MATHEMATICA Policy Research

REPORT

FINAL REPORT

The Equity Project Charter School: Impacts on Student Achievement

October 24, 2014

Joshua Furgeson
Moira McCullough
Clare Wolfendale
Brian Gill

Submitted to:
The Bill & Melinda Gates Foundation
PO Box 23350
Seattle, WA 98102

Project Officer: Steve Cantrell
Grant Number: OPP53404

Submitted by:
Mathematica Policy Research
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044

Project Director: Joshua Furgeson
Reference Number: 06635

ACKNOWLEDGEMENTS

This report reflects the contributions of many individuals not listed as authors.

At TEP, we thank Zeke Vanderhoek and Shelly Gupta. Zeke always made time to talk to us, ensured we had the information we needed, and reviewed a draft for accuracy (he only suggested changes to Chapters I and II). Shelly Gupta, TEP's Director of Finance & Operations, diligently and promptly provided us with lottery records and any other data we requested.

The New York City Department of Education provided us with the student administrative data we needed to conduct the analysis. Special thanks to Charmaine Lester.

At the Bill & Melinda Gates Foundation, we thank our supportive project officer Steve Cantrell.

At Mathematica, Hanley Chiang provided econometric assistance and helpful comments on a draft of the report. Lauren Harris provided able research assistance. John Kennedy superbly edited the report. Laura Sarnoski and Amy Berridge improved our presentation. Allie Clifford provided excellent formatting assistance.

CONTENTS

TABLES

FIGURES

EXECUTIVE SUMMARY

The Equity Project (TEP) charter school's unique approach to concentrating resources and attention on high quality teachers led the *New York Times* to label TEP one of the country's "most closely watched educational experiments." Located in New York City's Washington Heights neighborhood, TEP enrolled its first 5th-grade class during the 2009–2010 school year, and in 2013 that class graduated from TEP's 8th grade. By the 2012–2013 school year, TEP's four grades enrolled about 480 students. This report describes TEP's instructional and personnel strategies, examines the characteristics and attrition rates of TEP students, and measures TEP's impacts on students' achievement during the school's first four years of operation.

TEP recruits and rewards teachers with annual salaries of $125,000, plus weekly professional development and a bonus based on schoolwide performance. TEP receives the standard public allocation provided to New York City charter schools and pays its high teacher salaries by reducing costs elsewhere, having larger classes than typical New York City middle schools (about 31 students compared to about 27) and eliminating administrative positions. TEP teachers receive intensive professional development and have substantial administrative authority and responsibilities. TEP holds teachers accountable for their performance: more than a third of TEP teachers were not rehired after their first year.

TEP's students are similar to those of students in other schools in the neighborhood. TEP admits students using a lottery that favors students in the neighborhood and low-achieving students. More than 90 percent of TEP students are from low-income households (that is, they are eligible for free or reduced-price lunch) and 99 percent are African American or Hispanic. Relative to students who enroll in nearby schools, TEP students have similar 4th-grade test scores, are equally likely to be from low-income households, are equally likely to receive special education services, and are more likely to be Hispanic.

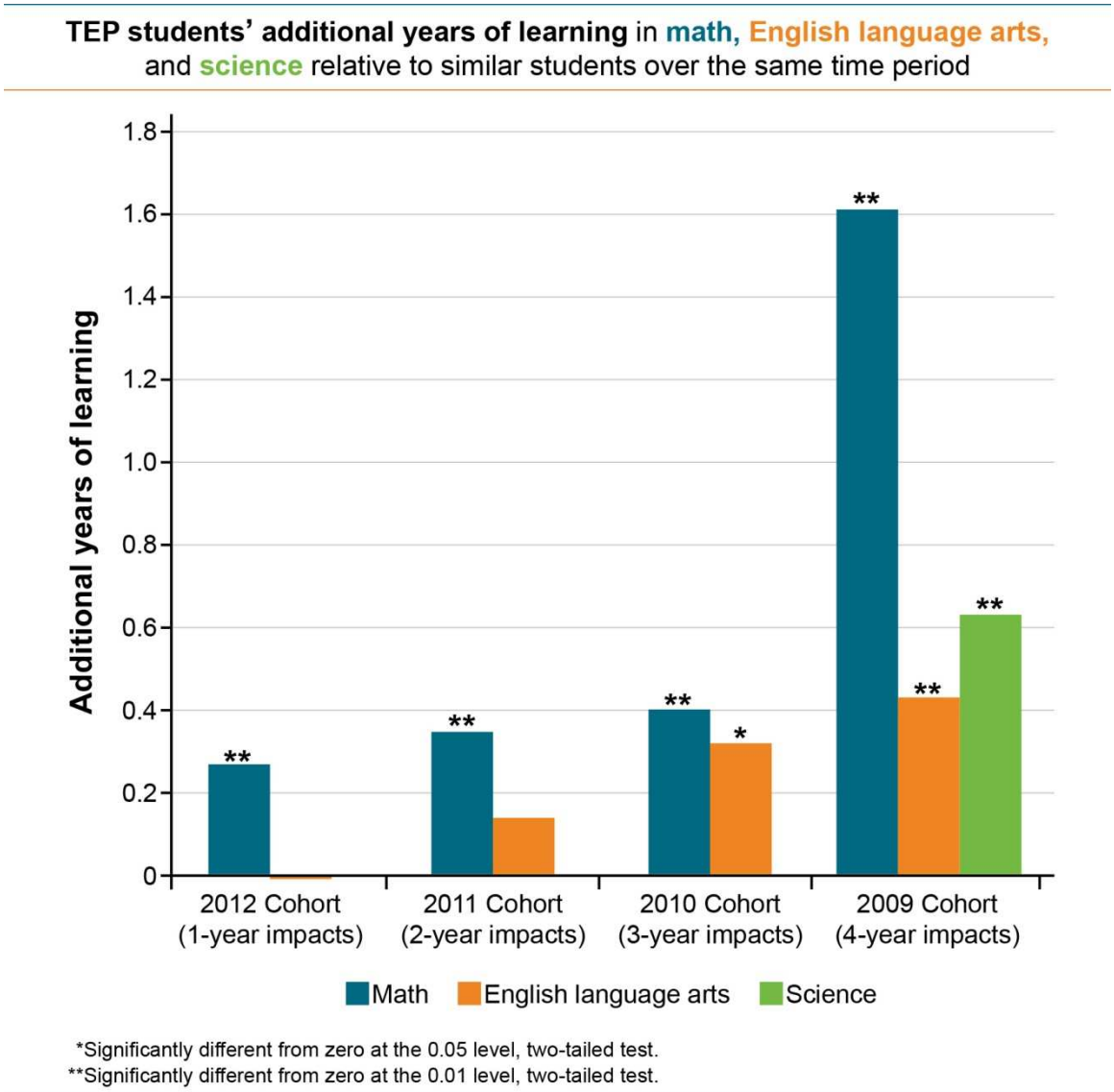
TEP's student attrition rate is similar to other comparable New York City schools. In the school's first four cohorts, the average student attrition rate at TEP was about 5 percent after one year, about 10 percent after two years, about 17 percent after three years, and about 20 percent after four years (the average rates for similar schools were 5, 13, 20, and 23 percent, respectively). During the first four years of operation, TEP did not expel any students and did not use out-of-school suspensions. Students who left TEP were similar to the students who remained.

By the 2012–2013 school year, TEP's impacts on students' achievement were consistently positive. After four years, students who enrolled at TEP had test score gains equal to an additional 1.6 years of school in math and an additional 0.4 years of school in English language arts. For TEP's first two student cohorts (enrolled in 2009 and 2010) during their first two years, TEP's impacts on achievement were largely negative. However, TEP's impacts consistently improved over time for those early cohorts and the later cohorts. By 2012–2013, TEP's cumulative impacts (ranging from one year for the cohort enrolling in 2012 to four years for the cohort enrolling in 2009) on math achievement were positive and statistically significant for all four cohorts of students. In reading, TEP's cumulative impacts by 2012–2013 were significantly positive for the first two cohorts and not statistically significant for the two newer cohorts with one or two years of TEP enrollment. For TEP's first cohort, cumulative

impacts could also be measured in science (science assessments are administered only in 8th grade), for which they were likewise positive and statistically significant.

Using benchmarks for average annual learning gains, TEP students in the first cohort, a full four years after enrollment at TEP, were ahead of their most similar peers in math and English language arts (see the following figure).

Figure E.1. TEP students' additional years of learning



Although this study examines only one school, the positive findings are of broader interest because widespread implementation of the TEP approach could cause systemic changes in teacher quality, a primary goal of current education policy. Using only standard charter school funding, TEP redefines the teacher's role with high salaries, ongoing development, more responsibilities, and accountability. Scaling the approach and creating many professionally and financially rewarding teacher roles might attract stronger applicants to teaching.

I. INTRODUCTION

The Equity Project (TEP) charter middle school's unique approach to concentrating resources and attention on high quality teachers—including a standard teacher salary of $125,000—has made TEP one of the country's "most closely watched educational experiments" (Gootman 2008). Using the standard public allocation provided to New York City (NYC) charter schools, TEP forgoes most administrators and requires teachers to teach larger classes. This comprehensive TEP model is thus potentially scalable and financially sustainable.[1]

Most TEP students are from the Washington Heights neighborhood where TEP is located, a predominantly Latino neighborhood in which about two-thirds of households speak Spanish at home, about 47 percent of households with children are headed by a female with no husband present, about 49 percent of the population is foreign-born, about 30 percent of adults 25 and older have a bachelor's degree or postgraduate degree, and the median family income is $40,671 (2012 dollars).[2] TEP has a student body that is 99 percent African American or Latino, and more than 90 percent of TEP students are eligible for a subsidized lunch. TEP students take classes in portable classrooms located on the campus of a district public school and have access to some of the facilities (such as the auditorium and athletic fields) on the campus.

TEP enrolled its first 5th-grade class during the 2009–2010 school year and that class graduated from TEP's 8th grade in 2013. By the 2012–2013 school year, TEP's four grades enrolled about 480 students.

This report rigorously estimates TEP's impacts on students' achievement during TEP's first four years, long enough to follow the first 5th-grade cohort through 8th-grade graduation. Chapter II describes TEP's practices from the 2009–2010 school year through the 2012–2013 school year. Chapter III reports TEP students' characteristics and the rates at which they leave TEP, and compares them with other NYC students at similar schools. Chapter IV presents our primary estimates of TEP's impact on student achievement in English/language arts (ELA), math, and science, as well as a brief description of the methods we used to estimate impacts (propensity-score matching to identify a similar comparison group of non-TEP students in NYC). Chapter V discusses implications of the findings and issues for future research.

[1] TEP submits tax form 990s to the Internal Revenue Service; the tax forms cover the school year—July 1 to June 30. TEP's submitted 990s were obtained for 2009–2010, 2010–2011, 2011–2012, and 2012-13 from GuideStar. These forms indicate that TEP operating costs are basically covered by recurring grants and per-pupil funding that all charter schools receive. In 2009–2010, government grants (federal grants for special education, Title 1 and Title 2; state grants) were $385,354 and program service revenue (the per-pupil revenue TEP receives from the government) was $1,848,053; total expenses were $2,292,134. In 2010–2011, government grants were $346,248 and program service revenue was $3,929,464; total expenses were $3,935,365. In 2011–2012, government grants were $344,910 and program service revenue was $5,971,718; total expenses were $5,672,577. In 2012-13, government grants were $349,923 and program service revenue was $7,894,311; total expenses were $7,125,889. TEP has received philanthropic contributions only for one-time capital expenses: a school building and an information system.

[2] These data were estimated by NYC using U.S. Census data for 2010–2012. The neighborhood data are for Community District 12 (Washington Heights, Inwood and Marble Hill). Data were accessed from http://www.nyc.gov/html/dcp/html/neigh_info/mn12_info.shtml on August 19, 2014.

II. TEP PRACTICES: 2009–2010 TO 2012–2013

In this chapter, we describe TEP's practices from 2009–2010 through the 2012–2013 school year, the period for which we have data to analyze TEP's effects on students' achievement. We also describe how TEP's practices changed during that time as the school matured.[3] We obtained the information on TEP practices from three sources: (1) documents or data provided by TEP, (2) telephone interviews with the TEP principal in 2014, and (3) TEP's website, accessed during the autumns of 2009 through 2012.

A. Approach to teachers

TEP concentrates resources and attention on hiring, developing, and rewarding high quality teachers, creating a comprehensive model that can, in principle, be replicated. The practices described in the section apply to regular TEP teachers.[4]

All TEP teachers received a base salary of $125,000; all returning teachers received an additional bonus—based on schoolwide performance—equal to 7 to 12 percent of salary.

TEP reported paying all teachers a base salary of $125,000 during the first four years.[5] Teachers who taught at TEP for at least two years and returned to teach at TEP the following year received a bonus. The bonus was based on school performance during the past year and previous years,[6] and so teachers who had taught at TEP for longer periods could receive higher bonuses. The maximum possible bonus was $25,000 in a teacher's second year, and the maximum bonus increased by $5,000 with each additional year at TEP. In 2010–2011, the bonus was $8,992; in 2011–2012 the bonus was either $11,150 or $12,162; and in 2012–2013, the bonus ranged from $12,309 to $14,759.[7]

[3] TEP has continued to adjust its practices since 2012–2013, but those changes do not affect TEP's impacts on student achievement during the period covered by this study and are not described in this report.

[4] Starting with the 2011–2012 school year, TEP also hired a few apprentice teachers, one in 2011–2012 and two more in 2012–2013. TEP created this role to hire promising teachers who typically had only two or three years of experience in hard-to-staff subjects—all three apprentice teachers were special education teachers. Apprentice teachers have the same responsibilities as other TEP teachers, but TEP had different expectations for these teachers during evaluation. TEP paid apprentice teachers $75,000 during the first year and $85,000 in the second year. After the second year, apprentice teachers were either promoted to regular teachers or not rehired.

[5] Teachers at TEP also receive medical, dental, and vision coverage; a 403b retirement plan; short- and long-term disability insurance; and term life insurance. Form 990s indicate that these benefits were valued from $9,689 to $15,933for the highest-paid teachers in 2012–2013.

[6] Specifically, the bonus was based on schoolwide student progress on state tests, schoolwide student progress on department goals, and schoolwide student survey results.

[7] TEP's form 990s indicate that teacher "reportable compensation" for the five highest paid teachers was $117,581 to $124,334 in calendar year 2010 (part of school year 2009–2010 and part of school year 2010–2011), $120,109 to $129,529 in calendar year 2011, and $127,501 to $134,036 in calendar year 2012. TEP reports that compensation is the $125,000 salary plus bonus, minus employee contributions toward medical/dental/vision benefits. Because the bonus is based on schoolwide performance, all teachers receive the same amount as others hired at the same time.

For comparison, during this period a teacher in NYC district public schools with five years of experience would have made $64,009 to $75,796, depending on academic credit and degrees; the maximum teacher salary (22 years of experience and master's degree and additional credit) was $100,049.[8] The median salary for NYC district teachers in TEP's geographic area was $75,092 in 2012–2013.[9]

TEP's teacher hiring process included a full-day teaching audition with TEP students; nearly all teachers hired by TEP had substantial prior teaching experience.

To be hired at TEP, teacher applicants completed a multistage application process focused on live teaching with TEP students—TEP's principal believes this is the most diagnostic information identifying successful teachers. Successful applicants completed a three-stage process: an application including a cover letter, resume, evidence of student learning, and a curricular tool; a lesson taught to a class of TEP students and an interview with the principal; and at least three classes taught at TEP while being observed by TEP's principal and teachers.[10]

TEP's principal designed the process to identify applicants with skills in four areas: (1) teaching expertise and experience, (2) subject-area knowledge, (3) curriculum development ability, and (4) verbal ability. To assess capability in each area, TEP requires applicants to submit the following information or perform the following tasks:

- **Teaching experience and expertise.** Applicants conducted a daylong teacher audition at TEP with TEP students. Applicants also submitted one of the following three items: an unedited video clip of a lesson, a portfolio of students' work that demonstrated two students' progress, and assessment data for at least an entire class of students. TEP recommended that applicants submit the video.[11]
- **Subject-area knowledge.** TEP assessed content knowledge during the teaching auditions.[12]
- **Curriculum development ability.** Applicants submitted an original curricular tool that they developed (for example, a worksheet, teaching technique, or learning technology).

[8] NYC Department of Education-certified teacher salary schedule, effective May 19, 2008.

[9] Salary information for full-time classroom teachers in New York City geographic district #6. Available at [http://www.p12.nysed.gov/irs/pmf/2013/2013_Stat-14.pdf]. Accessed August 19, 2014.

[10] In 2009–2010 and 2010–2011, the application process had four stages: (1) an initial application involving a cover letter, resume, description of relevant coursework and grades; (2), submission of written essays, a curricular tool, and evidence of student learning; (3) an in-person interview with TEP's principal; and (4) a teaching audition involving at least three separate lessons with TEP students at TEP. TEP's principal shifted the process to focus more on evaluating live teaching with TEP students.

[11] For TEP's first two years, applicants submitted two of the three pieces of evidence and an additional piece of evidence that they believed demonstrated student learning, as well as an essay about their pedagogical approach.

[12] During TEP's first two years, applicants also submitted an essay on any topic in the subject area, a written analysis of a pedagogical issue related to the subject area, the number of undergraduate and graduate courses completed in the subject area, and their overall grade point average. (In 2009–2010, TEP also required applicants to submit documentation that they had scored in at least the 90th percentile on a standardized test in the relevant subject area. TEP dropped this requirement for 2010–2011 because many qualified applicants had not taken relevant standardized tests.)

- **Language ability**. TEP assessed the quality of the written work submitted in the application and evaluated communication skills during the interviews and teaching audition.[13]

The teachers TEP hired had a median teaching experience of 6 years, and only 2 of the 42 teachers had 3 or fewer years of teaching experience.[14] TEP's principal reported that the teacher hiring process typically screened out applicants with little urban teaching experience. For comparison, in 2012–2013, the median teaching experience for district teachers in TEP's geographic area was 13 years.[15]

The average class size at TEP was about 31 students; most teachers teach the same class four times daily.

At TEP, the average 5th- through 7th-grade class size for core academic classes—for example, math and music—was about 31 students. Because TEP did not admit new 8th-grade students, class sizes were slightly smaller in that grade. In NYC district middle schools, the average class size from 2009–2010 to 2012–2013 was about 26 or 27.[16] However, there were several middle schools in TEP's neighborhood that have average class sizes for some grades similar to TEP's average.[17]

Most TEP teachers taught four classes of the same subject and the same grade (such as 5th-grade math) daily, with each class lasting 45 minutes. On a daily basis, teachers also provided literacy or math support, supervised students' lunch, or served as the second teacher in a class (typically for special education students). TEP teachers had two preparation periods and lunch.

Each TEP teacher had an administrative role that involved interacting with students, parents, or the community.

All TEP teachers had a daily administrative *whole-school service* role. Some roles, such as grade-level lead, require teachers to work on their whole-school activities during their preparation periods, while teachers in other roles fulfill their responsibilities from 4:00 to 5:00 p.m. These roles almost always related directly to students' development or parent and community involvement, and were intended to provide opportunities for increased teacher

[13] In 2009–2010, TEP also required that applicants submit documentation that they had scored in the 90th percentile or higher on the verbal section of the Graduate Record Examination, Graduate Management Admission Test, or Law School Admission Test. TEP dropped this requirement because many qualified applicants had not taken these tests.

[14] One was a physical education teacher who had no teaching experience and one was a social studies teacher with three years of experience.

[15] Total experience for all classroom teachers in New York City geographic district #6. Available at [http://www.p12.nysed.gov/irs/pmf/2013/2013-Stat-15.pdf]. Accessed September 20, 2014.

[16] Average class size is calculated by dividing the number of students in a program and grade by the number of official classes in that program and grade. NYC Department of Education. 2013. "2012-13 Updated Class Size Report." Available at [http://schools.nyc.gov/offices/d_chanc_oper/budget/dbor/DBOR_CLASS_SIZE/FY13_Data/20122013UpdatedClassSizeReport_20130214_final.pdf]. Accessed September 20, 2014

[17] School-level data for District 6 available at [http://schools.nyc.gov/AboutUs/schools/data/classsize/classsize20130215.htm]. Accessed September 20, 2014.

responsibility and to create a stronger school community (see Table II.1). Teachers helped determine the specific roles and accompanying responsibilities.

Table II.1. Number of teachers in each administrative role, 2009–2010 to 2012–2013

Role	2009–2010	2010–2011	2011–2012	2012–2013
Academic				
Reading and language specialist	1 teacher	5 teachers	3 teachers	3 teachers
English-learner lead	1 teacher	n.a.	n.a.	n.a.
Special education coordinator	n.a.	1 teacher	3 teachers	4 teachers
Math specialist	n.a.	n.a.	3 teachers	n.a.
Tutoring lead	n.a.	n.a.	3 teachers	n.a.
Coverage lead	n.a.	n.a.	3 teachers	4 teachers
Integrated Algebra Regents teacher	n.a.	n.a.	n.a.	1 teacher
Earth Science Regents teacher	n.a.	n.a.	n.a.	1 teacher
Music coach	n.a.	n.a.	n.a.	2 teachers
Literacy lead	n.a.	n.a.	n.a.	1 teacher
Math lead	n.a.	n.a.	n.a.	1 teacher
Operations involving students				
Dean of student discipline and incentives	1 teacher	2 teachers	n.a.	n.a.
Events or assembly coordinator	1 teacher	1 teacher	n.a.	n.a.
Attendance director	1 teacher	n.a.	n.a.	n.a.
Educational technology developer	1 teacher	n.a.	n.a.	n.a.
Lunch director	n.a.	1 teacher	n.a.	n.a.
High school placement director	n.a.	n.a.	3 teachers	n.a.
Incentives lead	n.a.	n.a.	n.a.	4 teachers
Grade-level lead	n.a.	n.a.	n.a.	4 teachers
Student activities coordinator	n.a.	n.a.	n.a.	1 teacher
Community or parent outreach				
Parent involvement coordinator	1 teacher	1 teacher	n.a.	1 teacher
Teacher recruiter	n.a.	1 teacher	n.a.	n.a.
Miscellaneous				
Assessment coordinator	1 teacher	2 teachers	1 teacher	n.a.
Advisory director	n.a.	2 teachers	n.a.	n.a.
Assistant principal	n.a.	n.a.	1 teacher	1 teacher
Teacher development lead	n.a.	n.a.	3 teachers	2 teachers
Basketball coach	n.a.	n.a.	n.a.	2 teachers
Equipment and field manager	n.a.	n.a.	n.a.	1 teacher

Source: TEP.

Note. Some teachers in all years had more than one role. Within each category, roles are ordered by the first year the role existed. Reading and language specialists planned and implemented reading cultural activities (such as book fairs) and reading professional development, planned and led monthly meetings of reading support teachers, and monitored students' reading progress. The coverage lead was the first substitute when a teacher in the same grade was absent. Incentive leads created and coordinated grade-level incentives (such as field trips) for positive behaviors and behaviors needing improvement, and planned culture-building activities for students and families. Grade-level leads were selected by the principal or assistant principal and worked with the school leadership on grade-level logistics, implemented teacher coverage, and met weekly with the other grade-level leaders and school leadership. The advisory director planned advisory meetings at which students met weekly in small groups to build relationships and develop character.

n.a. = not applicable (not an assigned role during the specific school year).

Aside from the principal, TEP had few administrative staff: a director of finance and operations and office manager, both starting before TEP opened; a business manager who started in 2010–2011; and an operations manager and a high school placement director (also alumni director) who started in 2012–2013. Each cohort at TEP also had a social worker who followed the cohort through all four grades. Teachers handled most administrative responsibilities. TEP's assistant principal—a new role created in 2011–2012—was a teacher who continued to teach four classes.[18]

During the school year, professional development involved weekly observations of other teachers and feedback from being observed by other teachers.

TEP's principal believes that teachers learn by watching others teach and by receiving feedback from peer observers. At least twice a week, TEP teachers were expected to observe one another teach.[19] Typically, teachers had an assigned partner who rotated every quarter; sometimes the teachers chose their partners and sometimes the principal or assistant principal did. Teachers were expected to observe their partners at least once a week and could observe any other teacher once a week. Teachers provided written feedback to their partners and were expected to meet with them weekly.

Almost all TEP teachers believed that this collaboration was productive—although similar percentages of all teachers in NYC seem to have felt the same away about their development. When surveyed by the NYC Department of Education (DOE) in 2011–2012 and 2012–2013, 87 to 96 percent of TEP teachers strongly agreed or agreed that their "professional development experiences this school year included opportunities to work productively with colleagues."[20] The average rate of agreement for teachers in all NYC schools was almost identical.

TEP teachers helped set school practices and received professional development during a six-week summer institute.

Teachers also received professional development during the summer, at the Summer Development Institute (SDI). During these institutes, the principal and teachers typically met daily (Monday through Friday) from 9:30 a.m. to 4:30 p.m. for six weeks.

During the SDIs, all incoming TEP teachers reviewed students' performance during the previous year and discussed appropriate policies for the upcoming year. Agendas guided the activities, which included individual, departmental, grade-level, and whole-school service planning. For many issues, teachers met in small groups to discuss and plan changes, which they then shared with the whole group.

[18] TEP's principal created this role, in part, to develop future leaders.

[19] In each quarter of 2009–2010, TEP teachers were expected to conduct daily observations. Teachers believed that the daily observations were too frequent, and the partnership model was revised for the 2010–2011 school year.

[20] This survey is administered to all teachers in NYC and does not ask about specific TEP practices. Teachers knew that their responses would affect their school's progress reports. Response rates among TEP teachers were always 100 percent; response rates for all NYC teachers were in the low 80s. Available at [http://schools.nyc.gov/Accountability/tools/survey/default.htm]. Accessed August 19, 2014.

TEP teachers felt involved in setting policy at TEP. When surveyed by NYC DOE from 2009–2010 to 2011–2012, 91 to 100 percent of TEP teachers strongly agreed or agreed that "school leaders invite[d] teachers to play a meaningful role in setting goals and making important decisions." For comparison, during the same years, the average rate of agreement for teachers in all New York City schools was 79 to 84 percent.

TEP teachers were evaluated on five primary performance dimensions and several secondary dimensions; teachers must receive high scores on all primary dimensions to be rehired.

Twice annually, the TEP principal or assistant principal evaluated each TEP teacher on five primary domains:[21]

- **Professional expectations.** Teachers were assessed on their attendance (three or fewer personal days and five or fewer sick days), punctuality, and whether the teacher submitted appropriate lesson plans, unit plans, and homework assignments to the principal or assistant principal.
- **Adherence to TEP staff norms.** Teachers were assessed using peer surveys—and compared to the average TEP teacher—on 10 professional standards, such as positivity, having productive "difficult" conversations, and modeling appropriate student behavior.
- **Classroom management.** Teachers were assessed using principal or assistant principal observations; student surveys administered twice annually;[22] and data regarding adherence to TEP discipline protocols (for example, checked whether the teacher assigned work to students receiving in-school suspension).
- **Instructional planning & delivery.** Teachers were assessed using principal or assistant principal observations; student perceptions related to instruction;[23] and the quality of instructional materials such as lesson plans, classroom handouts, and homework.
- **Assessment of student growth.** Teachers were primarily assessed on their students' achievement growth as measured on assessments developed by each TEP department (for example, math and social studies).[24]

Teachers were also evaluated on several secondary domains: teacher partnerships, whole school service, extended-day activities, hallway transitions, physical classroom environment, and administrative responsibilities.

[21] TEP's teacher evaluation process changed during the first four years, but the primary domains stayed about the same.

[22] Students answered four questions about their teachers' classroom management, such as "Students in this class treat this teacher with respect" or "Student behavior in this class is under control." These questions are based on student surveys developed by the Tripod Project for School Improvement.

[23] Students answered seven questions about their teachers' instruction, such as "In this class, we learn a lot almost every day," and "My teacher explains difficult things clearly." These questions are based on student surveys developed by the Tripod Project for School Improvement.

[24] For more information on these assessments, see page 11.

For most domains, teachers were rated on a three-point scale where a "1" indicated below TEP standard, a "2" indicated approaching TEP standard, and a "3" indicating meeting TEP standard (for the professional expectations domain, teachers can also receive a 0 or 4.) Whether teachers were rehired was determined by the total scores on the principal report for the end of the year. Teachers who received a total score between 13 and 15 points were typically rehired to teacher at TEP the following year (unless their evaluations on the secondary domains were low), those with a total score of 12 or 12.5 were rehired on a case-by-case basis (high performance on secondary domains was considered), and those with lower scores were not rehired.

About 35 percent of new TEP teachers were not rehired by TEP for a second year; an additional 12 percent resigned during or after their first year.

Of the 43 TEP teachers who were teaching at the beginning of a school year during TEP's first four years of operation, 20—or 47 percent—did not return to TEP after their first year (see Table II.2). About 35 percent were not rehired for a second year—effectively terminated—and about 12 percent resigned. (TEP signaled to teachers in advance whether they would be rehired, and several teachers resigned rather than not be rehired.) Of the 16 teachers in TEP's first three teaching cohorts who returned for a second year, one was not rehired and 7 resigned during or after their second year. For comparison, a study of NYC district public schools found that about 27 percent of middle school teachers who joined a school did not return for a second year (Marinel 2011).[25]

Table II.2. Number of TEP teachers who were not rehired or resigned, by teaching cohort

Year hired	Teachers hired	Year 1	Year 2	Year 3	Year 4
2009	8 teachers	2 not rehired 1 resigned[a]	1 not rehired 1 resigned	0 not rehired 0 resigned	0 not rehired 1 resigned
2010	10 teachers	3 not rehired[b] 3 resigned[c]	0 not rehired 1 resigned	0 not rehired 0 resigned	n.a.
2011	14 teachers	6 not rehired[d] 1 resigned[e]	0 not rehired 5 resigned	n.a.	n.a.
2012	11 teachers	3 not rehired 0 resigned	n.a.	n.a.	n.a.

Source: TEP.

Note. TEP reported that several of the teachers who resigned would not have been rehired. These frequencies include only those teachers employed by TEP at the start of the year and do not include the five teachers that TEP hired mid-year to replace teachers who were terminated or resigned during the school year. Unless otherwise noted, the resignations and nonrenewals occurred during the summer following the school year.

[a] This teacher resigned during the fall.

[b] One of these teachers was terminated during the fall.

[c] Two of these teachers resigned during the fall.

[d] One of these teachers was terminated during the fall.

[e] This teacher resigned in the spring.

n.a. = not applicable.

[25] The Marinel study used a broader definition of *attrition* that included teachers who became administrators at the school.

B. Academic and behavior practices

In this section, we describe TEP's curriculum, use of data to improve instruction, and student behavior policy. TEP's admission lottery is described in Appendix B.

TEP students took classes in ELA, social studies, math, science, music, and physical education; TEP replaced daily Latin with a second period of ELA in 2012–2013.

From 2009 to 2012, all TEP students took daily 45-minute classes of ELA, social studies, math, science, Latin, and music; in 2012–2013, an extra period of ELA replaced Latin.[26] For the first two school years, students also had physical education (PE) for three days a week—the other two days were assembly and student advisement—and in 2011–2012 and 2012–2013 students had PE daily. For TEP's first two years, students had an additional 30 minutes of daily language and math instruction; starting in 2011–2012, students received an additional 30 minutes of math support provided to groups of 10 to 20 students with similar achievement.

Class content typically followed the New York State standards with only minor changes during TEP's first three years. Starting in 2012–13, the content became more aligned with the Common Core State Standards, especially in ELA; focused on more complex texts and more argumentative writing; and used science and social studies to help teach literacy. TEP teachers determined the instructional approaches used in their classes.

TEP students participated in enrichment activities and received additional instructional support during the last period of the day.

During the last period of the day—TEP started at 8:00 a.m. and continued until 4:00 p.m.[27]—all students received enrichment and instructional support from TEP teachers for about an hour. The enrichment activities occurred three days a week. For example, during the 2010–2011 school year, teachers designed and implemented the following yearlong programs: Arts and Crafts, Headline News (journalism), Tricksters (circus techniques), Chess Club, TEP Vets (caring for animals), Team Spirit (girls sports club), Get Fit (fitness club), Team First (boys sports club), Board Games, Photography, Cooking, School Practice (similar to detention), and Running Rackets (track and racket sports). TEP's principal established the enrichment to enable students and teachers to participate in activities they enjoy in smaller groups. Teachers proposed the program, and it was approved if there was sufficient student interest. For the other two days of the week, students received academic support from TEP teachers for about an hour. The support students received was based on a teacher's determination of their needs.

[26] The TEP principal reported replacing Latin with a second period of ELA or literacy because TEP could not find enough Latin teachers who could effectively manage TEP classrooms. The principal added a second period of ELA or literacy because he felt that students, especially English learners, needed more language help and that additional language and reading instruction was a prerequisite for social studies and science achievement. Sixth and 8th grades had an extra period of literacy that focused on nonfiction texts and in which students were organized by achievement. Fifth and 7th grades had two periods of ELA taught by the same teacher.

[27] Initially, the school day ended at 5:00 p.m. In mid-October 2009, after about six weeks, the TEP principal shortened the school day by about an hour because he realized the day was too long for students and teachers.

Each academic department identified students' goals and tracked students' progress.

At TEP, every department, including PE, identified three core outcomes that students should achieve in each grade; these outcomes were tracked for every student using department-developed assessments. For example, to track performance on an outcome, the English department might have used three types of assessments: (1) assessments developed by a teacher targeted at instructional topics, (2) assessments aligned with state assessments, and (3) state assessments. Every assessment had a target, either mastery or growth. Over TEP's first four years, the assessments and outcomes changed and become more sophisticated.

TEP also administered the Northwest Evaluation Association's Measures of Academic Progress (math, ELA, and science) in the fall and spring of each year. Teachers used these data to identify students' needs and plan instruction accordingly.

TEP did not expel any students and did not use out-of-school suspensions; in 2011–2012, TEP shifted its behavior approach to emphasize interventions based on student-teacher relationships.

TEP's principal reports that the school did not expel or suspend (out of school) any student during its first four years. For severe disciplinary infractions such as fights, TEP students received in-school suspensions supervised by a social worker. The student completed regular classwork separate from his or her peers; worked to rebuild relationships; and, with family, met with the social worker and school leadership.

TEP significantly changed its disciplinary practices after 2010–2011, moving away from have a dean supervise detention toward detention supervised by the relevant teacher. During the school's first two years, students who misbehaved in classrooms received detention under deans of discipline. TEP's principal and teachers determined that this approach was not improving behavior, and decided that disciplinary consequences at TEP should be relational—consequences for misbehavior should build relationships between students or between a teacher and a student. For example, when a student was disrespectful to a teacher, the student has to spend the whole day with the teacher.

TEP teachers mostly believed that the school's disciplinary practices were effective. When surveyed by NYC DOE, 72 to 100 percent of TEP teachers *strongly agreed* or *agreed* that "order and discipline were maintained" at TEP. The rate of agreement dropped from 100 percent during the first two years to 84 percent in 2011–2012 and 72 percent in 2012–2013. For comparison, the average rate of agreement for teachers in all New York City schools was typically about 80 percent during this period.

III. TEP'S STUDENT POPULATION AND STUDENT ATTRITION

Some research has raised concerns that charter schools enroll fewer students with disabilities (for example, Government Accountability Office 2012) or attract higher-achieving students than district public schools. Critics have raised concerns that some charter schools selectively enroll students, a practice labeled cream skimming or creaming. TEP's lottery-based admission process, which favors low-achieving applicants, precludes obvious favoritism in the admissions process. Nonetheless, TEP might attract applicants who are less disadvantaged in particular ways.

Researchers have also raised concerns that charter schools might have higher rates of student attrition (for example, see Miron et al. 2011), possibly pushing out weaker or disruptive students. This attrition could result from a more challenging school environment, high expectations of students or parents, expulsion (see Ahmed-Ullah and Richards 2014), or other school practices. (TEP's principal reports that the school has not expelled any students.) If departing students are not replaced or are replaced by stronger students, this attrition could create a better peer-learning environment.

Attrition from TEP cannot bias our estimates of TEP's impact on student achievement, because students who leave TEP to enroll in other NYC schools continue to be counted as TEP students for the purposes of the impact analysis. However, evidence that TEP had high student attrition would alter the interpretation of any impacts, which could be due to better peer-learning created by selective attrition rather than TEP practices. Moreover, such attrition would indicate that the TEP approach is not scalable, because district public schools cannot push out students and must accept all applicants.

In this chapter, we examine TEP's enrollment patterns by comparing the 4th-grade achievement (the grade before TEP enrollment) and demographic characteristics of TEP students to (1) students in all NYC public schools who attended 4th grade in the same year (that is, who belong to the same cohort) and (2) the subset of students in the same cohort who attended 4th grade in schools that send students to TEP. We then compare student attrition between 5th and 8th grades at TEP to attrition at other schools that start in 5th grade and enroll students who attended schools in TEP's neighborhood. Finally, to examine whether TEP is replacing students who leave with stronger students, we compare the baseline characteristics of students who leave TEP with the characteristics of students who stay as well as the new students who replace the leavers.

All results in this chapter are based on NYC DOE student-level administrative data. For each student, the data include 4th-grade state assessment scores in ELA, mathematics, and science. The data also include information on each student's gender, race/ethnicity, special education status, English learner (EL) status, free or reduced-price (subsidized) lunch status, and home language. Finally, for every school a student attends, the data provide the date of enrollment and, when relevant, the date of discharge.

A. Student population

We identified 4th-grade achievement, attendance, and demographic characteristics for three groups of students: (1) students who attended TEP for at least one day;[28] (2) students who attended a school that sends students to TEP, labeled neighborhood schools; [29] and (3) all New York City students, including neighborhood school students. Table III.1 reports the average for each characteristic across all four cohorts, with each cohort weighted equally.

Table III.1. Baseline characteristics of TEP students, neighborhood school students, and NYC students, all cohorts

Characteristic	TEP	Neighborhood schools	All NYC
4th-grade math achievement	-0.31	-0.29	0.03*
4th-grade ELA achievement	-0.33	-0.30	0.03*
4th-grade science achievement	-0.44	-0.37	0.03*
Male	0.50	0.51	0.51
African American	0.11	0.15*	0.30*
Hispanic	0.87	0.77*	0.40*
Other race	0.02	0.07*	0.30*
Subsidized lunch	0.94	0.91	0.85*
Special education	0.17	0.17	0.19
English learners	0.32	0.35	0.16*
English home language	0.32	0.36	0.60*
Spanish home language	0.67	0.60*	0.24*
Other home language	0.01	0.04*	0.17*
Attended charter[a]	0.00	0.00	0.03*
Sample size	491	11,294	288,933

Source: NYC DOE administrative data.

Notes: For each group, the mean or average reported is the average for the group across all four cohorts with each cohort weighted equally. Students in neighborhood schools and all NYC schools were in 4th grade during the same year as the TEP cohort. Statistical significance was calculated using a t-test. Slightly more than 5 percent of neighborhood students and all NYC students were missing ELA achievement information; no other characteristic had missing data for more than 5 percent of students for any group. For data that can change over time, a student is classified as missing data if he or she was not enrolled in NYC schools in 4th grade or NYC data did not have that information for the student. For immutable characteristics (sex, race, and home language), a student is classified as having missing data if NYC did not have that information for that student; for students not enrolled in NYC schools in 4th grade, we used their 5th-grade data. Achievement scores are ranked and then standardized by subject and cohort to have a mean of 0. However, we assign students to a cohort by whether the student was in the appropriate cohort in grade 5, and some students who have test scores in the 4th grade are retained, skip a grade, or do not attend an NYC district or charter school after 4th grade. Because these students are excluded from these summary statistics, the reported means deviate slightly from 0. This analysis does not examine post 4th grade characteristics (outcomes), so these samples include students who do not have outcome data and, thus, differ from the other samples presented in this report.

[a] Aside from a few students in the 2010 cohort, no TEP or neighborhood school students attended a charter school in 4th grade.

*Significantly different from TEP student characteristics at the 0.05 level.

[28] DOE data indicate that a few students each year leave TEP (and other schools) during the first, second, or third day of the year. These students almost certainly did not actually attend TEP (or the other schools) because they had transferred to another school during the summer (that is, the attendance is an artifact of the data system). We do not classify these students as attending TEP.

[29] These schools are conventionally labeled feeder schools; to minimize jargon, we label these schools neighborhood schools. About two-thirds of these schools are in New York City District 6 (TEP's district), and almost all of the others are in Manhattan or the Bronx, both of which border District 6.

Relative to the citywide average, TEP students are lower-achieving, more likely to be Hispanic, and more likely to receive a subsidized lunch; TEP students are equally likely to receive special education services.

Compared with all NYC students, TEP students have significantly lower 4th-grade math, ELA, and science achievement scores; are more likely to receive a subsidized lunch; and are less likely to attend a charter school during 4th grade (Table III.1). TEP's student population was about 47 percentage points more Hispanic than the rest of the city—and less likely to be African American or other race—and consequently also more likely to be English learners and speak Spanish at home. TEP students were similar to NYC students overall on the percentage that are male and the percentage classified as special education. These trends are consistent across each cohort.

Relative to other students at neighborhood schools, TEP students are similar on baseline achievement and other demographic characteristics, except that TEP students are more likely to be Hispanic.

TEP students are similar to students at neighborhood schools (that is, TEP students are similar to their 4th-grade school peers). They have similar baseline achievement, subsidized-lunch participation rate, special education participation rate, and most other characteristics. TEP's student population was about 10 percentage points more Hispanic, and about 4 or 5 percentage points less likely to be either African American or other race (white, Asian, or Native American). These trends are consistent across each cohort.

To summarize, we find no evidence that students enrolling at TEP are advantaged on any dimension compared with their peers.

B. Student attrition

To estimate student attrition, we calculated the percentage of students—enrolled in TEP at the beginning of 5th grade—who had departed TEP by the end of one, two, three, and four years.[30] We estimated cumulative attrition rates for each TEP student cohort over different periods: after one year (four cohorts), two years (three cohorts, because the 2012 cohort had been enrolled at TEP for only one year by the end of the 2012–2013 school year), three years (two cohorts), and four years (one cohort). Our measure of attrition includes students who leave for reasons related and unrelated to TEP (for example, their families move to another city). To

[30] For a time period, the numerator in our primary measure of attrition is the number of students who were admitted to TEP before October 1 during the cohort's first year, were enrolled on the fourth day of school through September 30 during the cohort's first year, and were discharged on or after the fourth day of school. The denominator in our primary measure is the number of students admitted to the school before October 1 during the cohort's first year and still enrolled in that school by the fourth day of school of the cohort's first year or during any day in September after the fourth day. The time period begins on the fourth day of school of the cohort's first year and ends on the last school day of the period. Our primary measure includes students who enroll at TEP by the end of September, but excludes students who enroll but do not attend TEP. DOE data indicate that a few students each year leave TEP during the first, second, or third day of the year. TEP reported that these students did not actually attend TEP because they had transferred to another school during the summer (that is, the attendance is an artifact of the data system). A few students leave and reenroll at the school during the time period. These students are not considered attrition because they return to the school.

ensure that our attrition findings are robust to the definition of attrition, we also calculated attrition in two other ways.[31]

To provide a comparison, we also estimated attrition for all schools in NYC that (1) start in 5th grade and continue until at least 8th grade during 2009, 2010, 2011, and 2012; (2) have a 5th grade with at least 25 students each year; and (3) enroll a student from any of TEP's neighborhood schools (any cohort).[32] Of the 14 comparison schools that meet these criteria, 10 are charter schools. The students enrolled in these schools differed from TEP students,[33] and as a sensitivity check we estimated attrition controlling for the characteristics listed in Table III.1.

TEP's student attrition rate was similar to that of comparable schools.

Compared to the 14 other schools in NYC, TEP had similar student attrition rates and the differences were not statistically significant (Table III.2). This finding holds: using the other definitions of student attrition,[34] using the median attrition rate for comparison schools, and controlling for students' 4th-grade achievement and demographic characteristics. Across cohorts, the average student attrition rate at TEP was about 5 percent after one year, about 10 percent after two years, about 17 percent after three years, and about 20 percent after four years.

This low attrition rate is consistent with the high satisfaction that TEP students and parents reported in the annual NYC Schools Surveys. In 2009–2010 and 2010–2011, about 75 percent of responding TEP parents were very satisfied with the education their children received, and about 23 percent were satisfied. In 2011–2012 and 2012–2013, about 64 percent were very satisfied and about 33 percent were satisfied.[35]

[31] The first measure also excludes students who enroll but do not attend TEP, but includes students who enroll at TEP by the fourth day of the school year (rather than by September 30). The second measure includes all students who enroll at TEP, including students who do not actually attend TEP and students enrolled at any time during the year.

[32] Schools that end in the 5th grade will have 100 percent attrition after one year. K–8 schools or schools that start earlier than 5th grade are not an appropriate comparison because parents who enroll their children in an early grade and do not like the school would presumably leave in an earlier grade. Schools that had fewer than 25 students in the 2009, 2010, 2011, or 2012 5th-grade class were excluded. We limit the sample to schools attended by students who attended a neighborhood school to identify a more similar comparison group.

[33] Students at these comparison schools had significantly higher math, ELA, and science achievement in the 4th grade (between 0.28 and 0.40 standard deviation units), were 34 percentage points more likely to be African American (47 percentage points less likely to be Hispanic), and were significantly less likely to receive special education services (5 percentage points) or receive a subsidized lunch (8 percentage points). Minority and low-income students are more likely to change schools (Hanushek 2004).

[34] Using the alternative definitions of attrition, TEP had significantly lower attrition for one cohort/period, and significantly higher attrition for two cohorts/periods. The remaining cohorts and periods were not significantly different.

[35] More than 93 percent of surveyed parents responded each year. During this period, the percentage of all NYC parents who were very satisfied was consistently about 47 percent, and the percent satisfied was about 47 percent.

Table III.2. TEP's student attrition rates and average attrition rate of comparable schools, by cohort and duration

	2009–2010		2010–2011		2011–2012		2012–2013	
	TEP	Comp.	TEP	Comp.	TEP	Comp.	TEP	Comp.
2009 cohort	6%	8%	10%	17%	19%	21%	20%	23%
2010 cohort	n.a.	n.a.	10%	6%	17%	14%	17%	18%
2011 cohort	n.a.	n.a.	n.a.	n.a.	4%	5%	6%	9%
2012 cohort	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	4%	1%[a]

Source: NYC DOE administrative data.

Note: For all cohorts except the 2009 cohort, there are 14 comparison schools. In 2009, 9 comparison schools had more than five students who were in the DOE biographic data as attending the school, but did not have a DOE admission date for the school. These schools with attendance anomalies are not included, and there are 5 comparison schools for the 2009 cohort. The statistics were calculated using ordinary least squares regressions of attrition on an indicator variable for TEP enrollment. Students were weighted such that each comparable school had the same weight in the analysis. None of the differences between TEP and the comparable schools were statistically significant.

[a] Nine comparison schools had no student attrition during this school year.

n.a. = not applicable.

C. Characteristics of the TEP students who left early

To determine whether the TEP students who left differed from the remaining TEP students, and whether TEP replaced students who left with different types of students, we identified the 4th-grade test scores and demographics of students who left TEP—using our primary definition of attrition—as well as the characteristics of the remaining TEP students and the students who replaced those who left. For each cohort, a student was classified as a replacement if he or she enrolled in TEP after September 30 of the cohort's first year at TEP.[36]

The TEP students who left TEP were less likely to be Hispanic (more likely to be African American) but otherwise similar to the remaining TEP students; the students who replaced the leaving students were more likely to receive special education services and be Hispanic (less likely to be African American), and otherwise similar to the students they replaced.

The students who left TEP were similar to the students who remained—the only statistically significant different was that leavers were about 11 percentage points more likely to be African American (and 14 percentage points less likely to be Hispanic)—providing no evidence that TEP encouraged particular types of students to leave (Table III.3). Similarly, the replacements were similar to the students who left; the only differences were that they were more likely to receive

[36] For the 2010–2011 school year, TEP continued to enroll new students through the end of September to reach its target enrollment number. Classifying only those students who enrolled after September 30 as replacements ensures that these students are not considered replacements for students who left (as no students who had attended TEP had left). Because students who leave in the spring are not replaced until the summer, students must enroll by the last day of summer in 2013 to be considered a replacement for the 2010, 2011, or 2012 cohorts.

special education services and about 11 percentage points less likely to be African American. Across all cohorts, TEP replaced about 41 percent of students who left.[37]

Table III.3. Baseline characteristics of students who left TEP, remaining students, and replacements

Characteristic	Leavers	Remaining TEP	Replacements
4th-grade math achievement	-0.35	-0.31	-0.44
4th-grade ELA achievement	0.32	-0.33	-0.69
4th-grade science achievement	-0.51	-0.43	-0.73
Male	0.40	0.51	0.43
African American	0.27	0.09*	0.04*
Hispanic	0.71	0.89*	0.87*
Other race	0.02	0.02	0.09
Subsidized lunch	0.88	0.94	0.91
Special education	0.19	0.17	0.48*
English learners	0.21	0.33	0.30
English home language	0.48	0.31	0.30
Spanish home language	0.50	0.69	0.70
Other home language	0.02	0.01	0.01
Attended charter	0.02	0.00	0.00
Sample size	48	440	23

Source: NYC DOE administrative data.

Note: Statistical significance was calculated using a t-test.

*Significantly different from TEP student characteristics at the 0.05 level.

[37] If attrition is defined as students who left TEP after the third day of the school year and replacements as students who joined after the third day—consistent with one of our alternative definitions of attrition —the number of replacements is roughly equal to the number of students who left (51 replacements and 54 leavers) because of the new students that TEP enrolled in September 2010. The replacements are about 14 percentage points less likely to be African American (p = .06) and about 20 percentage points more likely to speak Spanish at home; there are no other statistically significant differences. Our other alternative definition of attrition—all students who enroll at TEP during the first year—precludes the concept of replacement students because the attrition measure includes all students enrolled during the year.

IV. IMPACT FINDINGS

TEP's approach and curriculum focus on helping students attain mastery in four core subjects: math, ELA, science, and social studies. To examine TEP's impact on achievement, we used student-level administrative data obtained from the NYC DOE and focused on students' achievement test outcomes in math and ELA. We also examined TEP's impact on science achievement in 8th grade—the only middle school grade in which students were tested—for the 2009 entry cohort.

We begin this chapter with a brief description of our methods for estimating impacts (see Appendix A for more information). We then present findings on TEP's cumulative impacts on student achievement separately by cohort and for each year that TEP has operated (2009–2010, 2010–2011, 2011–2012, and 2012–2013).

A. Data and methods

The four cohorts followed in this report entered TEP's 5th grade in fall 2009, fall 2010, fall 2011, and fall 2012.[38] To account for possible bias due to selective student attrition at TEP, we classify students as TEP students if they attended TEP for at least one day during the 5th grade,[39] regardless of whether they left TEP and enrolled at another school during the 5th-grade year or after completing it.[40] This approach produced a conservative estimate—biased toward zero—of the impact of continuously enrolling in TEP. As described in the preceding chapter, attrition from TEP was relatively low: about 80 percent of TEP's first cohort of 5th graders graduated from TEP four years later.

The primary achievement outcomes are standard scores on state ELA and math tests that have been converted to rank-based *z*-scores (Gill et al. 2005) to reduce the influence of unreliable outliers. We report TEP's impacts on science achievement for the first cohort, which was the only cohort that had completed 8th grade at the time of the analysis. For each year, grade, and subject combination, the mean rank-based *z*-score is 0 and the standard deviation is 1. A few students in each TEP cohort did not have outcomes and cannot be included in the analyses.

Because an experimental analysis using TEP's admissions lottery had several limitations (discussed in Appendix B), we decided before conducting analyses to only use a quasi-experimental matching approach to estimate impacts. Charter school impact estimates based on

[38] Five students from the 2009 cohort, 10 students from the 2010 cohort, and 5 students in the 2011 cohort entered TEP in the 6th or 7th grades. These students were excluded from the impact estimates because they received a different level of exposure to TEP than the rest of their cohort. Chapter III provides information about these students' characteristics in the 4th grade.

[39] Students who enrolled at TEP during the spring or summer but then withdrew before school started were often identified in the data as withdrawing on the first, second, or third day of school even though they never attended TEP. Accordingly, a student was classified as a TEP student if he or she attended TEP for at least one day after the third school day of the 5th-grade year.

[40] Students who leave the NYC district or charter schools entirely do not have observed outcomes and are not included in the analysis.

matching methods are very similar to experimental estimates as long as pre-treatment achievement measures are used in the analysis (Fortson et al. 2013; Gill et al. 2013).

Our quasi-experimental approach estimated propensity-score models (Rosenbaum and Rubin 1983) to identify a comparison group of students with similar characteristics and prior achievement to the students who enrolled at TEP. We first limited the potential comparison group to students who attended a school that was also attended by future TEP students during the 4th grade (a TEP neighborhood school). We then estimated propensity scores using baseline (4th-grade) test scores, pre-baseline (3rd-grade) test scores, and multiple demographic characteristics. These propensity scores measured a student's probability of enrolling at TEP, and we used them to select a matched comparison group. Each TEP student was matched to one or more non-TEP students with the most similar propensity for enrolling at TEP. For each cohort, the matched comparison group is similar to the TEP cohort on multiple baseline demographic characteristics and achievement test scores (see a detailed table of baseline differences in Appendix A).

Finally, we estimated impacts using a regression model that included covariates to control for any remaining observed baseline differences between the TEP cohorts and their matched comparison groups. We estimated impacts separately for each cohort and for each possible period, from a minimum of one year (for all four cohorts) to a maximum of four years after enrollment at TEP.

B. Findings

In this section, we describe impacts estimated using the primary matching approach and model. We present impacts estimated using several sensitivity analyses in Appendix C. The findings from the sensitivity analyses are consistent with the primary findings presented in this chapter.

Table IV.1 presents math impacts by student cohort and years after enrolling at TEP; Table IV.2 presents ELA impacts by student cohort and years of enrollment. Impacts are reported in standard deviation units, conventionally known as effect sizes.

Table IV.1. TEP impacts on math achievement, by cohort and duration

	1 year after enrolling at TEP	2 years after enrolling at TEP	3 years after enrolling at TEP	4 years after enrolling at TEP
2009 entering 5th graders	-0.30** (0.05) N = 981	-0.23** (0.05) N = 977	0.30** (0.05) N = 971	0.64** (0.05) N = 973
2010 entering 5th graders	0.02 (0.04) N = 1,058	-0.14** (0.05 N = 1,053	0.17** (0.05) N = 1,052	
2011 entering 5th graders	0.04 (0.04) N = 1,111	0.17** (0.04) N = 1,110		
2012 entering 5th graders	0.15** (0.05) N = 970			

Table IV.1 (*continued)*

Note: This table reports the coefficients from linear regressions of standardized math test scores on an indicator variable for TEP enrollment in 5th grade. Separate models were run for each cohort and outcome year combination. The comparison group consists of matched students from neighborhood schools who never enrolled in TEP; matching was conducted by cohort using the propensity scores predicted by the model as described in this chapter and Appendix A. Regression controls include two years of baseline test scores in math, ELA, and science, as well as indicator variables for baseline demographic characteristics reported in Appendix A. Robust standard errors are reported in parentheses. Analyses are weighted using the method described in Appendix A.

**Significantly different from zero at the 0.01 level, two-tailed test.

Table IV.2. TEP impacts on ELA achievement, by cohort and duration

	1 year after enrolling at TEP	2 years after enrolling at TEP	3 years after enrolling at TEP	4 years after enrolling at TEP
2009 entering 5th graders	-0.28** (0.05) N = 981	-0.09** (0.05) N = 977	0.10** (0.05) N = 971	0.13** (0.05) N = 974
2010 entering 5th graders	-0.21** (0.05) N = 1,058	0.05 (0.05) N = 1,053	0.10* (0.06) N = 1,052	
2011 entering 5th graders	-0.03 (0.04) N = 1,111	0.02 (0.05) N = 1,110		
2012 entering 5th graders	-0.01 (0.05) N = 970			

Note: This table reports the coefficients from linear regressions of standardized ELA test scores on an indicator variable for TEP enrollment in 5th grade. Separate models were run for each cohort and outcome year combination. The comparison group consists of matched students from neighborhood schools who never enrolled in TEP; matching was conducted by cohort using the propensity scores predicted by the model as described in this chapter and Appendix A. Regression controls include two years of baseline test scores in math, ELA, and science, as well as indicator variables for baseline demographic characteristics reported in Appendix A. Robust standard errors are reported in parentheses. Analyses are weighted using the method described in Appendix A.

*Significantly different from zero at the 0.05 level, two-tailed test.

**Significantly different from zero at the 0.01 level, two-tailed test.

TEP's impacts on achievement were largely negative for the first two cohorts during their first two years at TEP.

For the first two cohorts, the analysis found that two years after enrollment (that is, typically the end of 6th grade), the average math achievement outcomes of the TEP students were lower than those of the comparison group, and these differences were statistically significant. In ELA, TEP's two-year impacts were significantly negative for the 2009 cohort and not statistically distinguishable from zero for the 2010 cohort (Table IV.2). Negative achievement impacts are common in the first years after a charter school opens, before its operations stabilize (Gill et al. 2007).

TEP's impacts on student achievement consistently improved over time.

TEP's one-year (5th-grade) impacts consistently improved after TEP's first year of operation (the 2009–2010 school year). In math, TEP's one-year impact for the 2009 cohort was

negative and significant. The one-year impacts on math achievement for the 2010 and 2011 cohorts were statistically indistinguishable from zero. By the fourth cohort, TEP's impact on math achievement was positive and statistically significant. Each of these improvements relative to the first year of operation was statistically significant. A similar upward trend occurred in ELA, as each of the one-year impacts for the subsequent cohorts was significantly better than TEP's first-year impact. This finding of improving one-year impacts over the four successive cohorts is consistent with earlier literature indicating that charter school impacts improve as the age of the school increases (Gill et al. 2007; Carruthers 2012).

By the fourth year of operation (the 2012–2013 school year), TEP's achievement impacts were almost completely positive across all cohorts, despite the differences in duration after enrolling at TEP. In 2013, TEP's impacts on math achievement were positive and significant for every cohort. (These impacts reflect four years after enrollment for the 2009 cohort, three years after enrollment for the 2010 cohort, two years after enrollment for the 2011 cohort, and one year after enrollment for the 2012 cohort). Similarly, in 2013 TEP's impacts on ELA achievement were positive and significant for the 2009 cohort (four years of treatment) and for the 2010 cohort (three years of treatment). ELA impacts observed in 2013 were not statistically distinguishable from zero for either the 2011 cohort (two years of treatment) or the 2012 cohort (one year of treatment).

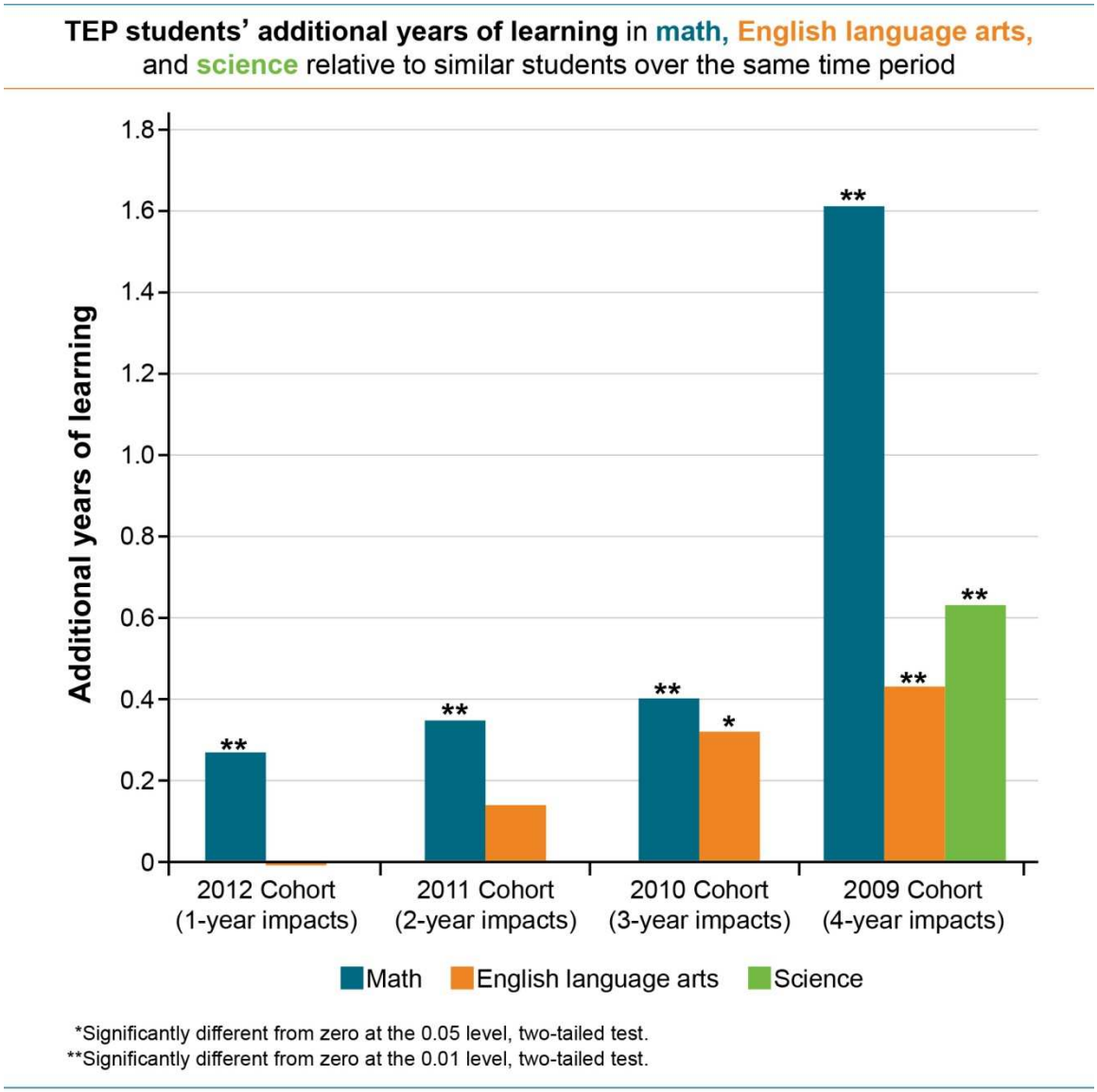
By three or four years after enrollment, TEP's impact on student achievement was consistently positive.

For students who were followed for at least three years, average achievement outcomes were consistently higher than the average outcomes of the comparison group by the third year after enrollment at TEP, and these differences were statistically significant. Three-year impacts on math and ELA were significantly positive for both the 2009 and 2010 cohorts, and four-year impacts on math and ELA for the 2009 cohort were significantly positive. Four-year impacts on science achievement for the 2009 cohort were also positive (0.19) and statistically significant (standard error = 0.06; N=967).

After four years, students who enrolled at TEP had test score gains equal to an additional 1.6 years of school in math, an additional 0.4 years of school in ELA, and an additional 0.6 years of school in science.

We converted TEP's impacts to years of additional learning, using the metrics estimated in Bloom et al. (2008). Using the benchmarks for average annual learning gains, TEP students in the first cohort, a full four years after enrollment at TEP, were ahead of their most similar peers in math by more than 1.5 years of learning. Using the benchmarks for average annual learning gains, TEP students in the first cohort, a full four years after enrollment at TEP, were ahead of their most similar peers in math by more than 1.5 years of learning. TEP students in the first cohort also surpassed their peers by about 0.4 years of learning in ELA and 0.6 years of learning in science. Figure IV.1 presents impacts for each cohort after its members' maximum possible number of years at TEP.

Figure IV.1. TEP students' additional years of learning



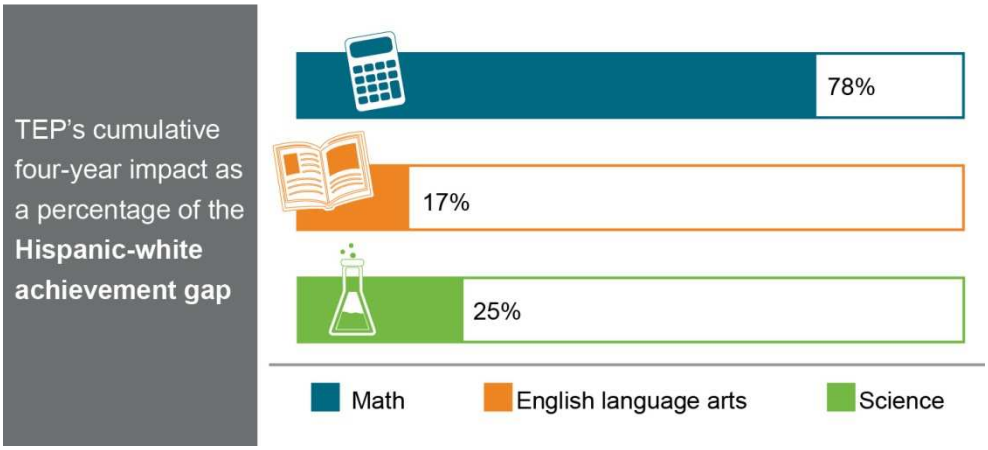
Note: This figure converts the effect sizes presented in Tables IV.1 and IV.2 to years of learning using the following benchmarks derived in Bloom et al. (2008) for average annual student achievement gains: from the end of grade 4 to the end of grade 8 (for four-year impacts)—0.40 standard deviations (SDs) in math, 0.30 SDs in ELA, 0.30 SDs in science; from the end of grade 4 to the end of grade 7 (for three-year impacts)—0.42 SDs in math and 0.32 SDs in ELA; from the end of grade 4 to the end of grade 6 (for two-year impacts)—0.49 SDs in math and 0.36 SDs in ELA; and from the end of grade 4 to the end of grade 5 (for one-year impacts)—0.56 SDs in math and 0.40 SDs in ELA. Effects are normalized such that the average annual achievement gains made by comparison students during the period of treatment equal zero.

TEP's cumulative effect on student achievement over four years is about 78 percent of the Hispanic-white achievement gap in math, 17 percent in ELA, and 25 percent in science.

About 87 percent of TEP students are Hispanic, and the Hispanic-white achievement gap in 8th grade (Bloom et al. 2008) provides another metric for evaluating TEP's impact on achievement. For the 2009 cohort, TEP's cumulative four-year impact on student achievement

was about 78 percent of the Hispanic-white achievement gap in math, about 17 percent in ELA, and about 25 percent in science (Figure IV.2).

Figure IV.2. TEP cumulative four-year impacts as a percentage of Hispanic-white achievement gap



Note: For math and ELA, this figure presents the percentage created when 4-year effect sizes listed in Tables IV.1 and IV.2 are divided by the 8th-grade Hispanic-white achievement gap—presented as an effect size—in Bloom et al. (2008). Bloom et al. do not report a science achievement gap, but the authors verified the process necessary to calculate the gap: (1) subtract the mean for Hispanics from the mean for whites on a standardized test, (2) divide the difference by the score standard deviation. This procedure was conducted on National Assessment of Educational Progress (NAEP) science scale score for grade 8 in 2011, yielding an achievement gap of 0.76. The figure presents the percentage created when the 4-year effect size in science (0.19) was divided by 0.76.

Results are robust across alternative specifications.

These achievement impact findings are largely unaffected by alternative specifications of the matching process and outcome measures (see Appendix C for detailed findings). To examine the sensitivity of the findings to alternative approaches, we also estimated impacts using (1) a different analysis sample without matching, (2) two alternative matching methods, and (3) standard *z*-scores rather than rank-based *z*-scores.

V. DISCUSSION

The findings in this report indicate that a more mature TEP positively affected students' achievement, and the effects were often substantial. During TEP's first two years, TEP students typically had lower achievement than similar comparison students, but these deficits were subsequently erased and reversed: TEP's impacts were almost uniformly positive during TEP's third and fourth years. We found no evidence that student selection or selective retention drove these impacts: the students admitted to TEP were low-achieving and disadvantaged, similar to other NYC students in the neighborhood, and TEP's student attrition rates were similar to those of similar schools.

A few teachers can strongly influence TEP's impacts, as TEP's underlying philosophy expects. All TEP students in a grade were typically taught math and ELA classes by only one or two teachers; consequently, TEP's impacts on specific cohorts in specific grades indicate the impacts of at most a few teachers. For example, the same teacher taught math to TEP's first cohort in both 7th and 8th grades, during which time the math achievement for this cohort improved dramatically relative to the comparison group (0.87 standard deviation units, slightly larger than the black-white or Latino-white test score gap reported in Bloom et al. 2008).[41] The importance of individual teachers does not mean, however, that the estimated impacts should be discounted as good luck, because hiring and developing effective teachers is the most important component of the TEP model.

Caution is warranted in drawing broader implications from TEP's success to date. TEP's approach of focusing resources and attention on teachers appears potentially scalable, but the impacts reported in this study are for only one school—proof-of-concept rather than proof that a model can be consistently effective across multiple schools. TEP's observed impacts on students' achievement reflect not only TEP practices but also TEP staff, and any interactions between the two. For example, we cannot distinguish the effect of TEP's founder and principal from the effect of TEP. Moreover, TEP is still changing (for example, the school has gone to a year-round calendar)[42] and the reported results are for a maturing school, not a mature school.

The rest of this chapter discusses the implications of the study findings and questions for future research.

A. Implications

New charter schools can improve. TEP might have needed time to develop a mature instructional approach. Although one widely cited study has found that, on average, new charter schools do not improve (Peltason 2013), several other studies have found that older charter schools are more effective than younger schools (Bifulco and Ladd 2006; Lavertu and Witte 2009; Carruthers 2012). TEP is a completely new school with a completely new model (rather

[41] The higher achievement could, of course, result from factors other than the teacher—TEP students had low achievement at the end of 6th grade, TEP implemented a new discipline policy in the summer of 2011, and TEP students also received 30 minutes of math tutoring daily starting in 2011–2012.

[42] Instead of one six-week summer institute, TEP now has three development institutes throughout the school year—two weeks during the summer, one week during the autumn, and one week during the spring.

than, for example, the fifth school opened by a charter management organization with an established model). TEP's new principal—who had not served in a leadership role in a school—might have needed some time to refine a novel educational model.

Even with a prioritized and intensive process, hiring effective teachers is challenging. TEP devotes substantial resources to identifying and developing effective teachers. TEP's high annual teacher termination rate suggests that the hiring process has not been fully successful in identifying effective teachers (although it might be more effective than other existing approaches). The scale of hiring—TEP had to hire a whole new grade of teachers each year plus replacements—might have overwhelmed the process. To make the school more sustainable and scalable, TEP has to improve at identifying teachers who fit with the TEP approach and who are effective teaching TEP students.

TEP is more ambitious than other incentive programs. TEP resembles other policy approaches, such as the U.S. Department of Education's Teacher Incentive Fund (TIF), that encourage strong teachers to teach in schools with low-income and low-achieving students. However, compared with most TIF-funded programs, TEP teachers receive a much larger financial reward for teaching at TEP, the rewards exist as long as the teacher is at TEP, TEP does not require any additional resources, and TEP teachers have more responsibilities and receive more development.

Currently, TEP attracts and rewards a few excellent existing teachers, but wider implementation could cause more systemic changes by attracting new teachers to the profession. TEP's principal focuses on hiring teachers who have successfully taught at schools with low-achieving and disadvantaged populations. Consequently, it is more an approach for rewarding excellent teachers at traditionally hard-to-staff schools than improving education broadly. However, scaling up the TEP approach—high salaries and more responsibilities—might have broader impacts by increasing the number of quality people willing to teach and drawing them away from other professions. There is no way to assess the possibility of such an impact without actually creating TEP-like schools on a larger scale.

B. Future research

This report rigorously estimated TEP's impacts on students' achievement during its first four years. The positive findings raise additional, important questions that future studies can address.

Does TEP's intensive development process improve teachers' performance?

This study measured school effectiveness and did not examine the extent to which TEP hired and developed effective teachers. A study that estimated teachers' value-added for TEP's teachers before and after they were hired at TEP could identify the effectiveness of new TEP teachers and how their effectiveness changes at TEP. Research that examined how much intensive development affected students' achievement would help identify where to focus improvement efforts.

What are the long-term impacts of the TEP intervention on outcomes such as attainment?

Recent research has found that charter schools and school choice can affect nonachievement outcomes, such as attainment (highest level of schooling completed) and criminal behavior (Furgeson et al. 2012; Angrist et al. forthcoming; Deming et al. 2011; Wolf et al. 2010; Dobbie and Fryer 2012). A comprehensive study of TEP would examine long-term impacts on high school graduation, college entry, and postsecondary degree completion.

Which replicable TEP practices differ from those of comparison schools?

This report did not measure how TEP practices differed from comparison schools. Qualitative tools such as surveys and observations at TEP and neighboring schools would identify the replicable components of the TEP model. Identifying the practices at comparison schools would also provide insight about the educational contexts where TEP's practices were effective.

REFERENCES

Ahmed-Ullah, N. S. and A. Richards. "CPS: Expulsion Rate Higher at Charter Schools." *Chicago Tribune*. February 26, 2014.

Angrist, J. A., Cohodes, S. R., Dynarski, S. M., Pathak, P. A., & Walters, C. D. "Stand and Deliver: Effects of Boston's Charter High Schools on College Preparation, Entry, and Choice." *Journal of Labor Economics,* Forthcoming.

Angrist, J. D., G. W. Imbens, and D. B. Rubin. "Identification of causal effects using instrumental variables." *Journal of the American Statistical Association*, vol. 91, no. 434, 1996, pp. 444-455.

Bifulco, R. and H.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.

Bloom, H. S., C. J. Hill, A. R. Black, & M. W. Lipsey. "Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions." *Journal of Research on Educational Effectiveness*, vol. 1, no. 4, 2008, pp. 289-328.

Carruthers, C.K. "New Schools, New Students, New Teachers: Evaluating the Effectiveness of Charter Schools." *Economics of Education Review*, vol. 31, no. 2, 2012, pp. 280–292.

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

Deming, D. J., J. S. Hastings, T. J. Kane, and D. O. Staiger. "School Choice, School Quality and Postsecondary Attainment." National Bureau of Economic Research, Working Paper No. 17438, 2011.

Fortson, K., N. Verbitsky-Savitz, E. Kopa, and P. Gleason. "Horseshoes, Hand Grenades, and Treatment Effects? Reassessing Bias In Nonexperimental Estimators." Mathematica Policy Research working paper, March 2013.

Furgeson, J., B.P. Gill, J. Haimson, A. Killewald, M. McCullough, I. Nichols-Barrer, B. Teh, N. Verbitsky-Savitz, M. Bowen, A. Demeritt, P. Hill, and R. Lake. "Charter-School Management Organizations: Diverse Strategies and Diverse Student Impacts." Washington, DC: Mathematica Policy Research and Center on Reinventing Public Education, January 2012.

Gill, B.P., J. Furgeson, H.S. Chiang, B. Teh, J. Haimson, N. Verbitsky-Savitz. "Replicating Experimental Impact Estimates with Nonexperimental Methods in the Context of Control Crossover." Mathematica Policy Research working paper, October 2013.

Gill, B.P., L. Hamilton, J.R. Lockwood, J. Marsh, R. Zimmer, D. Hill, and S. Pribesh. "Inspiration, Perspiration, and Time: Operations and Achievement in Edison Schools (MG-351-EDU)." Santa Monica, CA: RAND Corporation, 2005.

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

Gootman, Elissa. "At Charter School, Higher Teacher Pay." *New York Times*, March 7, 2008, A1.

Government Accountability Office. "Charter Schools: Additional Federal Attention Needed to Help Protect Access for Students with Disabilities." Report to Congressional Requesters. Washington, DC: United States Government Accountability Office, 2012.

Hanushek, E. A., J.F. Kain, and S.G. Rivkin. "Disruption versus Tiebout Improvement: The Costs and Benefits of Switching Schools." *Journal of Public Economics*, vol. 88, 2004, pp. 1721–1746.

Lavertu, S. and J. Witte. "The Impact of Milwaukee Charter Schools on Student Achievement." *The Brookings Institution's Issues in Governance Studies*, no. 23, 2009.

Marinell, W. H. "The Middle School Teacher Turnover Project: A Descriptive Analysis of Teacher Turnover in New York City's Middle Schools." Research Alliance for New York City Schools, 2011.

Miron, Gary, Jessica Urschel, and Nicholas Saxton. "What Makes KIPP Work? A Study of Student Characteristics, Attrition, and School Finance." Western Michigan University, March 2011.

Ni, Y. "The Sorting Effect of Charter Schools on Student Composition in Traditional Public Schools." *Educational Policy*, vol. 26, no. 2, 2012, pp. 215-242.

Peltason, E. H. and M. E. Raymond. "Charter School Growth and Replication, Vol. 1." Center for Research on Education Outcomes, 2013.

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

Wolf, Patrick, Babette Gutmann, Michael Puma, Brian Kisida, Lou Rizzo, Nada Eissa, and Matthew Carr. "Evaluation of the DC Opportunity Scholarship Program: Final Report (NCEE 2010-4018)." Washington, DC: U.S. Department of Education, Institute for Education Sciences, National Center for Education Evaluation and Regional Assistance, 2010.

APPENDIX A: DATA AND METHODS

Our descriptive and impact analyses used annual student-level administrative data obtained from the New York City (NYC) Department of Education (DOE). The data covered the 2006–2007 through 2012–2013 school years and included all NYC students in the relevant grades (as well as one grade higher and lower to track students who are retained or who skip a grade). Each student had a consistent, unique ID to permit longitudinal analyses. For each student, the data included grade-specific state assessment scores in ELA and mathematics (and science for 4th and 8th graders) and any test accommodations for each year the student was enrolled in a public school in NYC. The data also include annual indicators for each student's gender, race/ethnicity, special education status, English language learner (ELL) status, free or reduced-price lunch status, and home language. Finally, for every school a student attended, the data provided the date of enrollment and, when relevant, the date of discharge.

A. Primary treatment indicator and outcomes

The four cohorts followed in this report entered The Equity Project (TEP) in the 5th grade in fall 2009, fall 2010, fall 2011, and fall 2012.[43] To account for possible bias due to selective student attrition at TEP,[44] we classified students as TEP students if they enrolled at TEP for at least three days during the 5th grade.[45] These students are classified as part of the TEP treatment group regardless of whether they left TEP and enrolled at another school during the 5th-grade year or after completing the 5th-grade year.[46] This approach, analogous to an experimental intent-to-treat approach, produced a conservative estimate—biased toward zero—of the impact of continuously enrolling in TEP.

The primary achievement outcomes were standard scores on state English/language arts (ELA) and math tests that have been converted to reduce the influence of unreliable outliers.[47] An examination of the math and ELA achievement test scores revealed several outliers that could distort results. For example, in 2011, three TEP students in the 2009 cohort had math scores that were more than five standard deviations less than the district mean. Test documentation indicated that these standard scores were unreliable, with a standard error of measurement (SEM) that was 10 times as large as the SEM for other parts of the score distribution.

[43] Five students from the 2009 cohort, 10 students from the 2010 cohort, and 5 students in the 2011 cohort entered TEP in the 6th or 7th grades. These students were excluded from the impact estimates because they received a different level of exposure to TEP than the rest of their cohort. As described in Chapter III, these students were similar to the students they replaced on observable characteristics.

[44] Student attrition from TEP was lower than similar schools (see Chapter III).

[45] We require three days of enrollment to account for delays in the NYC enrollment data. Specifically, students who enrolled at TEP but then withdrew before school started are often identified in the data as withdrawing on the first, second, or third day of school even though they never attended TEP.

[46] Students who left the NYC data entirely did not have observed outcomes and were not included in the analysis.

[47] Test scores were standardized by subject, grade, and year using information from the entire district sample of students. Science tests were administered in the 8th grade in New York.

To address these unreliable outliers, we converted pre-baseline, baseline, and outcome ELA and math test scores to rank-based *z*-scores (Gill et al. 2005). Within each year, grade, and test subject combination, we first assigned percentile ranks to each scale score in the NYC district distribution. (Because each scale score can span multiple percentiles, we assigned the average percentile.) We then mapped each percentile rank to a standardized normal score, or *z*-score, using the inverse of the normal cumulative distribution function. For each year, grade, and subject combination, the mean rank-based *z*-score is 0 and the standard deviation is 1. As a sensitivity check, we report estimates using conventional *z*-scores in Appendix C. Those results do not substantially differ from the primary results.

A few students in each TEP cohort did not have outcomes and cannot be included in the analyses (Table A.1). Potential comparison students who did not have outcomes were also excluded.

Table A.1. TEP original and analysis samples for achievement impacts (maximum duration)

	2009 Cohort	2010 Cohort	2011 Cohort	2012 Cohort
Ever attended TEP (original sample)	126	124	120	122
Analysis sample for math outcomes	113	115	116	118
Analysis sample for ELA outcomes	114	115	116	118

Notes: For the 2009 cohort, the primary outcomes are the four-year math and ELA achievement test scores. For the 2010 cohort, the primary outcomes are the three-year math and ELA achievement test scores. For the 2011 cohort, the primary outcomes are the two-year math and ELA achievement test scores. For the 2012 cohort, the primary outcomes are the one-year math and ELA achievement test scores.

B. Methods for estimating impacts

In principle, TEP impacts can be estimated experimentally and quasi-experimentally. Experimental estimates use TEP's random admission lottery to identify a treatment group (admitted at the time of the lottery) and a control group (not admitted at the time of the lottery).

The lottery-based approach provides the greatest internal validity for purposes of causal inference, but has several limitations when using TEP's lotteries. The experimental analysis includes only TEP students admitted through the lottery, whose parents consented, and who did not have a 0 or 100 percent probability of admission (see Appendix B for more information). Because the experimental analysis must be based on assignment at the time of the lottery, many admitted students do not enroll and many nonadmitted students do enroll after they are admitted off the waiting list (see Appendix B). To account for substantial control crossover in the experimental analysis, the experimental estimates would have to use instrumental variables to estimate a treatment-on-treated estimate that pertains only to students who comply with their treatment assignment (Angrist et al. 1996), a small fraction of TEP students.

Given the limitations of the lottery-based approach, we decided before conducting analyses to rely exclusively on a quasi-experimental matching approach. Recent research has found that charter school impact estimates based on matching methods are very similar to experimental estimates as long as pre-treatment measures of the outcome of interest are used in the analysis

(Fortson et al. 2013; Gill et al. 2013). We use 3rd- and 4th-grade math and ELA test scores as pre-treatment matching variables and statistical controls. The next section describes the primary matching approach and model and Appendix C presents several sensitivity analyses and findings.

Propensity-score models identify comparison groups.

We estimated propensity-score models to identify students with a similar probability of enrolling at TEP (Rosenbaum and Rubin 1983), resulting in a comparison group of students with similar characteristics and prior achievement. Theoretically, propensity-score impact estimates are unbiased if the propensity scores accurately model TEP enrollment (that is, enrollment is unrelated to outcomes when controlling for baseline variables). In practice, propensity-score estimates using baseline achievement and demographic characteristics of charter impacts are very similar to experimental estimates (Fortson et al. 2013; Gill et al. 2013). Our propensity scores are estimated using baseline (4th-grade) test scores, pre-baseline (3rd-grade) test scores, and multiple demographic characteristics.

To further control for selection, we limited the potential comparison group to students who attended a school that was also attended by future TEP students during the 4th grade (a neighborhood school). By restricting the comparison group to neighborhood schools, we accounted for the location and type of school that a student attended before enrolling at TEP, an important consideration in nonexperimental evaluations (Cook et al. 2008). Because baseline versions of the outcome measure are crucial controls for selection, the analysis sample was also limited to students who had at least one baseline test score (ELA or math). Conditional on the students having outcomes, this requirement excluded six TEP students (one each in the 2009, 2010, and 2011 cohorts and three in the 2012 cohort) and a few hundred potential comparison students.[48]

To identify a matched comparison group, we estimated a logistic regression model (a propensity-score model) that predicts whether a student enrolls at TEP in 5th grade. The sample for the model pooled all four TEP cohorts. We performed a stepwise model selection procedure to identify the baseline characteristics that resulted in the best model fit. We expected baseline test scores to be strong predictors of outcome test scores (R^2 typically greater than 0.50) and therefore required that the model include pre-baseline—3rd grade—and baseline math and ELA test scores as well as missing test score indicators. Values for missing baseline test scores were imputed for the estimation of the propensity-score model. For baseline and pre-baseline test scores, we included a missing data indicator and set each missing test score to the state- or district-level mean, which is zero by design.[49] The remaining covariates were added or removed at each step in the model selection procedure to select the best fitting model; a significance threshold was set such that the selection procedure automatically dropped covariates with a *p*-value exceeding 0.20. The following characteristics were specified for this stepwise procedure:

[48] To be eligible, potential comparison students had to have outcomes for all years.

[49] When a baseline demographic characteristic was missing for a student, we set the baseline value equal to the student's most recent nonmissing value for the characteristic. When a baseline demographic characteristic was inconsistent across years of data for a student, we set the baseline value equal to the most common value across the years of data for the student. Ultimately, no TEP students or students in the matched comparison sample were missing values for any baseline characteristics.

baseline test accommodation indicator; sex; race/ethnicity; free or reduced-price lunch status; individualized education program status (special education); English learner status, home language (English, Spanish, or other); an indicator for whether the student was older than statutory age; two-way interaction terms among the covariates; and quadratic test score terms.

We used the Hosmer and Lemeshow (H-L) Goodness-of-Fit test and Akaike information criterion (AIC) to assess and compare model fit and make necessary adjustments in an iterative process. Such adjustments primarily involved combining cells with very few students; for example, because only one student who identified as neither white nor Hispanic nor African American ever attended TEP across the four cohorts, we collapsed white, non-Hispanic, and non-African American into a single racial/ethnic category. The final model maximized the AIC and yielded an H-L Goodness-of-Fit test *p*-value of 0.86, which indicates that the model fit the data well. (The null hypothesis for the H-L Goodness-of-Fit test is that after organizing students into deciles based on the probability of TEP enrollment predicted by the model, the number of students in each decile expected to enroll based on the model is not different from the observed frequency of TEP enrollment; higher *p*-values for the test indicate better model fits.)

Using the final best-fitting model, we calculated propensity scores for TEP entry for all students in the relevant cohort at neighborhood schools. The propensity scores measure a student's probability of enrolling at TEP and were used to select a matched comparison group. To increase statistical power to detect impacts and improve statistical precision, we allowed each TEP student to be matched to as many as 10 non-TEP comparison students with the closest propensity scores (most similar probabilities of enrolling at TEP) within a specified caliper (range) of 10^{-3}. To reduce bias and allow the best quality of match for each TEP student, we matched with replacement, allowing each comparison student to match to more than one TEP student. We conducted this matching procedure separately for each cohort.

We constructed analysis weights to account for each TEP student potentially having multiple matched comparison students and matched comparison students potentially being matched to more than one TEP student. Each TEP student was assigned a weight equal to one. When a TEP student had only one matched comparison student, the comparison student's weight was also one; when more than one comparison student was matched to a TEP student, the TEP student's weight was divided equally among the matched comparison students. When a comparison student was matched to multiple treatment students, the comparison student was assigned an analysis weight equal to the sum of the weights or weight-shares of all TEP students to whom he or she was matched. To facilitate interpretation, we then rescaled the weights to reflect the total number of TEP and matched comparison students in the analytic sample.

The matching process identified comparison students who are similar to TEP students. (We examined baseline differences without using any imputed values.) There were no significant differences at baseline for the 2009 and 2012 cohorts (see Table A.2). Among the students in the analytic sample for the 2010 cohort, TEP students were, on average, significantly more likely to use Spanish as their primary language at home. This difference remains significant regardless of whether the indicator for speaking Spanish as the primary home language is required to be included in the stepwise propensity model selection procedure. Among the students in the analytic sample for the 2011 cohort, TEP students were, on average, significantly more likely to be male. No other differences were statistically significant. When making comparisons for 12

student characteristics across four cohorts, the two statistically significant differences (of 48 possible differences) were no more than would be likely to occur by chance (that is, as false positives).

Table A.2. Baseline equivalence of analytic sample, by cohort

	2009 Cohort			2010 Cohort			2011 Cohort			2012 Cohort		
	TEP	Comp.	Diff.	TEP	Comp.	Diff.	TEP	Comp.	Diff.	TEP	Comp.	Diff.
4th-grade math achievement	-0.29	-0.27	-0.02	-0.21	-0.26	0.05	-0.41	-0.38	-0.03	-0.35	-0.33	-0.02
4th-grade ELA achievement	-0.27	-0.32	0.05	-0.21	-0.26	0.04	-0.40	-0.35	-0.05	-0.44	-0.39	-0.05
3rd-grade math achievement	-0.27	-0.30	-0.03	-0.24	-0.22	-0.02	-0.42	-0.40	-0.02	-0.36	-0.38	0.02
3rd-grade ELA achievement	-0.29	-0.29	0.00	-0.25	-0.30	0.05	-0.37	-0.31	-0.06	-0.34	-0.33	-0.01
Male	0.50	0.53	-0.03	0.44	0.51	-0.06	0.58	0.46	0.11*	0.47	0.53	-0.05
Subsidized lunch	0.88	0.86	0.02	0.97	0.96	0.01	0.97	0.97	0.00	0.96	0.95	0.01
Hispanic	0.89	0.87	0.02	0.88	0.84	0.04	0.82	0.88	-0.06	0.92	0.93	-0.02
African American	0.11	0.10	0.01	0.09	0.14	-0.05	0.16	0.11	0.05	0.06	0.05	0.01
English learners	0.30	0.29	0.00	0.38	0.36	0.02	0.36	0.35	0.02	0.27	0.33	-0.06
Special education	0.18	0.20	-0.02	0.16	0.16	0.00	0.19	0.18	0.01	0.14	0.16	-0.02
Spanish home language	0.53	0.65	-0.07	0.77	0.66	0.12*	0.61	0.68	-0.07	0.74	0.67	0.06

Note: ELA and math scores are rank-based *z*-scores standardized using the district means and standard deviations. All statistics are weighted using the analysis weights derived from the matching procedure. Due to rounding, some values in the difference columns might not equal the difference between the values in the TEP and Comp. columns. NYC changed coding for some characteristics between 2008–2009 and 2009–2010, and any differences between cohorts could be a function of different coding.

*Significantly different from zero at the 0.05 level, two-tailed test.

Impact regression models include baseline covariates to control for any remaining differences.

We estimated impacts using a regression model that included covariates to control for any remaining observable baseline differences. The following model estimated the impact of attending TEP for each cohort:

(1) $y_i = \alpha + X_i\beta + \theta TEP_i + \epsilon_i$

where y_i is the outcome rank-based test score for student i; X_i is a vector of pre-baseline and baseline test scores and baseline characteristics identified in advance of analysis;[50] TEP_i is the

[50] We aimed to maximize power by excluding extraneous characteristics that did not improve the model. Before analysis, using all students in the relevant cohorts at neighborhood schools (N = 10,194), we estimated regressions of the two dependent variables on baseline achievement, pre-baseline achievement, baseline attendance rate, and all characteristics included in the initial propensity-score model. We chose to include all characteristics that were statistically significant in at least one of the models–all of the characteristics other than whether a student attended a

treatment indicator for whether the student ever enrolled at TEP; β and θ are estimated parameters (θ estimates TEP's impact); and ε_i is an error term.[51] To account for possible heteroskedasticity (scores at the tails of the distribution have much larger standard errors), we estimated White's standard errors. Students were weighted using the analysis weights described.

Grade repetition presents one challenge for this analytic approach. A student who repeats a grade takes a different state assessment than do other students in her or his original cohort. If TEP retains students at a higher rate relative to comparison schools and we drop grade repeaters from our analyses, impact estimates might be biased.[52] At the same time, not accounting for grade repetition ignores that grade repeaters have one additional year to learn material covered by the state assessments in the repeated grade. We address this issue by following the approach taken in Furgeson et al. (2012), assuming that retained students perform at the same level relative to other students in their cohort in the repeated year as they did in the year before being retained.

We estimated all impacts separately by cohort and by length of treatment and present all eight combinations of cohort-length of treatment impacts as our primary findings. Although the earliest TEP cohorts (2009 and 2010) are important to examine because they have experienced the most years of TEP, TEP's NYC DOE school report card grades have improved over time, suggesting that more recent cohorts are more likely to represent TEP's future impacts. We do not estimate impacts per year of enrollment in TEP, because charter school impacts might not be linear by year (Furgeson et al. 2012; Gleason et al. 2010).

(continued)

charter school at baseline (only one TEP student across all cohorts attended a charter school at baseline) and English learner status (which is highly correlated with home language indicators).

[51] We assumed fixed site effects rather than clustering at the site level, because we are not making inferences about impacts of schools outside of our matched sample.

[52] In the 2009 cohort, three TEP students repeated grade 6 and two other students repeated grade 7. In the 2010 cohort, one TEP student repeated grade 6. In the 2011 cohort, four TEP students repeated grade 5.

APPENDIX B: LOTTERY-BASED IMPACT ANALYSIS

Studies often use charter school admission lotteries—which randomly admit students—as natural experiments to estimate school impacts on students' achievement (for example, Gleason et al. 2010; Angrist et al. 2014). Although The Equity Project (TEP) conducted admission lotteries to admit each 5th-grade class, we chose not to use the lotteries to estimate impacts for two reasons. First, many TEP students would be excluded from a lottery-based analysis because they applied after the lottery, participated in the lottery but were not admitted through it (for example, siblings of TEP students), or did not provide parental consent for Mathematica Policy Research to obtain identified students' data. Second, there was substantial treatment nonparticipation (students admitted who did not attend TEP) and control crossover (students admitted after the lottery who attend TEP). Consequently, any impacts estimated using TEP's lottery would not be representative of TEP's overall impact on all students. Instead, as described in Appendix A, we decided to use the propensity-score matching approach to estimate TEP's impacts on achievement.

In this appendix, we first describes TEP's admission lottery. In the following two sections, we describe why lottery-based impact estimates would not apply to most TEP students.

A. TEP's admission lotteries

At TEP's admission lotteries, conducted in April, folded index cards with each 5th-grade applicant's name and lottery information were randomly selected, one at a time, from a Plexiglas tumbler that was repeatedly spun. Each card had the name of a student who submitted an application to TEP before the lottery. An individual unaffiliated with TEP selected cards until all cards were drawn. The principal announced the name on each card after it was drawn and two TEP staff recorded the information. The information was projected onto a screen for parents and students who attended the lottery. Mathematica observed the 2009 and 2010 lotteries.

In each year, 120 students were admitted at the lottery, and the lottery determined the waiting list order for the remaining students. Mathematica obtained lottery records from TEP immediately after the lottery; the records obtained from TEP were consistent with Mathematica's records for the two lotteries Mathematica observed.

TEP's lottery process favored students at risk of academic failure,[53] students who live in New York City (NYC) School District 6 (the geographic area where TEP is located), and siblings of TEP applicants or TEP students. At each lottery, TEP admitted students in the following order:

1. Applicants at risk of academic failure from any NYC school district were admitted to fill 30 percent of the total seats (36 seats). The first at-risk student whose card was drawn was

[53] TEP defines a student as at risk of academic failure if he or she scored at *below proficient* or *well-below proficient* on the most recent New York State English Language Arts or math examination for which that student's score is available.

assigned lottery #1, the second was assigned lottery #2, and so on, until the last at-risk student admitted as an at-risk student was assigned lottery #36.[54]

2. Applicants residing in NYC School District 6 (regardless of at-risk status) were admitted in order of their cards being drawn.

3. After all students from NYC School District 6 were admitted in step 2, students residing in any NYC school district were admitted in order of their cards being drawn.[55]

The selection process also included preferences for siblings who were enrolled at TEP or applying to the same grade at the same time. In 2010, 2011, and 2012, siblings of enrolled TEP students were automatically admitted (to either at-risk or District 6 slots, as appropriate). Siblings applying together for 5th-grade seats at the lottery also received a preference. When one sibling was admitted, the other sibling automatically received the next lottery number in either the at-risk or District 6 lottery. Applicants with a sibling applying in the lottery could therefore be admitted by having their card selected early or could "win-by-sibling" if their card was not drawn but their sibling was admitted.[56]

B. TEP students ineligible for lottery-based analysis

For all school years except 2009–2010, students who applied after the lottery were admitted to TEP and attended during the 5th grade. Because these late applicants did not participate in the lottery—they applied afterward and were added to the waiting list—these TEP students were not randomly assigned admission to TEP and would be excluded from a lottery-based impact analysis based on the lottery.

Moreover, two types of TEP students who participated in the lottery would be excluded from a lottery-based analysis: (1) students whose probability of admission was 0 or 100 percent and (2) students whose parents declined to provide Mathematica with consent to participate in the study before the lottery or whose consent form TEP could not locate.[57]

1. **0 or 100 percent admission probability.** Some TEP students had a 0 percent probability of admission at the lottery (students not at-risk and outside District 6 were never admitted at the lottery) or 100 percent probability of admission (siblings of existing TEP student were

[54] When 36 at-risk students have been admitted, at-risk District 6 students can then be admitted through the District 6 lottery (if any District 6 seats remain). These lottery numbers were not the same as the order in which the cards were drawn. For example, if the first card drawn was a not-at-risk District 6 student, her lottery number was 37, because there were 36 at-risk seats. Students with lottery numbers less than or equal to 120 were admitted at the lottery.

[55] In all years, 2009 to 2012, none of these students were admitted at the time of the lottery.

[56] The lottery complications mean that applicants have differing probabilities of being admitted to TEP (for example, at-risk students from District 6 have a different probability of admission than not-at-risk students from District 6). A lottery-based impact analysis must account for these differences in probabilities because admission offers are random only within groups of students who have the same probability of admission (conventionally labeled risk sets).

[57] There were also three applicants who were not in the 4th grade at the time of the lottery. These students were also ineligible for a lottery-based analysis, and none attended TEP in the year after the lottery.

always admitted at the lottery) at the time of the lottery. To be included in a lottery-based analysis, students must have had an admission probability between 0 and 1.

2. **Declined or lost consent.** Some participants were excluded from the records collection and analysis because their parents did not consent for Mathematica to obtain their student records. In a few cases each year, TEP could not locate the consent forms. (TEP collected the parental consent forms for Mathematica.)

Due to these exemptions, 69 to 94 percent of the TEP students in each cohort were eligible for a lottery-based analysis (see second row in Table B.1).

C. Low compliance with lottery assignment

For the 2010, 2011, and 2012 cohorts, treatment for a lottery-based analysis must be defined as receiving an admission offer at the time of the lottery, because all students who participated in the lottery were eventually admitted to TEP (that is, there would be no control group if the treatment was defined as being ever admitted to TEP).[58] Although valid, the approach creates substantial control crossover because many students not admitted at the lottery were later admitted off the wait list and attended TEP. This crossover requires a treatment-on-treated impact estimate because the intent-to-treat estimate does not represent the impact of actually enrolling in TEP when there is control crossover (or when some treatment group students do not enroll).

This treatment-on-treated estimate estimates TEP's effect only on the subgroup of TEP students who would enroll only if they were admitted at the lottery and would not enroll if they were admitted after the lottery (Angrist et al. 1996). In contrast, our primary propensity-score impact estimate pertains to all TEP students included in the analysis. To estimate the coverage of the estimates, we calculated the approximate number of TEP students to whom a lottery-based, treatment-on-treated impact estimate pertains.[59]

The number of TEP students to whom a treatment-on-treated estimate pertained varied from 24 students for the 2009 cohort to 10 for the 2012 cohort (see third row in Table B.1).[60] Although a lottery-based estimate would pertain to 8 to 19 percent of TEP students in each cohort, our primary propensity-score analysis pertained to 97 or 98 percent of TEP students in each cohort.

[58] For the 2009 lottery, 51 lottery participants were never admitted to TEP.

[59] The impact estimate pertains to the students who complied with random assignment, conventionally known as *compliers*. *Always-takers* are TEP students who would have enrolled regardless of whether they were admitted at the lottery (they receive treatment regardless of whether they were assigned to receive treatment). Not all TEP students in the treatment condition were compliers, because some would have enrolled in treatment even if they were not assigned to treatment (always-takers). To estimate the percentage of compliers, we first identified the percentage of always-takers as the percentage of control students who enrolled at TEP. (These students enrolled even though they were not admitted at the lottery. Because of random assignment, the percentage of always-takers should be the same in expectation for treatment and control groups.) By subtracting this always-takers percentage from the percentage of treatment students who enrolled (compliers and always-takers), we obtained an estimate of the percentage of compliers. Multiplying this complier percentage by the number of eligible students assigned to treatment provided an estimate of the number of TEP students to whom the impact estimate pertained.

[60] The small differentials between treatment and control group enrollment at TEP would also significantly reduce the statistical power of a lottery-based impact analysis.

Table B.1. Percentage of TEP students included in lottery-based and primary analysis, by cohort

	Total	2009 cohort	2010 cohort	2011 cohort	2012 cohort
TEP students enrolled in the 5th grade	492	126	124	120	122
TEP students eligible for a lottery-based analysis	*390*	*119*	*85*	*100*	*86*
Estimated number of eligible TEP students to whom treatment-on-treated impact estimate pertains	*71*	*24*	*13*	*22*	*10*
Approximate percentage of TEP students to whom a lottery-based treatment-on-treated impact estimate pertains	**14**	**19**	**11**	**18**	**8**
Percentage of TEP students included in primary propensity-score analysis	**98**	**98**	**98**	**98**	**97**

Note: These percentages include TEP students who would be included in any impact analysis examining outcomes after any period. For example, for the 2009 cohort, the analysis includes any TEP student who had an outcome in the 5th, 6th, 7th, or 8th grade. The percentage who would be included in any specific impact analysis—grade 6 analysis for 2009 cohort—would be equal to or smaller than this percentage. The estimated number of eligible TEP students to whom treatment-on-treated impact estimate pertains was calculated by multiplying the number of eligible treatment students (students admitted at the lottery) by the percentage of students who *comply* with assignment (estimated as the percentage of treatment students who attend TEP minus the percentage of control students who attend TEP).

APPENDIX C: SENSITIVITY ANALYSES

This appendix presents impacts of The Equity Project (TEP) estimated using four alternative specifications of the achievement measures and matching process: (1) impact models involving all TEP students and all students in neighborhood schools and controlling for baseline differences statistically, rather than limited to only the matched comparison group; (2) a matching process using nearest-neighbor matching (without replacement) in which each TEP student is matched to the comparison student with the closest propensity score;[61] (3) a matching process using kernel density matching in which each TEP student is matched to a weighted average of all comparison students;[62] and (4) impact models with pre-baseline, baseline, and outcome achievement measured in standard *z*-scores rather than rank-based *z*-scores.[63]

The primary impacts are largely robust to alternative specifications.

When impact estimation used only statistical controls rather than matching—an ordinary least squares (OLS) model—estimated impacts were similar to the primary impact estimates in magnitude and identical in statistical significance (Tables C.1, C.2, and C.3).

[61] The propensity scores estimated for each TEP and potential comparison student were identical to the propensity scores estimated in the primary impact analyses. However, the matching process consisted of matching each TEP student to a single comparison student with the nearest propensity score. In addition, unlike the primary matching process in which a comparison student could be matched to multiple TEP students, the matching was done without replacement: Each comparison student was matched to only one TEP student.

[62] The propensity scores estimated for each TEP and potential comparison student were identical to the propensity scores estimated in the primary impact analyses. However, each impact estimate compares the observed achievement test score of each treatment student with a weighted average of all comparison students within the area of common support. The weight assigned to a comparison student is determined by a measure of the distance between the comparison student's and the treatment student's propensity score, with the lowest weights assigned to the comparison students farthest from the treatment student.

[63] The matching process used standard *z*-scores for pre-baseline and baseline achievement measures but was otherwise identical to the primary matching process.

Table C.1. TEP impacts on math achievement, by cohort and duration: Neighborhood school OLS model

	1 year after enrolling at TEP	2 years after enrolling at TEP	3 years after enrolling at TEP	4 years after enrolling at TEP
2009 entering 5th graders	-0.29** (0.05) N = 2,099	-0.24** (0.05) N = 2,095	0.30** (0.05) N = 2,089	0.61** (0.05) N = 2,091
2010 entering 5th graders	0.03 (0.04) N = 2,765	-0.11** (0.05) N = 2,760	0.19** (0.05) N = 2,759	
2011 entering 5th graders	0.03 (0.04) N = 2,845	0.15** (0.04) N = 2,844		
2012 entering 5th graders	0.17** (0.05) N = 2,480			

Note: This table reports the coefficients on linear regressions of standardized math test scores on an indicator variable for TEP enrollment in 5th grade. Separate models were run for each cohort and outcome year combination. The comparison group consists of all students from neighborhood schools who never enrolled in TEP. Regression controls include two years of baseline test scores in math, ELA, and science, as well as indicator variables for baseline demographic characteristics reported in Appendix A. Robust standard errors are reported in parentheses.

*Significantly different from zero at the 0.05 level, two-tailed test.

**Significantly different from zero at the 0.01 level, two-tailed test.

ELA = English/language arts; OLS = ordinary least squares.

Table C.2. TEP impacts of TEP on ELA achievement, by cohort and duration: Neighborhood school OLS model

	1 year after enrolling at TEP	2 years after enrolling at TEP	3 years after enrolling at TEP	4 years after enrolling at TEP
2009 entering 5th graders	-0.29** (0.05) N = 2,099	-0.10** (0.05) N = 2,095	0.09* (0.05) N = 2,089	0.11** (0.05) N = 2,092
2010 entering 5th graders	-0.20** (0.05) N = 2,765	0.04 (0.05) N = 2,760	0.10* (0.06) N = 2,759	
2011 entering 5th graders	-0.02 (0.04) N = 2,845	0.01 (0.05) N = 2,844		
2012 entering 5th graders	0.02 (0.05) N = 2,480			

Note: This table reports the coefficients on linear regressions of standardized ELA test scores on an indicator variable for TEP enrollment in 5th grade. Separate models were run for each cohort and outcome year combination. The comparison group consists of all students from neighborhood schools who never enrolled in TEP. Regression controls include two years of baseline test scores in math, ELA, and science, as well as indicator variables for baseline demographic characteristics reported in Appendix A. Robust standard errors are reported in parentheses.

*Significantly different from zero at the 0.05 level, two-tailed test.

**Significantly different from zero at the 0.01 level, two-tailed test.

ELA = English/language arts; OLS = ordinary least squares.

Table C.3. TEP impacts on science achievement: Neighborhood school OLS model

	4 years after enrolling at TEP
2009 entering 5th graders	0.22** (0.06) N = 2,085

Note: This table reports the coefficients on linear regressions of standardized science test scores on an indicator variable for TEP enrollment in 5th grade. Separate models were run for each cohort and outcome year combination. The comparison group consists of all students from neighborhood schools who never enrolled in TEP. Regression controls include two years of baseline test scores in math, ELA, and science, as well as indicator variables for baseline demographic characteristics reported in Appendix A. Robust standard errors are reported in parentheses.

**Significantly different from zero at the 0.01 level, two-tailed test.

ELA = English/language arts; OLS = ordinary least squares.

The second alternative specification, nearest-neighbor matching, resulted in largely similar impact estimates, but there were several differences in statistical significance (Tables C.4, C.5, and C.6). Estimates of impacts on third- and fourth-year English/language arts (ELA) achievement for the 2009 cohort, second-year math achievement and third-year ELA achievement for the 2010 cohort, and first-year math achievement for the 2012 cohort are no longer significant. The lack of statistical significance for these estimates is primarily due to less precision (larger standard errors); the estimated impacts were similar.

Table C.4. TEP impacts on math achievement, by cohort and duration: Nearest-neighbor matching

	1 year after enrolling at TEP	2 years after enrolling at TEP	3 years after enrolling at TEP	4 years after enrolling at TEP
2009 entering 5th graders	-0.29** (0.08) N = 222	-0.27** (0.08) N = 218	0.23** (0.08) N = 216	0.66** (0.08) N = 218
2010 entering 5th graders	-0.07 (0.07) N = 219	-0.11 (0.08) N = 215	0.17** (0.08) N = 214	
2011 entering 5th graders	-0.01 (0.07) N = 207	0.14* (0.08) N = 206		
2012 entering 5th graders	0.10 (0.07) N = 213			

Note: This table reports the coefficients on linear regressions of standardized math test scores on an indicator variable for TEP enrollment in 5th grade. Separate models were run for each cohort and outcome year combination. The comparison group consists of matched students from neighborhood schools who never enrolled in TEP; nearest-neighbor matching without replacement was conducted separately by cohort using the propensity scores predicted by the model as described in Appendix A. Regression controls include two years of baseline test scores in math, ELA, and science, as well as indicator variables for baseline demographic characteristics reported in Appendix A. Robust standard errors are reported in parentheses. Analyses are weighted using the method described in Appendix A.

*Significantly different from zero at the 0.05 level, two-tailed test.

**Significantly different from zero at the 0.01 level, two-tailed test.

ELA = English/language arts.

Table C.5. TEP impacts on ELA achievement, by cohort and duration: Nearest-neighbor matching

	1 year after enrolling at TEP	2 years after enrolling at TEP	3 years after enrolling at TEP	4 years after enrolling at TEP
2009 entering 5th graders	-0.34** (0.08) N = 222	-0.12** (0.06) N = 218	0.08 (0.07) N = 216	0.06 (0.08) N = 219
2010 entering 5th graders	-0.31** (0.08) N = 219	0.03 (0.07) N = 215	0.09 (0.08) N = 214	
2011 entering 5th graders	-0.07 (0.07) N = 207	-0.01 (0.07) N = 206		
2012 entering 5th graders	-0.02 (0.07) N = 213			

Note: This table reports the coefficients on linear regressions of standardized math, ELA, and science test scores on an indicator variable for TEP enrollment in 5th grade. Separate models were run for each cohort and outcome year combination. The comparison group consists of matched students from neighborhood schools who never enrolled in TEP; nearest-neighbor matching without replacement was conducted separately by cohort using the propensity scores predicted by the model as described in Appendix A. Regression controls include two years of baseline test scores in math, ELA, and science, as well as indicator variables for baseline demographic characteristics reported in Appendix A. Robust standard errors are reported in parentheses. Analyses are weighted using the method described in Appendix A.

*Significantly different from zero at the 0.05 level, two-tailed test.

**Significantly different from zero at the 0.01 level, two-tailed test.

ELA = English/language arts.

Table C.6. TEP impacts on science achievement: Nearest-neighbor matching

	4 years after enrolling at TEP
2009 entering 5th graders	0.16* (0.08) N = 214

Note: This table reports the coefficients on linear regressions of standardized science test scores on an indicator variable for TEP enrollment in 5th grade. Separate models were run for each cohort and outcome year combination. The comparison group consists of matched students from neighborhood schools who never enrolled in TEP; nearest neighbor matching without replacement was conducted separately by cohort using the propensity scores predicted by the model as described in Appendix A. Regression controls include two years of baseline test scores in math, ELA, and science, as well as indicator variables for baseline demographic characteristics reported in Appendix A. Robust standard errors are reported in parentheses. Analyses are weighted using the method described in Appendix A.

*Significantly different from zero at the 0.05 level, two-tailed test.

ELA = English/language arts.

The third alternative specification also used the same propensity-score estimation and impact model as the primary method of impact estimation but used kernel density matching instead of caliper matching. The impacts estimated using kernel density matching are similar to the primary estimates in magnitude and statistical significance with two exceptions: the three-year impacts on ELA for both the 2009 and 2010 cohorts are still positive but slightly smaller in size and no longer significant (Tables C.7, C.8, and C.9).

Table C.7. TEP impacts on math achievement, by cohort and duration: Kernel density matching

	1 year after enrolling at TEP	2 years after enrolling at TEP	3 years after enrolling at TEP	4 years after enrolling at TEP
2009 entering 5th graders	-0.31** (0.05) N = 1,993	-0.27** (0.05) N = 1,989	0.28** (0.05) N = 1,983	0.60** (0.05) N = 1,985
2010 entering 5th graders	0.01 (0.04) N = 2,703	-0.15** (0.06) N = 2,698	0.13** (0.05) N = 2,697	
2011 entering 5th graders	0.06 (0.05) N = 2,745	0.19** (0.04) N = 2,744		
2012 entering 5th graders	0.17** (0.04) N = 2,471			

Note: This table reports the coefficients on linear regressions of standardized math test scores on an indicator variable for TEP enrollment in 5th grade. Separate models were run for each cohort and outcome year combination. The comparison group consists of matched students from neighborhood schools who never enrolled in TEP; kernel density matching was conducted separately by cohort using the propensity scores predicted by the model as described in Appendix A. Regression controls include two years of baseline test scores in math, ELA, and science, as well as indicator variables for baseline demographic characteristics reported in Appendix A. Robust standard errors are reported in parentheses. Analyses are weighted using the method described in Appendix A.

**Significantly different from zero at the 0.01 level, two-tailed test.

ELA = English/language arts.

Table C.8. TEP impacts on ELA achievement, by cohort and duration: Kernel density matching

	1 year after enrolling at TEP	2 years after enrolling at TEP	3 years after enrolling at TEP	4 years after enrolling at TEP
2009 entering 5th graders	-0.31** (0.05) N = 1,993	-0.12** (0.04) N = 1,989	0.06 (0.04) N = 1,983	0.12** (0.05) N = 1,986
2010 entering 5th graders	-0.21** (0.05) N = 2,703	0.05 (0.05) N = 2,698	0.07 (0.06) N = 2,697	
2011 entering 5th graders	0.01 (0.03) N = 2,745	-0.01 (0.05) N = 2,744		
2012 entering 5th graders	-0.01 (0.04) N = 2,471			

Note: This table reports the coefficients on linear regressions of standardized math test scores on an indicator variable for TEP enrollment in 5th grade. Separate models were run for each cohort and outcome year combination. The comparison group consists of matched students from neighborhood schools who never enrolled in TEP; kernel density matching was conducted separately by cohort using the propensity scores predicted by the model as described in Appendix A. Regression controls include two years of baseline test scores in math, ELA, and science, as well as indicator variables for baseline demographic characteristics reported in Appendix A. Robust standard errors are reported in parentheses. Analyses are weighted using the method described in Appendix A.

**Significantly different from zero at the 0.01 level, two-tailed test.

ELA = English/language arts.

Table C.9. TEP impacts on science achievement: Kernel density matching

	4 years after enrolling at TEP
2009 entering 5th graders	0.19** (0.05) N = 1,979

Note: This table reports the coefficients on linear regressions of standardized science test scores on an indicator variable for TEP enrollment in 5th grade. Separate models were run for each cohort and outcome year combination. The comparison group consists of matched students from neighborhood schools who never enrolled in TEP; kernel density matching was conducted separately by cohort using the propensity scores predicted by the model as described in Appendix A. Regression controls include two years of baseline test scores in math, ELA, and science, as well as indicator variables for baseline demographic characteristics reported in Appendix A. Robust standard errors are reported in parentheses. Analyses are weighted using the method described in Appendix A.

**Significantly different from zero at the 0.01 level, two-tailed test.

ELA = English/language arts.

Finally, we also estimated impacts using the caliper-matching model but with standard *z*-score achievement measures rather than the rank-based *z*-score measures used in the primary impact analyses to address unreliable outliers (see Appendix A for additional information on the decision to use rank-based *z*-scores).[64] The impacts estimated using standard *z*-scores for all math

[64] The impact estimates presented in tables use standard *z*-scores and include all achievement outcome scores. We also estimated impacts using standard *z*-scores and removed outliers. In one iteration, we removed all outcomes with an absolute value *z*-score of greater than 3.0 standard deviations (SDs); in a second iteration, we removed all

and ELA achievement measures are similar to the primary impact estimates in magnitude, and statistical significance across the 2010, 2011, and 2012 cohorts (Tables C.10, C.11, and C.12), with one exception. The three-year ELA impact estimate for the 2010 cohort is no longer statistically significant. Also, there are some differences in estimates for the 2009 cohort. When using standard *z*-scores, the two- and three-year ELA impact estimates for the 2009 cohort are no longer significant; the three-year ELA impact estimate is also substantially smaller in size. There are also two notable differences in the size of estimated impacts on math achievement for the 2009 cohort: the negative impact on math achievement after two years of treatment is larger (more negative) and the positive impact on math achievement after three years of treatment is smaller.

Table C.10. TEP impacts on math achievement by cohort and duration: Standard *z*-score outcomes and baseline measures

	1 year after enrolling at TEP	2 years after enrolling at TEP	3 years after enrolling at TEP	4 years after enrolling at TEP
2009 entering 5th graders	-0.27** (0.06) N = 977	-0.30** (0.07) N = 973	0.21** (0.07) N = 967	0.58** (0.05) N = 969
2010 entering 5th graders	0.03 (0.04) N = 1,057	-0.13** (0.05) N = 1,052	0.13** (0.05) N = 1,051	
2011 entering 5th graders	0.05 (0.04) N = 1,071	0.18** (0.04) N = 1,070		
2012 entering 5th graders	0.16** (0.05) N = 1,034			

Note: This table reports the coefficients on linear regressions of standardized math test scores on an indicator variable for TEP enrollment in 5th grade. Separate models were run for each cohort and outcome year combination. The comparison group consists of matched students from neighborhood schools who never enrolled in TEP; matching was conducted by cohort using the propensity scores predicted by the model as described in this chapter and in Appendix A. Regression controls include two years of baseline test scores in math, ELA, and science, as well as indicator variables for baseline demographic characteristics reported in Appendix A. All achievement outcome, baseline, and pre-baseline measures—including measures used for matching—are standard *z*-scores rather than the rank-based *z*-scores used in the primary analyses. Analyses are weighted using the method described in Appendix A.

**Significantly different from zero at the 0.01 level, two-tailed test.

ELA = English/language arts.

(continued)

outcomes with an absolute value *z*-score of greater than 3.5 SDs. In both outlier sensitivity checks, the impact estimates were similar in size and significance to the analyses that included outliers across most years and cohorts, although the one- and two-year math and ELA achievement impacts were somewhat smaller in magnitude—that is, less negative.

Table C.11. TEP impacts on ELA achievement, by cohort and duration: Standard *z*-score outcomes and baseline measures

	1 year after enrolling at TEP	2 years after enrolling at TEP	3 years after enrolling at TEP	4 years after enrolling at TEP
2009 entering 5th graders	-0.18** (0.03) N = 977	-0.07 (0.04) N = 973	0.05 (0.05) N = 967	0.13** (0.05) N = 970
2010 entering 5th graders	-0.17** (0.05) N = 1,057	0.05 (0.05) N = 1,052	0.07 (0.06) N = 1,051	
2011 entering 5th graders	0.01 (0.04) N = 1,071	0.02 (0.05) N = 1,070		
2012 entering 5th graders	-0.04 (0.05) N = 1,034			

Note: This table reports the coefficients on linear regressions of standardized ELA test scores on an indicator variable for TEP enrollment in 5th grade. Separate models were run for each cohort and outcome year combination. The comparison group consists of matched students from neighborhood schools who never enrolled in TEP; matching was conducted by cohort using the propensity scores predicted by the model as described in this chapter and in Appendix A. Regression controls include two years of baseline test scores in math, ELA, and science, as well as indicator variables for baseline demographic characteristics reported in Appendix A. All achievement outcome, baseline, and pre-baseline measures—including measures used for matching—are standard *z*-scores rather than the rank-based *z*-scores used in the primary analyses. Analyses are weighted using the method described in Appendix A.

**Significantly different from zero at the 0.01 level, two-tailed test.

ELA = English/language arts.

Table C.12. TEP impacts on science achievement: Standard *z*-score outcomes and baseline measures

	4 years after enrolling at TEP
2009 entering 5th graders	0.23** (0.06) N = 963

Note: This table reports the coefficients on linear regressions of standardized science test scores on an indicator variable for TEP enrollment in 5th grade. Separate models were run for each cohort and outcome year combination. The comparison group consists of matched students from neighborhood schools who never enrolled in TEP; matching was conducted by cohort using the propensity scores predicted by the model as described in this chapter and in Appendix A. Regression controls include two years of baseline test scores in math, ELA, and science, as well as indicator variables for baseline demographic characteristics reported in Appendix A. All achievement outcome, baseline, and pre-baseline measures—including measures used for matching—are standard *z*-scores rather than the rank-based *z*-scores used in the primary analyses. Analyses are weighted using the method described in Appendix A.

**Significantly different from zero at the 0.01 level, two-tailed test.

ELA = English/language arts.

www.mathematica-mpr.com

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

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ WASHINGTON, DC

MATHEMATICA
Policy Research