32	0.11	0.132
Disciplinary Measures				
Number of days absent in baseline school year	6.03	5.26	0.77	0.018†
Student suspended in baseline school year	0.04	0.03	0.00	0.895
Demographic Characteristics				
White, Non-Hispanic[a]	0.58	0.54	0.04	0.067
Black, Non-Hispanic[a]	0.12	0.11	0.01	0.754
Other race, Non-Hispanic[a]	0.05	0.07	-0.02	0.180
Hispanic	0.28	0.30	-0.02	0.370
Male	0.46	0.48	-0.01	0.654
Age at start of school year (years)	11.58	11.56	0.02	0.490
Has Individualized Education Plan (IEP)	0.19	0.15	0.03	0.177
Limited English proficiency/ELL	0.11	0.08	0.03	0.070
Family Characteristics				
Income to poverty ratio 0 to 100 percent	0.12	0.13	-0.01	0.705
Income to poverty ratio 100 to 200 percent	0.22	0.21	0.01	0.548
Income to poverty ratio 200 to 300 percent	0.19	0.15	0.04	0.077
Income to poverty ratio >300 percent	0.47	0.51	-0.04	0.079
Two-parent family	0.78	0.78	0.00	0.902
Not two-parent family, but more than one adult	0.05	0.03	0.01	0.247
English main language spoken at home	0.89	0.90	0.00	0.915
Mother’s education: high school or less	0.23	0.26	-0.03	0.241
Mother’s education: some college	0.37	0.36	0.01	0.744
Mother’s education: college	0.41	0.39	0.02	0.490
Born in U.S.	0.93	0.92	0.00	0.850
Family received TANF or food stamps in past 12 months	0.05	0.05	0.00	0.710
Free or reduced-price lunch-eligible	0.36	0.37	-0.02	0.481
School Enrollment				
Enrolled in charter school at baseline	0.04	0.05	-0.01	0.288
Enrolled in private school at baseline	0.00	0.01	0.00	0.319
Enrolled in public school at baseline	0.95	0.94	0.01	0.356
Home schooled at baseline	0.01	0.00	0.01	0.279
Baseline school type unknown	0.00	0.00	0.00	0.441
Number of Students[b]	1,174	752		
Number of Sites	28	28		

Source: Gleason et al. (2010).

Note: Sample includes students in main sample for the analysis of impacts on Year 2 test scores (students with nonmissing baseline test score data and nonmissing second Year 2 test score data) in the sites included in this analysis.

[a] Race categories are mutually exclusive.

[b] Sample size differs for some of the individual baseline characteristics due to differential rates of missing data for different characteristics.

†† Difference between treatment and control group students significantly different from zero at 0.01 level, two-tailed test.

† Difference between treatment and control group students significantly different from zero at 0.05 level, two-tailed test.

Appendix Table 3. Covariates Included in Impact Aanalysis Models								
	Lottery Winners		Lottery Losers				Number of Observations[a]	
	Mean	Standard Deviation	Mean	Standard Deviation	Difference in Means	p-Value of Difference	Lottery Winners	Lottery Losers
Reading Achievement								
Baseline reading score (z-score units)	0.42	0.97	0.43	0.94	-0.01	0.796	1,381	924
Baseline reading proficiency—"high"	0.29	0.46	0.28	0.46	0.01	0.682	1,378	917
Baseline reading proficiency—"medium" or "high"	0.83	0.38	0.84	0.38	-0.01	0.592	1,378	917
Pre-baseline reading score (z-score units)	0.47	1.01	0.38	1.01	0.09	0.175	720	417
Pre-baseline reading proficiency—"high"	0.33	0.48	0.29	0.47	0.04	0.081	1,054	639
Pre-baseline reading proficiency—"medium" or "high"	0.83	0.41	0.82	0.41	0.02	0.493	1,054	639
Math Achievement								
Baseline math score (z-score units)	0.45	0.99	0.45	1.03	0.00	0.997	1,397	927
Baseline math proficiency—"high"	0.33	0.48	0.32	0.47	0.01	0.556	1,395	921
Baseline math proficiency—"medium" or "high"	0.78	0.42	0.76	0.43	0.01	0.467	1,395	921
Pre-baseline math score (z-score units)	0.47	1.02	0.32	1.08	0.15	0.03†	725	417
Pre-baseline math proficiency—"high"	0.31	0.47	0.29	0.47	0.02	0.419	1,044	607
Pre-baseline math proficiency—"medium" or "high"	0.81	0.40	0.75	0.43	0.06	0.011†	1,044	607
Disciplinary Measures								
Number of days absent in baseline school year	6.07	6.20	5.62	6.20	0.46	0.123	1,329	895
Student suspended in baseline school year	0.04	0.19	0.03	0.17	0.01	0.539	1,329	895
Demographic Characteristics								
White[b]	0.81	0.40	0.79	0.41	0.02	0.408	1,295	838
Black[b]	0.13	0.34	0.12	0.33	0.00	0.762	1,295	838
Other race[b]	0.10	0.31	0.11	0.32	-0.01	0.461	1,295	838
Hispanic[b]	0.27	0.45	0.28	0.46	-0.02	0.373	1,332	863
Male	0.46	0.51	0.48	0.51	-0.01	0.59	1,400	930
Age at start of school year	11.53	0.77	11.52	0.75	0.01	0.552	1,400	930
Young for grade	0.01	0.07	0.01	0.09	0.00	0.473	1,400	930
Old for grade	0.09	0.29	0.09	0.29	0.00	0.975	1,400	930
IEP status	0.18	0.39	0.16	0.38	0.02	0.56	1,104	789
Limited English Proficiency/ELL	0.10	0.31	0.08	0.28	0.02	0.095	1,334	894
Family Characteristics								
Income to poverty ratio 0 to 100 percentc	0.13	0.34	0.12	0.33	0.01	0.475	1,230	789
Income to poverty ratio 100 to 200 percent	0.21	0.41	0.19	0.40	0.02	0.362	1,230	789
Income to poverty ratio 200 to 300 percent	0.18	0.39	0.16	0.37	0.02	0.319	1,230	789
Income to poverty ratio >300 percent	0.49	0.51	0.54	0.51	-0.05	0.033†	1,230	789
Two-parent family	0.78	0.42	0.79	0.42	-0.01	0.704	1,293	837

(continued)

Appendix Table 3. Covariates Included in Impact Aanalysis Models *(continued)*

	Lottery Winners		Lottery Losers				Number of Observations[a]	
	Mean	Standard Deviation	Mean	Standard Deviation	Difference in Means	p-Value of Difference	Lottery Winners	Lottery Losers
Not two-parent family, but more than one adult	0.05	0.22	0.04	0.19	0.01	0.26	1,293	837
English main language spoken at home	0.89	0.32	0.90	0.31	-0.01	0.577	1,293	837
Mother's education: high school or less[c]	0.23	0.43	0.24	0.43	-0.01	0.755	1,331	867
Mother's education: some college	0.35	0.49	0.35	0.49	0.00	0.867	1,331	867
Mother's education: college	0.42	0.50	0.42	0.50	0.00	0.924	1,331	867
Born in U.S.	0.92	0.27	0.92	0.27	0.00	0.895	1,185	738
Family received TANF or food stamps in past 12 months	0.05	0.22	0.05	0.23	0.00	0.961	1,291	836
Free or reduced-price lunch-eligible	0.34	0.48	0.35	0.48	0.00	0.844	1,292	878
One child in household[c]	0.23	0.43	0.22	0.42	0.00	0.888	1,321	863
Two children in household	0.47	0.51	0.45	0.51	0.02	0.463	1,321	863
Three or more children in household	0.30	0.47	0.33	0.48	-0.02	0.354	1,321	863
School Enrollment								
Enrolled in charter school at baseline	0.05	0.21	0.06	0.24	-0.01	0.267	1,400	930
Enrolled in private school at baseline	0.00	0.07	0.01	0.09	0.00	0.254	1,398	929
Enrolled in public school at baseline	0.94	0.24	0.93	0.26	0.01	0.352	1,398	929
Changed schools midyear in baseline school	0.01	0.10	0.01	0.12	0.00	0.526	1,344	898
School Applications								
Applied to other charter school at baseline	0.20	0.41	0.19	0.40	0.00	0.857	1,257	802
Applied to private school at baseline	0.07	0.26	0.09	0.29	-0.02	0.182	1,148	739
Applied to other public school at baseline	0.19	0.40	0.21	0.42	-0.02	0.261	1,148	739
Other Information About Sample								
Baseline information form collected before lottery	0.43	0.50	0.45	0.51	-0.02	0.296	1,288	834
Student in cohort 2	0.48	0.51	0.48	0.51	0.00	0.81	1,400	930
Imputation Indicators								
Baseline reading score	0.01	0.10	0.01	0.08	0.00	0.262	1,400	930
Baseline math score	0.00	0.04	0.00	0.05	0.00	0.527	1,400	930
Baseline reading proficiency	0.02	0.12	0.01	0.11	0.00	0.552	1,400	930
Baseline math proficiency	0.01	0.09	0.01	0.09	0.00	0.841	1,400	930
Pre-baseline reading score	0.47	0.51	0.48	0.51	-0.01	0.625	1,400	930
Pre-baseline math score	0.47	0.51	0.48	0.51	-0.01	0.548	1,400	930
Pre-baseline reading proficiency	0.31	0.47	0.30	0.47	0.00	0.805	1,400	930
Pre-baseline math proficiency	0.34	0.48	0.33	0.48	0.00	0.968	1,400	930
Number of days absent in baseline school year	0.05	0.22	0.04	0.20	0.01	0.416	1,400	930
Student suspended in baseline school year	0.05	0.22	0.04	0.20	0.01	0.416	1,400	930
Race	0.07	0.26	0.08	0.28	-0.01	0.401	1,400	930

(continued)

Appendix Table 3. Covariates Included in Impact Aanalysis Models *(continued)*

	Lottery Winners		Lottery Losers				Number of Observations[a]	
	Mean	Standard Deviation	Mean	Standard Deviation	Difference in Means	p-Value of Difference	Lottery Winners	Lottery Losers
Ethnicity	0.05	0.22	0.05	0.23	-0.01	0.583	1,400	930
Gender	0.00	0.06	0.00	0.05	0.00	0.84	1,400	930
IEP status	0.19	0.40	0.19	0.40	0.00	0.876	1,400	930
Limited English Proficiency/ELL	0.04	0.21	0.03	0.18	0.01	0.227	1,400	930
Family structure (two-parent, two-adult, single-parent)	0.07	0.27	0.08	0.27	0.00	0.695	1,400	930
Mother's education	0.05	0.23	0.05	0.22	0.00	0.809	1,400	930
Born in U.S.	0.14	0.35	0.20	0.40	-0.06	0.002††	1,400	930
Family received TANF or food stamps in past 12 months	0.07	0.27	0.08	0.28	-0.01	0.693	1,400	930
Free or reduced-price lunch-eligible	0.06	0.24	0.06	0.25	0.00	0.936	1,400	930
Number of children in household	0.06	0.24	0.06	0.23	0.00	0.873	1,400	930
Type of school attended at baseline	0.00	0.03	0.00	0.04	0.00	0.602	1,400	930
Changed schools midyear in baseline school	0.04	0.19	0.04	0.20	0.00	0.872	1,400	930
Applied to other charter school at baseline	0.11	0.31	0.11	0.32	-0.01	0.616	1,400	930
Applied to private school at baseline	0.17	0.38	0.19	0.40	-0.02	0.302	1,400	930
Applied to other public school at baseline	0.16	0.36	0.17	0.37	-0.01	0.59	1,400	930
Baseline information form collected before lottery	0.08	0.27	0.08	0.28	0.00	0.89	1,400	930

Source: Gleason et al. (2010).

Note: Sample includes students in main analysis sample (students with nonmissing baseline test score data in the 29 sites included in the main impact analyses).

[a] Number of observations excludes imputed values.

[b] Race and ethnicity categories are not mutually exclusive.

[c] Omitted category in regression models.

†† Difference between lottery winners and losers significantly different from zero at 0.01 level, two-tailed test.

† Difference between lottery winners and losers significantly different from zero at 0.05 level, two-tailed test.

Appendix Table 4. Impacts on State Proficiency Levels						
Outcome: Proportion Achieving Proficiency on State Test	Impact of Admission Offer (ITT)				Impact of Attendance (TOT)	
	Mean, Lottery Winners	Mean, Lottery Losers	Difference (Impact Estimate)	p-Value	Adjusted Impact Estimate	p-Value
Reading Achievement—Year 1	0.71	0.72	0.00	0.813	-0.02	0.565
Reading Achievement—Year 2	0.73	0.71	0.01	0.497	0.02	0.649
Math Achievement—Year 1	0.59	0.61	-0.01	0.450	-0.03	0.395
Math Achievement—Year 2	0.60	0.60	0.00	0.861	-0.01	0.705
Number of Students	1,330	820	2,150		2,141	
Number of Sites			29		29	

Source: Gleason et al. (2010).
Note: Means, impact estimates, and effect sizes are estimated at the site-level and averaged across sites. Means for lottery losers are not regression adjusted; means for lottery winners are computed as the unadjusted mean for lottery losers plus the regression-adjusted impact estimate. The Benjamini-Hochberg procedure was used to adjust for multiple hypothesis testing. Sample sizes vary for individual outcomes.
ITT = Intent to treat.
TOT = Treatment on treated.
†† Difference between lottery winners and losers is statistically significant at the 0.01 level, two-tailed test.
† Difference between lottery winners and losers is statistically significant at the 0.05 level, two-tailed test.
** Difference between lottery winners and losers is statistically significant at the 0.01 level after adjusting for multiple hypothesis testing, two-tailed test.
* Difference between lottery winners and losers is statistically significant at the 0.05 level after adjusting for multiple hypothesis testing, two-tailed test.

About the Series

Policymakers require timely, accurate, evidence-based research as soon as it's available. Further, statistical agencies need information about statistical techniques and survey practices that yield valid and reliable data. To meet these needs, Mathematica's working paper series offers policymakers and researchers access to our most current work.

For more information about this study, please contact Melissa Clark, Senior Researcher, at mclark@mathematica-mpr.com.

This study was conducted for the Institute of Education Sciences of the U.S. Department of Education under Contract No. ED-01-C0039/0010. We thank Hanley Chiang, Tom Cook, Brian Gill, Jay Greene, Laura Hamilton, Rick Hanushek, Paul Hill, Helen Ladd, Rebecca Maynard, Robert Meyer, Larry Orr, and Peter Schochet for helpful comments and feedback.

Improving public well-being by conducting high-quality, objective research and surveys

Princeton, NJ • Ann Arbor, MI • Cambridge, MA • Chicago, IL • Oakland, CA • Washington, DC

MATHEMATICA
Policy Research

www.mathematica-mpr.com