NBER WORKING PAPER SERIES

ARE HIGH QUALITY SCHOOLS ENOUGH TO CLOSE THE ACHIEVEMENT GAP? EVIDENCE FROM A SOCIAL EXPERIMENT IN HARLEM

Will Dobbie
Roland G. Fryer, Jr

Working Paper 15473
http://www.nber.org/papers/w15473

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 2009

We are extraordinarily grateful to Geoffrey Canada, Betina Jean-Louis and Bessie Wilkerson (Harlem Children's Zone), Joel Klein (Chancellor, New York City Department of Education), Jennifer Bell-Ellwanger and Joanna Cannon (Research and Policy Support Group, New York City Department of Education), Aparna Prasad and Gavin Samms (EdLabs), and Scott Walker (Harvard Map Collection) for their endless cooperation in collecting the data necessary to perform our analysis and for hours of discussions and comments. We would also like to thank George Akerlof, Josh Angrist, David Autor, Melody Barnes, Roland Benabou, Esther Duflo, Arne Duncan (Secretary of Education), Fred Frelow, Edward Glaeser, Michael Greenstone, Lawrence Katz, Daniel Koretz, Steven Levitt, Lindsey Mathews, Kathleen McCartney, Jeannie Oates, Ben Olken, Orlando Patterson, Michelle Rhee (Chancellor, Washington D.C. Department of Education), James H. Shelton III (Assistant Deputy Secretary for Innovation and Improvement), Grover Whitehurst, William Julius Wilson, and seminar participants at Boston University, Brookings Institute, Canadian Institute for Advanced Research, Government Accounting Office, Harvard University (Economics, Graduate School of Education, Kennedy School of Government, and Sociology), Massachusetts Institute of Technology, and Princeton for detailed comments and feedback. Eduard Bogel, Vilsa Curto, Peter Evangelaki and Jonathan Scherr provided exceptional research assistance. Support from the Eli and Edythe Broad Foundation, through the Education Innovation Laboratory at Harvard University (EdLabs), is gratefully acknowledged. Correspondence can be addressed to either of the authors by mail: 44 Brattle Street, 5th Floor, Cambridge, MA, 02138; or by email: wdobbie@edlabs.harvard.edu (Dobbie) or rfryer@fas.harvard.edu (Fryer). The usual caveat applies. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

Are High Quality Schools Enough to Close the Achievement Gap? Evidence from a Social Experiment in Harlem
Will Dobbie and Roland G. Fryer, Jr
NBER Working Paper No. 15473
November 2009
JEL No. I20,J01

ABSTRACT

Harlem Children's Zone (HCZ), which combines community investments with reform minded charter schools, is one of the most ambitious social experiments to alleviate poverty of our time. We provide the first empirical test of the causal impact of HCZ on educational outcomes, with an eye toward informing the long-standing debate whether schools alone can eliminate the achievement gap or whether the issues that poor children bring to school are too much for educators alone to overcome. Both lottery and instrumental variable identification strategies lead us to the same story: Harlem Children's Zone is effective at increasing the achievement of the poorest minority children. Taken at face value, the effects in middle school are enough to close the black-white achievement gap in mathematics and reduce it by nearly half in English Language Arts. The effects in elementary school close the racial achievement gap in both subjects. We conclude by presenting four pieces of evidence that high-quality schools or high-quality schools coupled with community investments generate the achievement gains. Community investments alone cannot explain the results.

Will Dobbie
Education Innovation Laboratory
Harvard University
44 Brattle Street, 5th Floor
Cambridge, MA 02138
dobbiew@nber.org

Roland G. Fryer, Jr
Department of Economics
Harvard University
Littauer Center 208
Cambridge, MA 02138
and NBER
rfryer@fas.harvard.edu

At nine months old, there are no detectable cognitive differences between black and white babies (Fryer and Levitt, forthcoming). Differences emerge as early as age two, and by the time black children enter kindergarten they are lagging whites by 0.64 standard deviations in math and 0.401 in reading (Fryer and Levitt, 2004).[1] On every subject at every grade level, there are large achievement differences between blacks and whites that continue to grow as children progress through school (Campbell, Hombo, and Mazzeo, 2000; Neal, 2006).[2] Even accounting for a host of background factors, the achievement gap remains large and statistically significant (Jencks and Phillips, 1998).[3]

There have been many attempts to close the achievement gap. Early childhood interventions such as Head Start, Nurse-Family Partnership, and the Abecedarian Project boost kindergarten readiness, but the effects on achievement often fade once children enter school (Currie and Thomas, 1995; Krueger and Whitmore, 2001; Anderson, 2008).[4] More aggressive strategies that place disadvantaged students in better schools through busing (Angrist and Lang, 2004) and school choice plans (Rouse, 1998; Krueger and Zhu, 2002; Cullen et al., 2005; Hastings et al., 2006), have also left the racial achievement gap essentially unchanged.[5] There are several successful charter schools and charter-management organizations, but the bulk of the evidence finds only modest success (Hanushek et al., 2005; Hoxby and Rockoff, 2004; Hoxby and Murarka, 2009).[6]

School districts have tried an array of strategies to close the achievement gap, including smaller schools and classrooms (Achilles et al., 1993; Nye et al., 1995; Krueger, 1999; Krueger and Whitmore, 2001; Jepsen and Rivkin, 2002), mandatory summer school (Jacob and Lefgren, 2004) merit pay for principals, teachers and students (Podgursky and Springer, 2007; Fryer, 2009a, 2009b), after-school

[1]Early measures of math and reading achievement are more accurately described as measures of math and reading readiness. The ECLS-K reading test, for example, includes questions designed to measure basic skills (print familiarity, letter and word recognition, beginning and ending sounds, and rhyming sounds), vocabulary, listening and reading comprehension, knowledge of the alphabet, phonetics, and so on. The math test evaluates number recognition, counting, comparing and ordering numbers, solving word problems, and interpreting picture graphs.
[2]This fact was first established by Coleman et al. (1966). For more recent analysis, see Campbell et al. (2000), Carneiro and Heckman (2002), Fryer and Levitt (2004, 2006), Neal (2006), or Phillips et al., (1998).
[3]Approximately two-thirds of the achievement gap cannot be explained by background characteristics.
[4] There is some evidence that Head Start, Perry Preschool and Nurse-Family Partnership may have positive long-term impacts on outcomes such as crime, high-school graduation, and labor-market outcomes (Currie and Thomas, 2000; Ludwig and Miller, 2007; Olds, 2006; Deming, 2009).
[5] In a recent review of education policy focused on poor children, Jacob and Ludwig (2008) find that targeted investment in early childhood education, smaller class sizes, and bonuses for teachers in hard-to-staff schools all pass a cost-benefit analysis, but are unlikely to eliminate the racial and social class disparities in education outcomes by themselves.
[6] Abdulkadiroglu et al. (2009) is a notable exception, finding that students enrolled in oversubscribed Boston charter schools with organized lottery files gain about 0.17 standard deviations per year in ELA and 0.53 standard deviations per year in math. However, the typical middle school applicant in Abdulkadiroglu et al. (2009) starts 0.173 and 0.277 standard deviations higher in fourth grade math and reading than the typical Boston student, and the typical high school applicant starts 0.089 and 0.179 standard deviations higher in eighth grade math and reading tests. It is unclear whether average or below average students are as well served by the charter schools considered in the study.

programs (Lauer et al., 2006; Redd et al., 2002), budget, curricula, and assessment reorganization (Borman and Hewes, 2003; Borman et al., 2007; Cook et al., 2000), and policies to lower the barrier to teaching via alternative paths to accreditation (Decker et al., 2004; Kane et al., 2008). These programs have thus far been unable to substantially reduce the achievement gap in even the most reform-minded of districts.

The lack of progress has fed into a long-standing and rancorous debate among scholars, policymakers, and practitioners as to whether schools alone can close the achievement gap, or whether the issues children bring to school as a result of being reared in poverty are too much for even the best educators to overcome.[7] Proponents of the school-centered approach refer to anecdotes of excellence in particular schools or examples of other countries where poor children in superior schools outperform average Americans (Chenoweth, 2007). Advocates of the community-focused approach argue that teachers and school administrators are dealing with issues that actually originate outside the classroom, citing research that shows racial and socioeconomic achievement gaps are formed before children ever enter school (Fryer and Levitt, 2004; 2006) and that one-third to one-half of the gap can be explained by family-environment indicators (Phillips et al., 1998; Fryer and Levitt, 2004).[8] In this scenario, combating poverty and having more constructive out-of-school time may lead to better and more-focused instruction in school. Indeed, Coleman et al. (1966), in their famous report on equality of educational opportunity, argue that schools alone cannot treat the problem of chronic underachievement in urban schools.

Harlem Children's Zone is a 97-block area in central Harlem, New York, that combines reform-minded charter schools with a web of community services designed to ensure the social environment outside of school is positive and supportive for children from birth to college graduation. This provides a rich laboratory to understand whether communities, schools, or a combination of the two are the main drivers of student achievement. The answer to this question is of tremendous importance for public policy around the world as it goes to the heart of how communities and public goods should be allocated to alleviate racial and economic inequality. Organizations in Albany, Allentown, Atlanta, Boston, Camden, Charlotte, Detroit, Hartford, Indianola, Jacksonville, Kansas City, Los Angeles, Miami, Newark, New Orleans, Orlando, Philadelphia, Pittsburgh, Richmond, Rochester, San Francisco, Santa Fe, Savannah,

[7] Interest groups such as the Education Equality Project (http://www.educationequalityproject.org/) and the Broader, Bolder Approach (http://www.boldapproach.org/) embody this debate.

[8] The debate over communities or schools often seems to treat these approaches as mutually exclusive, evaluating policies that change one aspect of the schools or a student's learning environment. This approach is potentially informative on the various partial derivatives of the educational production function, but is uninformative on the net effect of many simultaneous changes. The educational production function may, for example, exhibit either positive or negative interactions with respect to various reforms. Smaller classes and more time on task matter more (or less) if the student has good teachers; good teachers may matter more (or less) if the student has a good out of school environment, and so on.

Washington, DC, Israel, Netherlands, Uganda and South Africa are developing plans similar to the HCZ model.

We use two separate statistical strategies to account for the fact that students who attend HCZ schools are not likely to be a random sample. First, we exploit the fact that HCZ charter schools are required to select students by lottery when the number of applicants exceeds the number of available slots for admission. In this scenario, the treatment group is composed of students who are lottery winners and the control group consists of students who are lottery losers. This allows us to provide a set of causal estimates of the effect of being offered admission into the HCZ charter schools on a range of outcomes, including test scores, attendance, and grade completion.

There are two important caveats to our lottery identification strategy. First, the HCZ middle school was not significantly oversubscribed in its first year of operation, and the HCZ elementary schools have never been significantly oversubscribed, making it more difficult to estimate the effect of being offered admission for these groups. Second, results from any lottery sample may lack external validity. The counterfactual we identify is for students who are already interested in charter schools. The effect of being offered admission to HCZ for these students may be different than for other types of students.

To complement the lotteries, our second identification strategy uses the interaction between a student's home address and her cohort year as an instrumental variable. This approach takes advantage of two important features of the HCZ charter schools: (1) anyone is eligible to enroll in HCZ's schools, but only students living inside the Zone are actively recruited by HCZ staff; and (2) there are cohorts of children that are ineligible due to the timing of the schools' opening and their age. Our identification is driven by the between-cohort comparison of outcomes within the Zone, using the outcomes of children outside the Zone to control for natural year-to-year variation in test scores. If the interaction between a student's address and cohort only affects his or her outcome through its effect on enrollment in the charter school, this provides another set of causal estimates.

Both statistical approaches lead us to the same basic story. Harlem Children's Zone is effective at increasing the achievement of the poorest minority children. Students enrolled in the sixth grade gain more than four fifths of a standard deviation in math and between one-quarter and one-third of a standard deviation in English Language Arts (ELA) by eighth grade. Taken at face value, these effects are enough to close the black-white achievement gap in mathematics and reduce it by about half in ELA. Students in the HCZ elementary school gain approximately four fifths to one and a half a standard deviation in both math and ELA by third grade, closing the racial achievement gap in both subjects. HCZ students are also less likely to be absent. These results are robust across identification strategies, model specifications, and subsamples of the data. Surprisingly, students of all ability levels obtain roughly the same benefit from attending HCZ charters.

Community programs, such as The Baby College and Harlem Gems, do not admit participants by lottery or admit in such a way that one can clearly identify their causal impact. Additionally, the IV strategy employed earlier does not have enough power in small samples. Using least squares in multiple samples and datasets we fail to find any empirical evidence that either program is positively associated with achievement at any age, but these results should be considered suggestive because they are measured imprecisely.[9]

Beyond the least squares results, there are four additional pieces of evidence that suggest that the combined effects of the community programs alone are not responsible for the gains in achievement. First, our IV strategy purges the effects of the community bundle, yet the estimates are larger than the available lottery estimates. Second, students outside the Zone garner the same benefit from the schools as the students inside the Zone, suggesting that proximity to the community programs is unimportant. Third, siblings of HCZ lottery winners who were themselves ineligible for the lottery because of their age, show no gains in achievement and modest decreases in absenteeism. Fourth, the Moving to Opportunity (MTO) experiment relocated individuals from high-poverty to low-poverty neighborhoods (while keeping children in roughly similar schools) and showed small positive results for girls and negative results for boys (Sanbonmatsu et al., 2006; Kling et al., 2007). Additionally, and more speculative, there is substantial anecdotal evidence that Harlem Children's Zone was unsuccessful in raising achievement in the years before opening the charter schools. *We cannot, however, disentangle whether communities coupled with high-quality schools drive our results, or whether the high-quality schools alone are enough.*

The paper is structured as follows. Section II provides a brief overview of Harlem Children's Zone. Section III introduces the data and econometric framework. Section IV presents estimates of the impact of four key elements of HCZ on educational outcomes. Section V discusses whether communities, schools, or both are most responsible for the results. Section VI concludes. There are four appendices: Appendix A outlines each program offered by Harlem Children's Zone. Appendix B is a data appendix that details our sample and variable construction. Appendix C empirically examines the possibility of cheating at HCZ. Appendix D conducts a back of the envelope cost-benefit exercise.

II. Harlem's Children Zone

Harlem Children's Zone began in 1970 as an amalgam of after-school programs, truancy-prevention services, and anti-violence training for teenagers in schools, named the Rheedlen Centers for Children and Families. The disintegration of central Harlem during the crack epidemic of the 1980s and

[9] Dobbie and Fryer (2009) finds scant evidence that any community programs in the Harlem Children's Zone are effective at increasing academic achievement.

1990s prompted some within the organization to question their piecemeal methodology – several programs, but no comprehensive strategy. In response, the president of Rheedlen, Geoffrey Canada, created the Harlem Children's Zone to address all the problems that poor children were facing – from crumbling apartments to failing schools and violent crime to chronic health problems – through a "conveyor belt" of services from birth to college. The approach is based on the idea that we must improve both communities and schools. Starting with a 24-block area in central Harlem, the Zone expanded to a 64-block area in 2004 and a 97-block area in 2007. Figure 1 provides an aerial map of Harlem Children's Zone and its expansion path.

HCZ offers a number of programs, which we have partitioned into "community" investments and "school" investments. Community programs are available to anyone living near HCZ, and serve 8,058 youth and 5,291 adults. School programs are only available to the approximately 1,300 students who attend the HCZ public charter schools.

Community Investments

HCZ has over 20 programs designed to help and empower individuals in their 97 blocks. These investments include early childhood programs (Head Start, e.g.), public elementary-, middle- and high-school programs (i.e. karate, dance, after-school tutoring), a college-success office, family, community and health programs, foster-care prevention services, and so on. Appendix A provides a description of all programs run by Harlem Children's Zone. HCZ's vision is to "create a tipping point" in the neighborhood so that children are surrounded by an enriching environment of college-oriented peers and supportive adults. This is consistent with the vision articulated by scholars and policymakers who argue that communities, not schools, are responsible for the achievement gap. Below, we describe, in greater detail, two of the most ambitious community programs. For a more detailed analysis of HCZ community programs, see Dobbie and Fryer (2009).

The Baby College, a signature community initiative and the first program to be started, is a nine-week parenting workshop for expectant parents and those with children up to three years old. The curriculum, informed by the work of T. Berry Brazelton, covers a broad range of subjects, including brain development, discipline, immunization, safety, asthma, lead poisoning, parental stress, and parent-child bonding (Brazelton, 1992). The College, held on Saturday mornings with weekly follow-up visits at home, is free of charge and offers parents a number of incentives to attend, including free breakfast, lunch, and daycare, as well as drawings for prizes such as gift certificates at local stores or a month's free rent. Anyone with children up to three years old is eligible to attend The Baby College, but tremendous effort is made to recruit parents from inside the Zone (Tough, 2008). Thirteen HCZ employees canvass apartment buildings, seek suggestions for good candidates from past Baby College participants, and visit laundromats, supermarkets and check-cashing outlets within the Zone in search of possible enrollees, including revisiting people who indicated they were interested, but never enrolled.

Another major community initiative is Harlem Gems, an intensive, all-day, pre-kindergarten program with a 4:1 child-to-adult ratio that focuses on kindergarten preparation. The curriculum is designed to increase socialization skills, build routines, and begin development of the pre-literacy and language skills students need in kindergarten. The Gems program incorporates a number of nontraditional subjects such as Spanish and French, and strongly encourages parents to volunteer at the school and become more involved in their child's education. Each Gems classroom employs a Family Worker, whose primary responsibility is ensuring that parents are active and involved. There are monthly parent meetings, book nights, and field trips that parents are encouraged to attend, and between one and five monthly parenting workshops on a wide range of topics, including some designed purely for parental involvement (e.g. a knitting club).[10]

School Investments

The HCZ Promise Academy public charter schools began in the fall of 2004 with the opening of the HCZ Promise Academy elementary and middle schools, followed in the fall of 2005 with the opening of the HCZ Promise Academy 2 Elementary School.[11] The HCZ Promise Academy will enroll a new kindergarten and sixth-grade cohort each year until it is a full K-12 school, while HCZ Promise Academy 2 will enroll a new kindergarten cohort each year until it is a full K-12 school.[12][13]

The HCZ Promise Academies have an extended school day and year, with coordinated after-school tutoring and additional classes on Saturdays for children who need remediation in mathematics and English Language Arts skills. Our rough estimate is that HCZ Promise Academy students that are behind grade level are in school for twice as many hours as a traditional public school student in New York City.

---

10 There is also a Head Start version of Harlem Gems, which follows a similar model except that (1) children can enter the Head Start at three or four years of age, thus a sizable proportion of the participants receive two years of instruction, and (2) because of income guidelines, students in Head Start tend to come from families with lower socioeconomic status than that of Universal Pre-K participants. In all versions of the program, the students are most often referred from The Baby College or recruited from within the HCZ borders.

11 Promise Academy 2 held both a kindergarten and first-grade lottery their first year, enrolling 40 students in each grade. After their first year of operation, Promise Academy 2 relocated, and in the process lost a number of students. To simplify our analysis and abstract from the issues created by this relocation, we focus our analysis on Promise Academy.

12 In the fall of 2007, Promise Academy did not enroll a new sixth-grade cohort, and expelled the entire existing eighth-grade cohort (the 2004 lottery cohort). HCZ was unhappy with the performance of the middle school to this point, and the decision was made to focus on the existing sixth- and seventh-grade cohorts in reforming the school. Test data, which were made public months after the decision was made, showed that the 2004 cohort did quite well (Tough, 2008).

13 In 2008, the lottery was moved back to fifth grade (Tough, 2008).

Students who are at or above grade level still attend the equivalent of about fifty percent more school in a calendar year.

Both schools emphasize the recruitment and retention of high-quality teachers and use a test-score value-added measure to incentivize and evaluate current teachers. The schools have had high turnover as they search for the most effective teachers: 48 percent of HCZ Promise Academy teachers did not return for the 2005 – 2006 school year, 32 percent left before 2006 – 2007, and 14 percent left before 2007 – 2008. Each teacher has an annual meeting with Geoffrey Canada to discuss their performance, and is supported by myriad behind-the-scenes efforts to make sure their time is spent primarily on teaching and not administrative tasks.

The schools provide free medical, dental and mental-health services (students are screened upon entry and receive regular check-ups through a partnership with the Children's Health Fund), student incentives for achievement (money, trips to France, e.g.), high-quality, nutritious, cafeteria meals, support for parents in the form of food baskets, meals, bus fare, and so forth, and less tangible benefits such as the support of a committed staff. The schools also make a concerted effort to change the culture of achievement, surrounding students with the importance of hard work in achieving success. These types of school investments are consistent with those that argue high-quality schools are enough to close the achievement gap.

Anecdotal Evidence of Success

A slew of anecdotal evidence suggests that the HCZ approach is working. For six straight years, 100 percent of pre-kindergarteners in the Harlem Gems program were school-ready. Eighty-one percent of parents who have attended The Baby College report reading to their child more often than they previously read to them. Older students in the HCZ public charter schools typically outperform their peers at neighboring schools, and, most recently, three young women from the HCZ chess program won the national championship for their age group. In the latest progress report from the New York City Department of Education, the Promise Academies received an A.

One has to take the above evidence with a grain of salt – children who participate in HCZ are not a random sample of students. Children may enroll in the Zone's programs because they are already struggling or under-performing in traditional public schools, because they are identified as being particularly at risk by a teacher or counselor, or because they or their parents are highly motivated to seek out the best opportunities. In other words, students served by HCZ are likely to be self-selected, either positively or negatively, and results that simply compare outcomes of HCZ children to the outcomes of other children in Harlem may be biased in one direction or another.

III. Data and Econometric Framework

We merge data from two sources: information from files at Harlem Children's Zone and administrative data on student demographics and outcomes from the New York City Department of Education (NYCDOE).

The data from Harlem Children's Zone consist of enrollment files from The Baby College and Harlem Gems and lottery results from the HCZ Promise Academy public charter schools.[14] A typical student's data include her first name, last name, birth date, parents' or guardians' names, home address, and the program to which she is applying or attending. With the exception of The Baby College, the HCZ data are remarkably complete. The Baby College data are more sparse because almost 18 percent of children had not been born at the time of data collection. In these cases "expected named" and due date were collected.

The NYCDOE data contain student-level administrative data on approximately 1.1 million students across the five boroughs of the NYC metropolitan area. The data include information on student race, gender, free and reduced-price lunch status, behavior, attendance, and matriculation with course grades for all students and state math and ELA test scores for students in grades three through eight. The data also include a student's first and last name, birth date, and address. We have NYCDOE data spanning the 2003 – 2004 to 2008 – 2009 school years.

The state math and ELA tests, developed by McGraw-Hill, are high-stakes exams conducted in the winters of third through eighth grade.[15] Students in third, fifth, and seventh grades must score level 2 or above (out of 4) on both tests to advance to the next grade. The math test includes questions on number sense and operations, algebra, geometry, measurement, and statistics. Tests in the earlier grades emphasize more basic content such as number sense and operations, while later tests focus on advanced topics such as algebra and geometry. The ELA test is designed to assess students on three learning standards – information and understanding, literary response and expression, critical analysis and evaluation – and includes multiple-choice and short-response sections based on a reading and listening section, along with a brief editing task.[16]

All public-school students, including those attending charters, are required to take the math and ELA tests unless they are medically excused or have a severe disability. Students with moderate disabilities or who are Limited English Proficient must take both tests, but may be granted special accommodations (additional time, translation services, and so on) at the discretion of school or state

[14] The exception being the 2004 middle-school cohort, for which the lottery numbers were lost, and the 2006 elementary-school lottery cohort, which is still being compiled by HCZ staff.
[15] Sample tests can be found at http://www.emsc.nysed.gov/osa/testsample.html
[16] Content breakdown by grade and additional exam information is available at http://www.emsc.nysed.gov/osa/pub/reports.shtml

administrators. In our analysis the test scores are normalized to have a mean of zero and a standard deviation of one for each grade and year.[17]

We also construct measures of absenteeism and matriculation using the NYCDOE data. Absenteeism is measured as the total number of absences a student accumulates during the first 180 days of the school year. After the first 180 days, the NYCDOE no longer collects absence data from schools. Matriculation is an indicator for whether a student is "on-time" given her expected grade. In our elementary-school sample a student's expected grade is imputed from her initial kindergarten year. If a student enters kindergarten in 2003 – 2004, she is on time if enrolled in first grade in 2004 – 2005, second grade in 2005 – 2006, and so on. If we do not observe an elementary student in kindergarten, we impute an expected grade from the student's birth date. In our middle-school sample a student's expected grade is imputed from her initial sixth-grade year.[18] If we do not observe a student in sixth grade, we impute an expected grade from the nearest observed grade.

The HCZ data were matched to the New York City administrative data using the maximum amount of information available. Match keys were used in the following order: (1) last name, first name, date of birth with various versions of the names (abbreviations, alternative spellings, hyphenated vs. non-hyphenated); (2) last name, first name, and various versions of the date of birth (most often the month and day reversed); (3) last name, first name, prior school, and prior grade with various likely adjustments to prior grade; (4) name, date of birth, and prior grade. Once these match keys had been run, the remaining data were matched by hand considering all available variables. Match rates were 93.1 percent for the Harlem Gems data (N=451), 92.6 percent for the winners of the kindergarten lottery (N=203), 89.2 percent for the losers of the kindergarten lottery (N=195), 90.5 percent for the winners of the middle-school lottery (N=200), and 85.5 percent for the losers of the middle-school lottery (N=352).[19] These numbers are comparable to the match rates achieved by others using similar data (Hoxby and Murarka, 2009).

Match rates for The Baby College were significantly lower. Recall, approximately 18 percent of the Baby College sample was not born yet, necessitating that we match on expected name and due date.

---

17 Results are of similar magnitude and statistical significance when raw or percentile scores are used instead of standardized scores.

18 In our sample 24 percent of all students and 27 percent of black students have already been held back by sixth grade. As a result, a matriculation measure based on kindergarten enrollment is less informative.

19 If nonrandom attrition is driving the difference in match rates in our middle-school sample, our results may be biased. We explore the sensitivity of our results by creating pseudo-lottery losers to equalize the match rates for each cohort. These pseudo-students are assigned the highest test score observed in the control group for that year (about two standard deviations above the mean). The results of this exercise are available in Appendix Table 2. The estimated effect of attending the Promise Academy Middle School remain qualitatively similar, with ITT and TOT effects only slightly smaller than the main lottery results reported in Table 3. We also note that our IV estimates using the interaction between a student's cohort and address should be unaffected by the difference in match rates as it is based entirely on the NYCDOE data.

Another 20 percent of The Baby College sample did not have age or birth date information. Finally, a number of Baby College participants were too young to be in our NYCDOE dataset. Within participants with complete information and of appropriate age, we were able to match 61.29 percent (N=987) of our sample. As Baby College participants have three to six years to attrite before we observe them in the NYCDOE data, this is consistent with our expected match rate.

Using the student addresses provided by the NYCDOE, we also calculated the distance from each student's home to the nearest point on the boundary of the Harlem Children's Zone using arcGIS. When multiple addresses were available for a single student, we use the earliest available address.[20] A student is defined as living "in the Zone" if they live completely inside or touching the boundaries of the original 24-block Zone.

Summary statistics for the variables that we use in our core specifications are displayed in Table 1. Students who entered the elementary- or middle-school lottery are more likely to be black, but all other variables, including eligibility for free or reduced-price lunch, are similar to the typical New York City student. Students enrolled in the middle school lottery score 0.287 standard deviations and 0.266 standard deviations below the typical New York City student in fifth grade math and ELA respectively, but outscore the typical student living in the Zone by 0.186 in math (p-value = 0.001) and 0.062 in ELA (p-value = 0.289).

Econometric Approach to Estimating the Causal Impact of Harlem Children's Zone

The simplest and most direct test of any HCZ program would be to examine the outcome of interest regressed on an indicator for enrollment in the particular program (HCZ_i) and controls for basic student characteristics (X_i):

$$outcome_i = \alpha + \beta X_i + \gamma HCZ_i + \varepsilon_i .$$

If students select into the programs because of important unobserved determinants of academic outcomes, such estimates may be biased. To confidently identify the causal impact of HCZ enrollment, we must compare students with different enrollment statuses who would have had the same academic outcomes had they both been enrolled or not enrolled in the program. By definition, this involves a counterfactual we cannot observe.

In our analysis, we construct the counterfactual in two ways. First, we exploit the fact that HCZ public charter schools are required to select students by lottery when demand exceeds supply, and,

[20] Another approach is to use the student's address closest to the date of the lottery. The results are not sensitive to this alternative. Our view is that the earlier address was more likely to be exogenous.

second, by using the interaction between a student's home address and cohort year as an instrumental variable.

New York law dictates that over-subscribed charter schools allocate enrollment offers via a random lottery. Restricting our analysis to students who entered the HCZ lottery, we can estimate the causal impact of being offered admission into the public charter school by comparing average outcomes of students who 'won' the lottery to the average outcomes of students who 'lost' the lottery. The lottery losers therefore form the control group corresponding to the counterfactual state that would have occurred for students in the treatment group if they had not been offered a spot at the public charter school.

Let Z_i be an indicator for a winning lottery number. The mean difference in outcomes between the lottery winners (Z_i = 1) and lottery losers (Z_i = 0) is known as the "Intent-to-Treat" (ITT) effect, and is estimated by regressing student outcomes on Z_i. In theory, predetermined student characteristics (X_i) should have the same distribution within the lottery winners and losers because they are statistically independent of group assignment. In small samples, however, more precise estimates of the ITT can often be found by controlling for these student characteristics (X_i). The specifications estimated are of the form:

(1) $$outcome_i = \alpha_1 + X_i\beta_1 + Z_i\gamma_1 + \varepsilon_{1i}$$

where our vector of controls, X_i , includes an indicator variable for gender, a mutually inclusive and collectively exhaustive set of race dummies, an indicator for free lunch, and predetermined measures of the outcome variable when possible (i.e. pre-lottery test scores).[21]

The ITT is an average of the causal effects for students who enroll in the public charter school and those that do not. The ITT therefore captures the causal effect of being offered a spot in the HCZ public charter school, not of actually attending.

Under several assumptions (that the treatment group assignment is random, lottery losers are not allowed to enroll, and winning the lottery only affects outcomes through public charter school enrollment), we can also estimate the causal impact of actually enrolling in the HCZ public charter

[21] A student is income-eligible for free lunch if her family income is below 130 percent of the federal poverty guidelines, or categorically eligible if (1) the student's household receives assistance under the Food Stamp Program, the Food Distribution Program on Indian Reservations (FDPIR), or the Temporary Assistance for Needy Children Program (TANF); (2) the student was enrolled in Head Start on the basis of meeting that program's low-income criteria, (3) the student is homeless, (4) the student is a migrant child, or (5) the student is a runaway child receiving assistance from a program under the Runaway and Homeless Youth Act and is identified by the local educational liaison. A student is eligible for reduced-price lunch if family income is between 130 and 185 percent of federal poverty guidelines. Approximately 70 percent of our sample is eligible for free lunch, with another ten percent eligible for reduced-price lunch.

school.[22] This parameter, commonly known as the "Treatment-on-Treated" (TOT) effect, measures the average effect of enrollment on lottery winners who choose to attend the HCZ public charter school. The TOT parameter can be estimated though a two-stage least squares regression of student outcomes on enrollment (HCZ_i) with lottery status (Z_i) as an instrumental variable for enrollment:

$$outcome_i = \alpha_2 + \beta_2 X_i + \gamma_2 HCZ_i + \varepsilon_{2i} \tag{2}$$

The TOT is the estimated difference in outcomes between students who actually enroll in the public charter school and those in the control group who would have enrolled if they had been offered a spot.

A key concern in estimating the TOT is how to account for students who leave the charter school. Over time, it is inevitable that students leave the HCZ public charter schools, either through natural attrition or because they are asked to leave. First- (and second-) year attrition from the HCZ Promise Academy is 8.5 (17.6) percent for the elementary school and 15.6 (26.4) percent for the middle school, compared to 9.7 (17.0) percent for KIPP Star – a nationally recognized charter middle school two blocks from the western boundary of the Zone – and 16.4 (30.3) for the middle schools most demographically similar to HCZ. Demographically similar elementary schools all experience first-year attrition of over 40 percent.[23]

If attrition is nonrandom, the TOT estimates may be biased if we estimate the effect of current enrollment. We therefore let HCZ_i be an indicator for a student ever having been enrolled at the public charter school, regardless of her current enrollment status. In the current paper, this will serve as a lower-bound estimate of the true effect. If we used the indicator as current enrollment we get TOT estimates that are approximately three and a half percent higher than those reported and if we use years of enrollment (as others in the literature have done, see, for example, Abdulkadiroglu et al., 2009) we get estimates that are approximately two percent higher in seventh grade and three percent higher in eighth grade.

To compliment the lotteries, our second statistical strategy exploits the interaction between a student's home address and cohort year as an instrumental variable. Two forces drive our identification.

---

[22] Four to five students in each middle school cohort enroll despite having "losing" lottery numbers due to clerical errors (private correspondence with HCZ), but otherwise the lottery is binding. To insure that the TOT assumptions are valid, we code these individuals as not having received treatment. Coding these individuals as having received treatment would imply that our estimates are a LATE, and results in somewhat larger point estimates. The kindergarten lotteries were not significantly oversubscribed in any year, and as a result, nearly every student was eventually offered a spot at the charter school. As a result we cannot estimate ITT and TOT effects for these cohorts. We are able to use a student's position on the waitlist as an instrumental variable and estimate a local average treatment effect of attending the school. This is discussed further in the results section.

[23] First-year attrition in elementary school is defined as the percent of students who leave between kindergarten and first grade. In the HCZ and public middle schools, attrition is defined as the percent of students who leave between sixth and seventh grades, and in KIPP Star it is defined as the percent that leave between fifth and sixth grades. Demographically similar public schools were chosen by minimizing the weighted difference in available demographic variables and previous test scores between all NYC schools and the HCZ Promise Academy. Weights were determined by regressing a previous cohort's demographic variables on test scores.

First, we compare outcomes between children living in the Zone who were eligible for its public charter schools and students living in the Zone who were not eligible. For example, students who started kindergarten in 2003 were ineligible for the HCZ public charter school, which began enrolling kindergarten students in 2004. As the 2003 cohort is likely to be quite similar to the 2004 cohort, they provide a plausible counterfactual. Second, we compare the outcomes of children living outside the Zone in the two cohorts to adjust for year-to-year variation that may come about through broad city-wide reforms.

The basic logic and evidence for our IV strategy is illustrated in Figure 2, which plots the percentage of students enrolled in the HCZ public charter schools and mean test scores for two successive cohorts.[24] In both panels, the vertical axis represents enrollment rates (the denominator is all children in the specified area) and the horizontal axis represents distance in meters to the nearest point on the Zone's border. The left panel is the 2004 cohort who are age-eligible for the HCZ Promise Academy; the right panel contains the 2003 cohort who was born a year too early and were already in kindergarten when the lottery commenced. Our identification strategy compares the outcomes of children in the Zone between the eligible and ineligible cohorts, letting the age-ineligible students form the counterfactual of what would have occurred had the age-eligible children not enrolled in the public charter school. We also compare children outside the Zone to adjust for year-to-year achievement shocks.

Let enrollment in HCZ programs be a function of student characteristics (X_i), home address ($inZone_i$), cohort year ($Cohort_i$), and the interaction between address and cohort year:

(3) $$HCZ_i = \alpha_3 + \beta_3 X_i + \gamma_3 inZone_i + \delta_3 Cohort_i + \eta_3 \left(inZone_i * Cohort_i \right) + \varepsilon_{3i}$$

The residual of this representative "first-stage" equation (3) captures other factors that are correlated with enrollment in HCZ that may be related to student outcomes. The key identifying assumptions of our approach is that (1) the interaction between address (in or out of Zone boundaries) and cohort year is correlated with enrollment, and (2) the interaction between address and cohort year only affects student outcomes through its effects on the probability of enrollment, not through any other factor or unobserved characteristic.

The first assumption is econometrically testable. Appendix Table 1 presents results from representative first-stage regressions, where enrollment in the HCZ public charter school at kindergarten and sixth grade are regressed on student race, lunch status, address, cohort year and the interaction

[24] In our elementary-school sample a student's cohort is imputed from her initial kindergarten year. If we do not observe an elementary student in kindergarten, we impute the expected grade from the student's birth date. In our middle-school sample a student's cohort is imputed from her initial sixth-grade year. If we do not observe a student in sixth grade, we impute an expected grade from the nearest observed grade.

between address and cohort year. For kindergarten enrollment, five of the eight interactions are significant at the one-percent level and the remaining three are significant at the five-percent level. For sixth-grade enrollment, four of the seven interactions are significant at the one-percent level. For both regressions, a joint F-test with the null that the interactions are jointly equal to zero is strongly rejected (p-value = 0.000).

The validity of our second identifying assumption – that the instrument only affects student outcomes through the probability of enrollment - is more difficult to assess. To be violated, the interaction between a student's address and cohort year must be correlated with her outcomes after controlling for the student's background characteristics, address and cohort year. This assumes, for instance, that parents do not selectively move into the Children's Zone based on their child's cohort. Given that all children, regardless of their address, are eligible for HCZ programs, this seems a plausible assumption. Motivated parents can enroll their children in the programs no matter where they live; the relationship between distance to the Zone and enrollment comes about primarily through increased knowledge about the programs or cost of attending, not eligibility. We also assume that shocks either affect everyone in a given cohort regardless of address, or affect everyone at a given address regardless of cohort. If there is a something that shifts achievement test scores for third graders living inside the Children's Zone, but not third graders outside the Zone or fourth graders inside the Zone, our second identifying assumption is violated.[25]

Under these assumptions (and a monotonicity assumption that being born into an eligible cohort does not make a student less likely to enroll) we can estimate the causal impact of enrolling in the HCZ public charter school. The identified parameter, commonly known as the local average treatment effect (LATE), measures the average effect of treatment for students induced into enrollment by the instrument (Imbens and Angrist, 1994). The LATE parameter is estimated though a two-stage least squares regression of student outcomes on enrollment (HCZ_i) with the interaction between address and cohort as an instrumental variable for enrollment.

IV. The Impact of Four Key Elements of Harlem Children's Zone on Student Achievement

HCZ Promise Academy – Middle School

[25] For our second identifying assumption to be valid, there should be no difference in the relationship between eligible and ineligible students living inside the Zone and the relationship between eligible and ineligible students living outside the Zone ($Eligible_{in}$-$Ineligible_{in}$= $Eligible_{out}$-$Ineligible_{out}$). Appendix Table 3 presents a partial test of this assumption. Columns 1 through 4 and 6 through 9 present basic demographic statistics for the IV sample. Columns 5 and 10 present p-values from a test with the null hypothesis that the difference between eligible and ineligible cohorts is the same for students in the Zone and out of the Zone. Ineligible students living outside the Zone are more likely to be eligible for reduced price lunch. There are no other statistically significant differences.

Figures 3-5 provide a visual representation of our basic results from HCZ Promise Academy – Middle School. Figure 3A plots yearly, raw, mean state math test scores, from fourth to eighth grade, for four subgroups: lottery winners, lottery losers, white students in New York City public schools and black students in New York City public schools. Lottery winners are comprised of students who either won the lottery or who had a sibling that is already enrolled in the HCZ Promise Academy. Lottery losers are individuals who lost the lottery and did not have a sibling already enrolled.[26] Technically, these represent ITT estimates.

In fourth and fifth grade, before they enter the middle school, math test scores for lottery winners, losers, and the typical black student in New York City are virtually identical, and roughly 0.75 standard deviations behind the typical white student.[27] Lottery winners have a modest increase in sixth grade, followed by a more substantial increase in seventh grade and even larger gains by their eighth-grade year.

The TOT estimate, which is the effect of actually attending the HCZ Promise Academy Middle School, is depicted in panel B of Figure 3. As we directly observe test scores of lottery winners who enroll and we have a TOT estimate, we can estimate the mean test score for those in the control group who would have enrolled if they had won the lottery (Kling et al., 2007). The TOT results follow a similar pattern, showing remarkable convergence between children in the middle school and the average white student in New York City. After three years of "treatment," HCZ Promise Academy students have nearly closed the achievement gap in math – they are behind their white counterparts by 0.121 standard deviations (p-value = 0.113). If one adjusts for gender and free lunch, the typical eighth grader enrolled in the HCZ middle school outscores the typical white eighth grader in New York City public schools by 0.087 standard deviations though the difference is not statistically significant (p-value = 0.238). Put differently, the typical student before entering the HCZ public charter school was at the 20th percentile of the white distribution. After three years, the typical student is at the 45th percentile before adjusting for gender and free lunch and the 54th percentile after adjusting for gender and free lunch.

Figure 4A plots yearly state ELA test scores, from fourth to eighth grade. Treatment and control designations are identical to Figure 3A. In fourth and fifth grade, before they enter the middle school, ELA scores for lottery winners, losers, and the typical black student in NYC are not statistically different, and roughly 0.65 standard deviations behind the typical white student.[28] Lottery winners and losers have very similar ELA scores from fourth through seventh grade. In eighth grade, HCZ Promise Academy

[26] Lottery winners were defined using information provided by HCZ regarding the last student called off the waitlist – number 10 in 2005 and number eighteen in 2006. Given the size of the estimated treatment effect, our results are robust to other definitions of "lottery winner."
[27] This is similar in magnitude to the math racial achievement gap in nationally representative samples (0.882 in Fryer and Levitt, 2006, and 0.763 in Campbell et al., 2000), and samples with an urban focus (Fryer 2009c).
[28] This is smaller than the reading racial achievement gap in some nationally representative samples (0.771 in Fryer and Levitt, 2006, and 0.960 in Campbell et al., 2000), but similar to data with an urban focus (Fryer 2009c).

students distance themselves from the control group. These results are statistically meaningful, but much less so than the math results. The TOT estimate, depicted in panel B of Figure 4, follows an identical pattern with marginally larger differences between enrolled middle-school students and the control group. Adjusting for gender and free lunch pushes the results in the expected direction.

Interventions in education often have larger impacts on math scores as compared to reading or ELA scores (e.g. Decker et al., 2004; Rockoff, 2004; Jacob, 2005). This may be because it is relatively easier to teach math skills, or that reading skills are more likely to be learned outside of school. Another explanation is that language and vocabulary skills may develop early in life, making it difficult to impact reading scores in adolescence (Hart and Risley, 1995; Nelson, 2000).

Thus far, we have concentrated solely on the average student. It is important to know which students benefit the most from being offered a spot at the HCZ Promise Academy middle school. The distributions of test scores, by treatment status, are plotted in Figures 5A and 5B. The modest gains in sixth-grade math scores appear to come primarily from the left tail of the distribution. By seventh and eighth grade, however, the lottery winner and lottery loser score distributions have noticeably separated across all ex ante achievement levels. In stark contrast, gains in eighth-grade ELA scores appear to come primarily from the left tail of the distribution.

Table 2 provides some concrete numbers, and, more importantly, standard errors to the figures just described. Columns 1 and 2 present ITT and TOT estimates of the pooled sample of both cohorts. Columns 3 through 6 provide identical estimates for the 2005 and 2006 cohorts respectfully. Each row represents a different outcome of interest, including math and ELA achievement scores from standardized statewide exams, an indicator for whether or not a student is scoring on grade level, absences, and whether or not a student matriculates on time. The sample is restricted to students with outcome data in 2008 - 2009, the most recent available year.[29]

In the pooled sample, a lottery winner scores 0.249 (0.053) standard deviations higher in sixth-grade mathematics, 0.275 (0.054) standard deviations higher in seventh grade, and 0.546 (0.076) standard deviations in eighth-grade math relative to lottery losers. Lottery winners are 11.6 (4.8) percent more likely to be performing on grade level in mathematics in sixth grade, 17.9 (4.3) percent more likely in

[29] There is not a statistically significant difference in the likelihood of lottery losers and lottery winners being in our final dataset. Regressing an indicator for being included in the final dataset on an indicator for winning the lottery and basic demographic controls, the coefficient on the being a lottery winner is 0.0248 (0.0301) in the Middle School sample and 0.0446 (0.0326) in the Elementary School sample. If nonrandom attrition is driving the differences in the probability of being in our final data set, our results may be biased. We can explore the sensitivity of our results by creating pseudo-lottery losers to equalize the rates for each group, similar to the sensitivity exercise conducted in Appendix Table 2. Even when these pseudo-students are assigned the highest test score observed in the control group for that year (about two standard deviations above the mean) the results are nearly identical to those presented below.

seventh grade, and 27.5 (4.3) percent in eighth grade.[30] ELA achievement is strikingly similar between lottery winners and losers until their eighth-grade year, where admission into HCZ Promise Academy middle schools causes a 0.190 (0.058) standard deviation increase in ELA achievement.

The effect of actually attending the HCZ Promise Academy – the TOT estimate – is proportionately larger. In the pooled sample, an enrollee scores 0.348 (0.073) standard deviations higher in sixth-grade math, 0.394 (0.076) higher in seventh-grade math, 0.765 (0.100) higher in eighth-grade math, and 0.266 (0.081) higher in eighth-grade ELA relative to non-enrollees. As before, the effect of attending HCZ Promise Academy on sixth and seventh grade ELA scores is essentially zero (0.047 (0.075) and 0.041 (0.072), respectively).

Lottery winners are absent less in the first 180 days than the control group in every grade – 2.984 (0.704) days in sixth grade, 2.202 (1.290) days in seventh grade, and 3.960 (1.323) days in eighth grade. There are no statistically significant differences in matriculation between the lottery winners and lottery losers. For sixth and seventh grade, the magnitude of the results in the 2006 cohort are bigger than the 2005 cohort; for eighth grade the magnitude of the results are larger for the 2005 cohort. There is no obvious explanation for these patterns. The pooled sample, mechanically, falls in between these two sets of estimates. The advantage of the pooled sample is that the standard errors are slightly lower.

Table 3 presents instrumental variable results for our middle-school cohort. Our instrumental-variables strategy is most useful in examining programs that do not have oversubscribed lotteries, but it can also provide a useful robustness check of our lottery estimates, tests the external validity of our lottery estimates, and allows us to include the 2004 middle-school cohort whose lottery numbers were apparently lost. A key assumption in our instrumental-variables analysis is that students outside the Zone provide an effective control for year-to-year variation (changes in the test, new community programs, and so on). If shocks are local, it is preferable to restrict our sample to students who live very close to the Zone. If shocks are more widespread, however, expanding the number of students in our sample will increase precision without biasing our results. We do not have strong prior beliefs on the proximity of the shocks, so we estimate a number of specifications, including samples that are restricted to students within 800, 1600 and 2400 meters of the border of the Zone. Appendix Figure 1 displays poverty rates of the Zone and the surrounding areas. Addresses within 800 meters of the Zone are quite similar. As we get farther away from the original Zone the sample includes more and more affluent areas. In practice, the results are qualitatively similar across the reported samples. To improve precision we drop students with test scores greater than three standard deviations from the mean and students with more than 50 absences. Results are qualitatively similar if all students are included in the regressions.[31]

---

[30] A student that scores either a 3 or 4 (on a 4-point scale) on the state test is considered “at or above grade level.”

[31] Students with more than 50 absences or greater than three standard deviations away from the mean make up about

Each entry in Table 3 is an estimate of the effect of being enrolled at the HCZ Promise Academy Charter School in sixth grade on the reported outcome, with standard errors clustered at the cohort level in parentheses.[32] Columns 1 through 3 of Table 3 demonstrate that the effect on math achievement is large, positive, and increasing as children age. Point estimates range from 0.066 (0.562) to 0.217 (0.568) standard deviations for sixth-grade scores, 0.902 (0.299) to 1.208 (0.405) for seventh-grade scores, and 1.130 (0.153) to 1.298 (0.191) for eighth-grade scores. Consistent with these results, HCZ Promise Academy middle-school students are 0.512 (0.124) to 0.636 (0.127) more likely to be at or above grade level by eighth grade in math. Consistent with our lottery estimates, the treatment effect of HCZ on achievement is smaller for ELA scores. Point estimates range from -0.010 (0.402) to 0.052 (0.311) standard deviations for sixth grade, 0.209 (0.461) to 0.318 (0.385) for seventh grade, and 0.136 (0.303) to 0.273 (0.257) for eighth grade. None of the ELA estimates are statistically significant. That the LATE estimates in Table 3 are so similar to the TOT estimates in Table 2 gives us added confidence that our IV strategy is valid.

Let us put the magnitude of our estimates in perspective. Jacob and Ludwig (2008), in a survey of programs and policies designed to increase achievement among poor children, report that only three often practiced educational policies pass a simple cost-benefit analysis: lowering class size, bonuses for teachers for teaching in hard-to-staff schools, and early childhood programs. The effect of lowering class size from 24 to 16 students per teacher is approximately 0.22 (0.05) standard deviations on combined math and reading scores (Krueger, 1999). While a one-standard deviation increase in teacher quality raises math achievement by 0.15 to 0.24 standard deviations per year and reading achievement by 0.15 to 0.20 standard deviations per year (Rockoff, 2004; Hanushek and Rivkin, 2005; Aaronson, Barrow, and Sander, 2007; Kane and Staiger, 2008), value added measures are not strongly correlated with observable characteristics of teachers making it difficult to ex ante identify the best teachers. The effect of Teach for America, one attempt to bring more skilled teachers into struggling schools, is 0.15 standard deviations in math and 0.03 in reading (Decker et al., 2004). The effect of Head Start is 0.147 (0.103) deviations in applied problems and 0.319 (0.147) in letter identification on the Woodcock-Johnson exam, but the effects on test scores fade in elementary school (Currie and Thomas, 1995; Ludwig and Phillips, 2007). An average charter school in New York City raises math scores by 0.09 (0.01) standard deviations per year and ELA scores by 0.04 (0.01) standard deviations per year (Hoxby and Muraka, 2009). All these

5 percent of our overall sample. According to the NYCDOE, these students are most likely those that did not fill in their test sheet at all, and those with missing or corrupt data. It also clearly includes habitual truants with correct data.

[32] Bertrand et al. (2004) find that clustering standard errors when the number of groups is small can lead to over-rejection of the null hypothesis of no effect. In Monte Carlo simulations using CPS data, they find that the null is rejected falsely 8 percent of the time with 10 groups, and 11.6 percent of the time with 6 groups. Our IV results should be interpreted with this caveat in mind. We thank Josh Angrist and Esther Duflo for raising this issue.

effects are a small fraction of the impact of being offered admission into HCZ public charter schools. Abdulkadiroglu et al. (2009) find effect sizes closest to our own, with students enrolled in a set of Boston charter middle schools gaining about 0.539 (0.078) standard deviations per year in math and about 0.171 (0.052) standard deviations per year in ELA.

Table 4 explores the sensitivity of our estimated treatment effects on a variety of subsamples of the data. We report only the TOT estimates from our lottery results and their associated standard errors. The first column of the table reports baseline results from the pooled regressions in Table 3. Columns 2 and 3 present coefficients for two mutually exclusive and collectively exhaustive subsamples (i.e. boys versus girls). The last column provides the p-value on a test of equality between the coefficients.

Surprisingly, there are no statistically meaningful differences among subsets including students who entered below or above the median test score, free lunch status, and students who live within 400 meters of the boundary of Harlem Children's Zone compared to students that do not.[33] In other words, we cannot reject the null hypothesis that every subgroup benefits equally from attending.

Although the above results provide great optimism for the possibility of reform in urban public schools, one worries that improvements on state exams may be driven by test specific preparatory activities at the expense of more general learning. Jacob (2005), for example, finds evidence that the introduction of accountability programs increases high stakes test scores without increasing scores on low-stakes tests, most likely through increases in test-specific skills and student effort. It is important to know whether the results presented above are being driven by actual gains in general knowledge or whether the improvements are only relevant to the high-stakes state exams.[34]

To provide some evidence on this question, we present data from the Iowa Test of Basic Skills (ITBS) that was administered to every student enrolled at the HCZ public charter schools. The ITBS is a standardized, multiple-choice exam administered in grades 3 through 8. The ITBS is used by nearly every school district in Iowa and many districts nationally.[35] We report National Percentile Rank scores, which compares a student to a national sample of students who are in the same grade and who were tested at the same time of year (fall or spring). As a result of this standardization, having the same score across time implies "normal" academic progress. There are two important caveats to our ITBS results. First, scores are not available for the control group, so the evidence is only suggestive. Second, we do not have scores for any other schools in NYC, making comparisons between improvements on the New York specific state exam and improvements on the ITBS exam impossible.

---

[33] We also investigated the effects of being 200 and 600 meters from the Zone and obtained the same results.

[34] Whether general learning or the willingness and ability to prepare for an important exam is most correlated with longer term outcomes (health, education, crime, wages) is a crucial open question (Duckworth et al., 2006; Duckworth et al., 2007; Segal, 2007).

[35] More information on the ITBS is available at http://www.education.uiowa.edu/itp/itbs/

With these important caveats in mind, figures 6A and 6B report ITBS math and reading comprehension scores for HCZ middle school students. ITBS scores are not available in 6th grade for the 2005 cohort, so we plot scores of the two cohorts separately. The sample is restricted to students who have valid scores in each grade. The ITBS trends are similar, if more modest, than the results using state test scores. The 2006 cohort scored 40.8 and 34.7 on math and reading respectively in the fall of 6th grade, and 48.7 and 41.1 in the spring of 7th grade. To put this in context, the typical Black 8th grade student in New York City scored at about the 25th percentile of the 2007 NAEP math exam, and about the 23rd percentile of the 2007 NAEP reading exam.

HCZ Promise Academy -- Elementary School

The HCZ Promise Academy elementary-school lotteries, which admitted students in kindergarten for Fall 2004, 2005, and 2006, have never been significantly oversubscribed.[36] As a result, nearly all lottery "losers" are eventually offered admission into the public charter school. This prevents us from estimating the ITT and TOT effects as cleanly as we did for the middle-school students. Instead, we use two statistical techniques. First, we construct IV estimates using a student's address interacted with cohort as an instrument similar to those presented in Table 3. Second, we use a student's waitlist number as an instrument in the sample of students who entered the lottery. Under a number of plausible identifying assumptions (a low waitlist number is correlated with enrollment, one's position on the waiting list is random, being higher on the waiting list does not make one more likely to enroll, and that a low waitlist position only influences outcomes through its effect on enrollment), we can identify the causal effect of attending the HCZ Promise Academy elementary public charter school on student outcomes. Our variation in enrollment comes primarily from the timing of the enrollment offer. Students with a winning lottery number or a very low waitlist number were typically offered a spot at the HCZ school in the early summer, while students with a very high waitlist number are typically not offered a spot until late summer or early fall, after the school year had already commenced.

Table 5 presents estimates from our main IV strategy. The reported coefficient is an indicator of being enrolled at the HCZ public charter school in kindergarten, which is instrumented for using the interaction between a student's address (in or out of the Zone's boundaries) and cohort. As before, we restrict the sample to students with test scores no greater than three standard deviations from the mean and no more than 50 absences, but results are qualitatively similar if all students are included in the regressions. Standard errors, clustered at the cohort level, are reported in parentheses.

The effect of being enrolled at the elementary public charter school on third-grade test scores -- the first year that children in New York take standardized exams – is large, with point estimates ranging

[36] We are still waiting on Harlem Children's Zone to assemble the lottery files from the 2006 kindergarten cohort.

from 1.117 (0.581) to 1.296 (0.417) standard deviations in math and 1.389 (0.601) to 1.570 (0.521) in ELA. Results for fourth grade math are more modest, ranging from 0.446 (0.242) to 0.669 (0.197), while our estimates for fourth grade ELA are consistently large (1.052 (0.418) to 1.479 (0.376)). Even our modest estimates suggest that the HCZ elementary school eliminates racial gaps in both math and ELA. The relatively large gains in ELA are particularly noteworthy in light of our middle-school results, suggesting that deficiencies in ELA might be addressed if intervention occurs relatively early in the child's life. This is consistent with developmental research that shows that the critical period for language development occurs early in life, while the critical period for developing higher cognitive functions extends into adolescence (Hopkins and Bracht, 1975; Newport, 1990; Pinker, 1994; Nelson, 2000; Knudsen et al., 2006).

Table 6 presents a series of estimates for the elementary school lottery sample using a student's waitlist number as an instrumental variable.[37] The IV estimates suggest that children who attend the elementary school gain 0.644 (0.436) standard deviations in math and 0.994 (0.410) standard deviations in ELA by third grade, and 1.032 (0.569) standard deviations in math and 0.476 (0.514) standard deviations in ELA by fourth grade, but imprecision makes definitive conclusions impossible. Students have 8.131 (4.617) fewer absences in first grade, 10.002 (4.334) fewer in second and 6.915 (4.486) fewer in third grade. HCZ students also appear to be more likely to matriculate on time in the lottery results. Given that the IV results in Table 5 indicate that students are less likely to matriculate on time and have about the same number of absences, we suggest interpreting both sets of results with caution.

Iowa Test of Basic Skills (ITBS) test scores of HCZ elementary students are plotted in Figures 7A and 7B. As with the middle school cohort, scores are not available for the control group, so the evidence is only suggestive. We report National Percentile Rank scores, which compares the score of a student to a national sample of students who are in the same grade and who were tested at the same time of year (fall or spring). As a result of this standardization, having the same score across time implies "normal" academic progress. ITBS scores are not available in Kindergarten for the 2004 cohort, so we plot scores of the two cohorts separately. The sample is restricted to students who have valid scores in each grade. Students do not take the reading comprehension portion of the exam until first grade, so we report the average of the word analysis and vocabulary subset scores instead.

Relative to the middle school cohort, ITBS scores for the elementary cohorts are noisier, making interpretation of the results more difficult. It appears that students gain relative to their national peers in

[37] We use a dichotomous instrument indicating whether a student received an offer before the start of classes in the fall of their Kindergarten year. This information was available in the HCZ lottery files, which indicated the date a student was called from the waitlist. In the pooled first stage regression, being offered a spot at HCZ before the start of the school year is associated with a 0.24 (0.05) increase in the probability of enrolling at the HCZ Promise Academy.

all subject areas in their first year to year and a half at the Promise Academy, but make only normal progress after that. Elementary students start with a much smaller deficit than their middle school counterparts and are at or above the national average after one to two years of attending the Promise Academy. To put this in context, the typical Black 4th grade student in New York City scored in about the 31st percentile of the 2007 NAEP math exam, and about the 30th percentile of the 2007 NAEP reading exam.

The Baby College and Harlem Gems

The Baby College (the nine-week parenting program) and Harlem Gems (the preschool program) do not admit students in such a way that one can confidently identify their impact and our IV strategy employed earlier does not have enough power in such small samples.[38]

Panel A of Table 7 fits least squares models to the entire New York City sample. This estimate represents the partial correlation of attending Harlem Gems or The Baby College on our set of third grade outcomes. To the extent that more motivated parents put their children in programs such as The Baby College and Harlem Gems, our correlations are an underestimate of the true effect. If, on the other hand, HCZ staff recruits less-motivated parents, our correlations are overestimates. As such, we strongly caution against any causal interpretation of the forthcoming estimates. We control for gender, race and lunch status.[39] Consistent with our main results, the association between third grade test scores and attending the Promise Academy is positive and statistically significant. Harlem Gems and The Baby College, on the other hand, are associated with a *decrease* in third grade math and ELA scores, though the neither estimate is statistically significant.

Results from Panel A may be driven by negative selection into Harlem Gems or The Baby College. To partially account for this, Panel B of Table 7 presents results restricting the sample to students living in or around the Zone and Panel C presents results restricting the sample to students who entered the 2004 or 2005 Kindergarten Lottery. Results from these two samples are qualitatively similar to the full New York City sample. Point estimates on the Harlem Gems range from -0.081 (1.140) to 0.116 (0.185) for math and -0.080 (0.168) to 0.036 (0.138) for ELA. Estimates for The Baby College range from -0.223 (0.156) to -0.049 (0.168) for math and -0.172 (0.148) to -0.108 (0.155) for ELA.[40] None of the coefficients are statistically significant. Taken at face value, our point estimates suggest that

[38] The Baby College and Harlem Gems began earlier than the Promise Academies, and as a result we have very few "untreated" cohorts to compare to.

[39] Using a student's address to calculate their block-group level income and adding this income as a covariate does not alter the results.

[40] Results are qualitatively similar in all samples if school or teacher fixed effects are included.

neither Harlem Gems nor The Baby College positively impact math or reading scores. The estimates are so imprecise, however, that no definitive conclusions can be drawn.

The results from Table 7 are a bit surprising. An obvious explanation is that the effects of Harlem Gems and The Baby College, like many other early investments, fade by the time children are in third grade. To understand whether this is driving the meager results, Figures 8 and 9 show Iowa Test of Basic Skill test score trends for the 2005 Kindergarten Cohort separated by Harlem Gems and Baby College enrollment. Panels A and B of Figure 8 show results for Harlem Gems. We adjust scores for gender and free lunch and plot the predicted scores.[41] There are no observable differences in adjusted ITBS at any point between students who attended the Harlem Gems and those that did not. Panels A and B of Figure 9 repeat the analysis for The Baby College. Similar to the Harlem Gems results, there are no observable differences in adjusted ITBS at any point students who attended The Baby College and those that did not.

To dig deeper into The Baby College results, we also examine the partial correlation between attending The Baby College and the Bracken School Readiness Assessment for the subset of children who attended the 2001-2007 cohorts of Harlem Gems. The Bracken test was administered before students attended Gems (three years before the test outcomes in Table 7), and has been normalized to have mean zero and a standard deviation of one. Table 8 presents these results. The raw correlation between attending The Baby College and Bracken Scores is 0.045 (0.111) standard deviations. Controlling for cohort effects, age in months, parental education and household income only makes the partial correlation of The Baby College on Bracken smaller. Accounting for all of our available controls, attending The Baby College is associated with a 0.009 (0.111) standard deviation gain in school readiness.[42]

We fail to find any empirical evidence that Harlem Gems or The Baby College is positively associated with achievement. There are a few potential explanations. Perhaps the most obvious is that Harlem Gems and The Baby College are simply not effective programs. The Nurse-Family Partnership, one of the more successful early childhood programs with a parenting component, involves weekly visits from a registered nurse beginning in a mother's first trimester. These visits continue through a child's second birthday (though less frequently as a child ages), with over 50 total visits before mothers exit the program. In contrast, The Baby College uses paraprofessionals in a treatment that lasts only nine weeks. Second, our estimates are not causal and there may be thorny issues of selection that remain even after controlling for family demographics.[43] A final explanation is that assessments of academic achievement

---

[41] There is positive selection on observables into Harlem Gems. As a result the unadjusted graphs follow the same trends, but with Harlem Gems scores somewhat higher at each grade than the non-Harlem Gems scores.

[42] Estimates from our main IV strategy using the interaction between a student's address and cohort year also suggest that the impact of The Baby College on third-grade test scores is zero, but the results are sensitive to small specification changes and have a weak first stage.

[43] We find no evidence of negative selection into The Baby College or Harlem Gems based on observable characteristics. Regressing an indicator for The Baby College or Harlem Gems enrollment on demographic

may be the wrong outcomes to assess the effectiveness of an early education or parenting program. It is quite possible, for instance, that The Baby College is more likely to assist parents when kids are absorbing higher-order thinking and will help them engage with their children so they are less likely to engage in criminal activities, and so on. More definitive conclusions about the true impact of Harlem Gems and The Baby College will be better understood with the passage of time, broader data collection, and planned experimentation.

V. Communities, Schools, or Both?

HCZ students are exposed to a web of community services along with education innovations ranging from a longer school day and year to after-school programs to mental and physical health services. These programs fall into three broad categories: (1) community programs available to anyone in the 97 blocks that comprise the Zone; (2) student-family programs that are only available to HCZ students and their families, and (3) student programs that are only available to HCZ students. Thus far, we have loosely referred to (1) as the community programs and (2) and (3) as school programs. Let us consider each bundle in turn.

The community services – community centers, the truancy-prevention programs, the network of targeted programs such as the asthma and obesity initiatives, The Baby College, Harlem Gems, and so on – are available to any child in HCZ. These programs may plausibly affect student outcomes in any number of ways, from mitigating the physical and emotional barriers to success that typical urban students face, to providing a more supportive out-of-school learning environment.

Beyond The Baby College and Harlem Gems results, there are three pieces of evidence that suggest that the combined effect of the community bundle alone is not responsible for the gains in achievement. First, our main IV strategy compares children inside the Zone's boundaries relative to other children in the Zone that were ineligible for the lottery so the estimates are purged of the community bundle. Recall, our IV estimates are larger than the lottery estimates, however, suggesting that communities alone are not the answer. Second, in our analysis of subsamples, children inside the Zone garnered the same benefit from the schools as those outside the Zone, suggesting that proximity to the community programs is unimportant. Third, the Moving to Opportunity experiment, which relocated individuals from high-poverty to low-poverty neighborhoods while keeping the quality of schools roughly constant, showed null results for girls and negative results for boys (Sanbonmatsu et al., 2006; Kling et

characteristics and indicators for being eligible for free and reduced price lunch, we find that Harlem Gems enrollees are somewhat *less* likely to be eligible for free lunch than other students in the Kindergarten Lottery, the Zone, or New York as a whole, and Baby College enrollees are neither more or less likely to be eligible for free lunch than other students in the available samples. In the Harlem Gems sample, Baby College enrollees are, if anything, somewhat more advantaged, being more likely to have a parent that is college educated.

al., 2007). This suggests that a better community, as measured by poverty rate, does not significantly raise test scores if school quality remains essentially unchanged. Additionally, and far more speculative, there is substantial anecdotal evidence that the Children's Zone program was unsuccessful in the years before opening the public charter schools.[44] Indeed, the impetus behind starting the schools was the lack of test-score growth under the community-only model.

The bundle of student-family services, such as the provision of nutritious fruits and vegetables, pre-made meals, money and travel allowances to ensure kids get to school, and general advice on how to support their child's learning, are only available to HCZ students and their families. A simple test of this hypothesis is to compare outcomes of children in the HCZ Promise Academy public charter schools to their siblings who are in regular public schools. If the student-family services provide inputs to the entire family, the siblings of the charter-school students will benefit.

We instrument a student's enrollment status with her lottery number.[45] Under several assumptions (that the treatment-group assignment is random, lottery losers are not allowed to enroll, and winning the lottery only affects outcomes through charter-school enrollment) we can estimate the causal impact of a student enrolling in the public charter school on her sibling's outcomes. We define siblings as any two students who share the same last name and address at any point in time.

Table 9 presents our primary sibling-peer results in the middle-school sample.[46] Each regression controls for the sibling's pre-lottery outcomes, gender, race and lunch status. The reported coefficient is an indicator for whether the student was enrolled in the HCZ Promise Academy in sixth grade, using that student's lottery number as an instrument. Outcomes are for two years after a student was entered in the HCZ lottery, the most recent data we have available.[47] Standard errors are clustered at the family level.[48] The sample is restricted to siblings that did not attend the HCZ Promise Academy themselves. Results are qualitatively similar if all siblings are used.

---

[44]In *Whatever it Takes: Geoffrey Canada's Quest to Change Harlem and America*, Paul Tough writes, "A few years into the life of the Zone, Canada hit a snag. The problem was the schools. His original plan called for his staff to work closely with the principals of Harlem's local public schools, providing them with supplemental services like computer labs and reading programs…To Canada's surprise and displeasure, principals sometimes resisted the help, turning down his requests for classroom space or kicking out the tutors that the organization supplied. Even in the schools where the programs were running smoothly, they didn't seem to be producing results: the neighborhood's reading and math scores had barely budged" (Tough, 2008).

[45] The siblings sample is too small to execute our main instrumental variables strategy; the corresponding first stage is weak. The main IV results generally correspond to the results reported here, but are less precise and less robust to specification changes.

[46] Most siblings of the HCZ elementary students do not have valid test scores (e.g. are too young) or are also enrolled in the HCZ schools. We therefore concentrate our analysis on siblings of the middle-school cohort.

[47]Results are qualitatively similar if we allow the effect of enrollment to vary by dosage and use outcomes from the same calendar year.

[48]To address concerns that the subset of HCZ students with siblings may be different than the subset without siblings, we replicate our main results from Table 3 for both groups. Students with and without siblings experience nearly identical gains to HCZ enrollment, and are quite similar on observable characteristics.

The effect of a student's enrollment in HCZ appears to be relatively small or even negative on her sibling's test scores; siblings of charter-school students gain approximately 0.053 (0.337) standard deviations in math and lose 0.125 (0.275) standard deviations in ELA, but measured so imprecisely that we cannot make firm conclusions.[49] There is a large effect of a student's enrollment on their sibling's absences, with siblings missing 7.961 (5.062) fewer days on average, though the effect is statistically significant only at the 10 percent level. Taken at face value, this suggests that the student-family bundle of services may decrease absences, but there is no effect on achievement test scores.

We have provided some evidence that HCZ's success in raising test scores is unlikely to be driven by the bundle of community services, either directly or indirectly, and that the effects of the student-family programs on test-scores are, at best, modest. This suggests that either the HCZ Promise Academy public charter schools are the main driver of our results or the interaction of the schools and community investments is the impetus for such success.

An important remaining question is why HCZ charters are so effective at educating the poorest minority students. With trepidation, we will hazard a guess. With the available data it is impossible to disentangle which particular school input is driving the results, but it is likely not teacher incentives alone (Fryer, 2009b), social workers alone (Kendziora et al., 2008), or student incentives alone (Fryer, 2009a). Each one of these initiatives may have a positive return on investment, but they are not likely to produce such dramatic results in isolation.[50]

HCZ teachers or administrators, who are partly rewarded based on student achievement in various ways, could be cheating. There have been documented cases of cheating in California (May, 1999), Massachusetts (Marcus, 2000), New York (Loughran and Comiskey, 1999), Texas (Kolker, 1999), Great Britain (Hofkins, 1995; Tysome, 1994) and Chicago (Jacob and Levitt, 2003). Using an algorithm similar to Jacob and Levitt (2003), we implement two statistical tests of cheating at HCZ. First, we test whether HCZ students are more likely to have an unusual block of consecutive identical answers, and second we test whether HCZ students systematically underperform on easy questions while overperforming on hard questions.[51] We find that neither approach suggests that there is cheating at HCZ. On 3rd and 8th grade ELA and Math for both cohorts (8 total measures), HCZ is between the 42nd and 72nd percentiles in the

[49] Comparing the standardized scores of students across grades imposes a number of functional form restrictions, so we also consider the effect of enrollment on a sibling's percentile score rank. Results are qualitatively similar, with point estimates small and not statistically significant.

[50] There may also be important interactions between different elements of the HCZ school bundle. Analogous to the discovery of HAART that treats HIV/AIDS, each element of the bundle may not be effective individually but together one receives both the direct effect of the program and the interactions between elements of the bundle.

[51] Jacob and Levitt (2003) also search for large, unexpected increases in test scores one year, followed by very small test score gains (or even declines) the following year. Their identification strategy exploits the fact that these two types of indicators are very weakly correlated in classrooms unlikely to have cheated, but very highly correlated in situations where cheating likely occurred. We cannot use this second measure, as administrative cheating at HCZ would imply large gains that are maintained as long as students attend.

"suspiciousness" of its answer strings. Adjusting for school size to account for the number of strings we search over, HCZ is between the 43rd and 60th percentiles. To test whether HCZ students answer "too many" hard questions correctly – a possible indication of cheating if teachers are changing those answers - we plot residuals from regression models predicting performance on each question in Appendix Figures 2A through 2H. Conditional on overall test performance, HCZ students appear no more likely to answer hard questions correctly than other students in the city. Additional results and details are available in Appendix C.

It is plausible that high-quality teachers in the public charter schools are responsible for the results as the estimates are within the variance of teacher quality in Rockoff (2004), Hanushek and Rivkin (2005), Aaronson et al. (2007), and Kane and Staiger (2008), which range from 0.15 to 0.24 standard-deviations in math and 0.15 to 0.20 in reading. Recall, in their quest to find high-quality teachers, the teacher-turnover rate at the HCZ charter was 48 percent in their first year.[52]

Second, a linear combination of good policy choices can explain the results. In their analysis of New York City charter schools, Hoxby and Murarka (2009) estimate the relationship between a series of school policy choices (time on task, e.g.) and the success of the charter school. Plugging in HCZ's combination of policies into the regression equation utilized in Hoxby and Murarka (2009) predicts yearly gains of 0.54 standard deviations, which would account for our results.[53] Indeed, this estimate is larger than any other charter school in their sample, suggesting that either HCZ has a unique blend of policies or the community investments serve as a vital technology shifter in the production of achievement.

VI. Conclusion

The racial achievement gap in education is one of America's most pressing social concerns. The typical black seventeen year-old reads at the proficiency level of the typical white thirteen year-old (Campbell et al., 2000). On the Scholastic Aptitude Test, the gateway to America's institutions of higher

52 The ideal test of this hypothesis would be to assign disadvantaged students to above average teachers in grades 6 through 8. Unfortunately, this rarely happens in practice. In our data less than one percent of black students are consecutively assigned to 6th, 7th and 8th grade teachers who have value added scores that are at least one standard deviation above the mean.

53 We use the regression coefficients reported on page 59 of Hoxby and Muraka (2009) and HCZ survey responses to construct our estimate. HCZ has the following characteristics (in order listed by Hoxby and Muraka, 2007, with associated regression coefficients in parentheses): has been operating for 4 years (-0.009), is a Charter Management Organization (-3.660), has 210 instructional days (0.021) and 8 instructional hours per day (-0.077), conducts Saturday school (0.153) and an optional after school program (0.058), uses a combination of original curriculum and the Core Knowledge curriculum (0.137/2 + 0.072/2), has an average class size of 20 (0.002), administers internal evaluations (-0.085), requires school uniforms (-0.179) and a dress code (0.139), does not have a broken windows disciplinary policy, does require a parent contract (-0.234) and reserve seats for parents on the board (0.233) and has one school leader (0.199).

learning, there is little overlap in the *distribution* of scores (Card and Rothstein, 2007). In the past decades, there has been very little progress in solving the achievement gap puzzle.

HCZ is enormously successful at boosting achievement in math and ELA in elementary school and math in middle school. The impact of being offered admission into the HCZ middle school on ELA achievement is positive, but less dramatic. High-quality schools or community investments coupled with high-quality schools drive these results, but community investments alone cannot.

As the Obama administration and other governments around the world decide whether and how to use the Harlem Children's Zone model to combat urban poverty, costs is an important consideration. The New York Department of Education provided every charter school, including the Promise Academy, \$12,443 per pupil in 2008-2009. HCZ estimates that they added an additional \$4,657 per-pupil for in-school costs and approximately \$2,172 per pupil for after-school and "wrap-around" programs. This implies that HCZ spends \$19,272 per pupil. To put this in perspective, the median school district in New York State spent \$16,171 per pupil in 2006, and the district at the 95th percentile cutpoint spent \$33,521 per pupil (Zhou and Johnson, 2008). The total budget for the HCZ – including its community and school investments – is roughly \$50 million per year.

Taken at face value, the achievement gains of HCZ students will translate into dramatically improved life trajectories. Our middle school lottery estimates – the most modest estimated - suggest a 9.7 to 12.5 percent increase in earnings (Neal and Johnson, 1996; Currie and Thomas, 1999), a 2.8 to 3.7 percent decrease in the probability of committing a property or violent crime (Levitt and Lochner, 2001), and a 12.5 to 18.8 decrease in the probability of having a health disability (Auld and Sidhu, 2005; Elias, 2005; Kaestner, 2009). If HCZ affects educational attainment as dramatically as achievement, the implied benefits are enormous (e.g. Angrist and Krueger, 1991; Grossman and Kaestner, 1997; Card, 1999; Lantz et al., 1998; Card, 2001; Lochner and Moretti, 2004; Pettit and Western, 2004; Lleras-Muney, 2005; Belfield, 2006; Cutler and Lleras-Muney, 2006; Rouse, 2006; Levin et al., 2007; Oreopoulos, 2007; Kaestner, 2009). The public benefits alone from converting a high school dropout to graduate are more than \$250,000.[54]

We hope that our analysis provides a sense of optimism for work on the achievement gap. The HCZ model along with recent results in Abdulkadiroglu et. al. (2009) demonstrate that the right cocktail of investments can be successful. The challenge is to find even more efficient ways to achieve similar achievement gains in regular – non charter – public schools, so that all children, independent of their race or their parent's income, can receive a high quality education.

[54] See Appendix D for details on all of the cost-benefit calculations.

References

Aaronson, Daniel, Lisa Barrow and William Sander (2007) "Teachers and Student Achievement in the Chicago Public High Schools." *Journal of Labor Economics*, 25(1): 95 – 135.

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

Achilles, Charles M., Barbara A. Nye, Jayne B. Zaharias, and B. DeWayne Fulton (1993) "The lasting benefits study (LBS) in grades 4 and 5 (1990-91): A legacy from Tennessee's four-year (K-3) class size study (1985-1989), Project STAR." HEROS, Inc.

Anderson, Michael (2008) "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." *Journal of the American Statistical Association*, forthcoming.

Angrist, Joshua, and Alan Krueger (1991) "Does Compulsory School Attendance Affect Schooling and Earnings?" *The Quarterly Journal of Economics*, 106 (4): 979-1014.

Angrist, Joshua D., and Kevin Lang (2004) "How Important Are Peer Effects? Evidence from Boston's Metco Program," *American Economic Review*, 94: 1613-34.

Auld, Christopher and Nirmal Sidhu (2005), "Schooling and Cognitive Ability." *Health Economics*, 14(10): 1013-1034.

Belfield, Clive R. (2006) "The consequences of raising the graduation rate for black males: The effects on crime," Working paper, Teachers College, New York.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan (2004) "How Much Should We Trust. Differences-in-Differences Estimates?" *Quarterly Journal of Economics*, 119(1): 249-275.

Black, Sandra, Paul Devereux, Kjell Salvanes. "Staying In the Classroom and Out of the Maternity Ward? The Effects of Compulsory Schooling Laws on Teenage Births." *Economic Journal*, 118(530): 1025 – 1054.

Brazelton, T. Berry (1992) *Touchpoints: Your Child's Emotional and Behavioral Development*. Reading Massachusetts: Perseus Books.

Borman, Geoffrey D., and Gina M. Hewes (2003) "The long-term effects and cost-effectiveness of Success for All." *Educational Evaluation and Policy Analysis*, 24: 243-266

Borman, Geoffrey D., Robert E. Slavin, Alan C.K. Cheun, Anne M. Chamberlain, Nancy A. Madden, and Bette Chambers (2007) "Final reading outcomes of the National Randomized Field Trial of Success for All." *American Educational Research Journal*, 44(3): 701-731.

Campbell, Jay, Catherine Hombo, and John Mazzeo (2000) "NAEP 1999 Trends in Academic Progress: Three Decades of Student Performance." NCES 2000-469. Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Card, David (1999) "The Causal Effect of Education on Earnings," in Card, David, and Ashenfelter, Orley (eds.), *Handbook of Labor Economics,* Vol. 3, (1802-1859).

Card, David (2001), "Estimating the Returns to Schooling: Progress on Some Persistent Econometric Problems," *Econometrica,* 69 (5): 1127-1160.

Card, David, and Jesse Rothstein (2007) “Racial Segregation and the Black-White Test Score Gap.” *Journal of Public Economics* 91(11-12): 2158-2184.

Carneiro, Pedro, and James Heckman (2002) “Human Capital Policy,” University of Chicago working paper.

Chenoweth, Karin (2007) *It's Being Done: Academic Success in Unexpected Schools*. Cambridge, MA: Harvard Education Press.

Coleman, James, Ernest Campbell, Carol Hobson, James McPartland, Alexander Mood, Frederic Weinfeld, and Robert York (1966) Equality of Educational Opportunity. Washington, DC: U.S. Government Printing Office.

Cook, Thomas D., H. David Hunt, and Robert F. Murphy (2000) “Comer’s School Development Program in Chicago: A theory-based evaluation.” *American Educational Research Journal* 37(2): 535-597.

Cullen, Julie Berry, Brian A. Jacob, and Steven D. Levitt (2005) “The Impact of School Choice on Student Outcomes: An Analysis of the Chicago Public Schools.” *Journal of Public Economics*, Volume 89(5-6): 729– 760.

Currie, Janet, and Duncan Thomas (1995) “Does Head Start Make a Difference?” *American Economic Review* 85(3), 341–364

Currie, Janet and Duncan Thomas (2000) "School Quality And The Longer-Term Effects Of Head Start," *Journal of Human Resources*, 35(4): 755-774.

Currie, Janet, and Duncan Thomas (2001), “Early Test Scores, School Quality and SES: Long Run Effects on Wage and Employment Outcomes,” in Polachek, Solomon W. (Ed.), *Worker wellbeing in a changing labor market*, Kidlington, Oxford: Elsevier Science Ltd.

Cutler, David, and Adriana Lleras-Muney (2006). NBER Working Paper No. 12352.

Decker, Paul, Daniel Mayer and Steven Glaserman (2004) “The Effects of Teach for America on Students: Findings from a National Evaluation.” Mathematica Policy Research Report No. 8792-750, June 9, 2004.

Deming, David (2009) “Early Childhood Intervention and Life-Cycle Skill Development: Evidence from Head Start.” Forthcoming in *American Economic Journal: Applied Economics*

Duckworth, Angela, Christopher Peterson, Michael Matthews and Dennis Kelly (2006) “Grit: Perseverance and Passion for Long-Term Goals." Mimeo University of University of Pennsylvania.

Duckworth, Angela, Patrick Quinn, Donald Lynam, Rolf Loeber, Magda Stouthamer-Loeber, Terrie Moffit, and Avshalom Caspi (2007) “What intelligence tests test: Individual differences in test motivation and IQ.” Mimeo University of University of Pennsylvania.

Dobbie, Will and Roland Fryer (2009) "Do Community Programs Help Kids in the Community? Evidence from the Harlem Children's Zone" Unpublished Draft.

Elias, Julio (2005), "The Effects of Ability and Family Background on Non-monetary Returns to Education." University of Chicago Working Paper.

Fryer, Roland (2009a) "Financial Incentives and Student Achievement: Evidence from Randomized Trials." Unpublished draft.

Fryer, Roland (2009b) "School-Based Merit Pay: An Experiment in 200 New York City Public Schools." Unpublished draft.

Fryer, Roland (2009c) "Understanding the Racial Achievement Gap in Urban Schools" Unpublished draft.

Fryer, Roland, and Steven Levitt (2004) "Understanding the Black-White Test Score Gap in the First Two Years of School." *The Review of Economics and Statistics*, 86(2): 447-464.

Fryer, Roland, and Steven Levitt (2006) "The Black-White Test Score Gap Through Third Grade." *American Law and Economic Review*, 8(2): 249-281.

Fryer, Roland, and Steven Levitt (forthcoming) "Testing for Racial Differences in the Mental Ability of Young Children." *American Economic Review*

Hanushek, Eric, John Kain, Steven Rivkin, and Gregory Branch (2005) "Charter School Quality and Parental Decision Making with School Choice." NBER Working Paper 11252.

Hanushek, Eric and Steven Rivkin (2005) "Teachers, Schools and Academic Achievement." *Econometrica*, 73(2): 417 – 458.

Hart, Betty and Todd Risley (1995). *Meaningful Differences in the Everyday Experience of Young American Children.* Baltimore, MD: Brookes Publishing.

Hastings, Justine, Thomas Kane, and Douglas Staiger (2006) "Gender and Performance: Evidence from School Assignment by Randomized Lottery." *American Economic Review*, 96(2): 232-236.

Hofkins, Diane (June 16, 1995) "Cheating 'rife' in national tests." New York Times Educational Supplement.

Hopkins, Kenneth and Glenn Bracht (1975) "Ten-Year Stability of Verbal and Nonverbal IQ Scores." *American Educational Research Journal*, 12(4): 469–477.

Hoxby, Caroline, and Sonali Murarka (2009) "Charter Schools In New York City: Who Enrolls and How They Affect Their Students' Achievement." NBER Working Paper 14852.

Hoxby, Caroline, and Jonah Rockoff (2004) "The Impact of Charter Schools on Student Achievement." Working Paper

Imbens, Guido W., and Joshua D. Angrist (1994) "Identification and Estimation of Local Average Treatment Effects." *Econometrica*, 62(2): 467 – 475.

Jacob, Brian (2004) "Public Housing, Housing Vouchers, and Student Achievement: Evidence from Public Housing Demolitions in Chicago." *American Economic Review*, 94(1), 233-258.

Jacob, Brian (2005) "Accountability, Incentives and Behavior: Evidence from School Reform in Chicago." *Journal of Public Economics*, 89(5-6): 761-796.

Jacob, Brian, and Lars Lefgren (2004) "Remedial Education and Student Achievement: A Regression-Discontinuity Analysis." *Review of Economics and Statistics*, 86(1): 226-244

Jacob, Brian, and Steven Levitt (2003) "Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating." *Quarterly Journal of Economics* 117(3): 843-878.

Jacob, Brian, and Jens Ludwig (2008) "Improving Educational Outcomes for Poor Children." NBER Working Paper 14550.

Jepsen, Christopher, and Steven G. Rivkin (2002) "What is the Tradeoff Between Smaller Classes and Teacher Quality?" NBER Working Paper 9205.

Jencks, Christopher, and Meredith Phillips (1998) "America's Next Achievement Test: Closing the Black-White Test Score Gap," *American Prospect*, 40: 44–53.

Kaestner, Robert (2009), "Adolescent Cognitive and Non-Cognitive Correlates of Adult Health," NBER Working Paper 14924.

Kane, Thomas, Jonah Rockoff and Douglas Staiger (2008) "What does certification tell us about teacher effectiveness? Evidence from New York City." *Economics of Education Review*, 27(6): 615-631.

Kane, Thomas, and Douglas Staiger (2008) "Estimating Teacher Impacts on Student Achievement: An Experimental Validation." NBER Working Paper 14607.

Katz, Lawrence, Jeffrey Kling, and Jeffrey Liebman (2001) "Moving to Opportunity in Boston: Early Results of a Randomized Mobility Experiment." *Quarterly Journal of Economics*, 116(2): 607-654.

Kendziora, Kimberly, Wehmah Jones, Delphinia Brown, David Osher, Marc Rudolph, Katje King. Shefali Trivedi, and Pam Cantor (2008) "Safe Schools, Successful Students Initiative: Final Report to the United Way."

Kling, Jeffrey, Jeffrey Liebman, and Lawrence Katz (2007) "Experimental Analysis of Neighborhood Effects." *Econometrica*, 75(1): 83–119.

Knudsen, Eric, James Heckman, Judy Cameron, and Jack Shonkoff (2006) "Economic, neurobiological, and behavioral perspectives on building America's future workforce." *Proceedings of the National Academy of Sciences*, 103(27): 10155–10162.

Kolker, Claudia (April 14, 1999) "Texas Offers Hard Lessons on School Accountability." Los Angeles Times.

Krueger, Alan (1999) "Experimental Estimates of Education Production Functions." *Quarterly Journal of Economics*, 114(2): 497-532

Krueger, Alan, and Pei Zhu (2004) “Another Look at the New York City School Voucher Experiment.” *American Behavioral Scientist*, 47(5): 658-698

Krueger, Alan, and Diane Whitmore (2001) The Effect of Attending a Small Class in the Early Grades on College Test-Taking and Middle School Test Results: Evidence from Project STAR. *Economic Journal*, 111 (January), 1-28.

Lantz, P. M., House, J. S., Lepkowski, J. M., Williams, D. R., Mero, R. P., and Chen, J. (1998) “Socioeconomic factors, health behaviors, and mortality: Results from a nationally representative prospective study of US adults.” *Journal of the American Medical Association*, 279: 1703–1708.

Lauer, Patricia, Motoko Akiba, Stephanie Wilkerson, Helen Apthorp, David Snow, and Mya Martin-Glenn (2002) “Out-of-School-Time Programs: A Meta-Analysis of Effects for At-Risk Students.” *Review of Educational Research*, 76(2): 275-313

Levin, Henry M., Clive Belfield, Peter Muennig, and Cecilia Rouse (2007) "The public returns to public educational investments in African-American males," *Economics of Education Review* 26 (2007): 700-709.

Levitt, Steve, and Lance Lochner (2001) “The determinants of juvenile crime,” in J. Gruber (Ed.), *Risky Behavior Among Youths: An Economic Analysis* (327-373). Chicago: University of Chicago Press.

Lleras-Muney, Adriana (2005) "The Relationship between Education and Adult Mortality in the United States," *Review of Economic Studies*, 72 (1): 189-221.

Lochner, Lance, and Enrico Moretti (2004) “The effect of education on crime: Evidence from prison inmates, arrests, and self-reports,” *American Economic Review*, 94: 155–189.

Loughran, Regina and Thomas Comiskey (1999) “Cheating the Children: Educator Misconduct on Standardized Tests.” Report of the City of New York Special Commissioner of Investigation for the New York City School District.

Ludwig, Jens and Deborah Phillips (2007) “The Benefits and Costs of Head Start.” NBER Working Paper 12973.

Ludwig, Jens and Douglas Miller (2007) "Does Head Start Improve Children's Life Chances? Evidence from a Regression Discontinuity Design." *The Quarterly Journal of Economics*, 122(1): 159-208

Marcus, John (February, 2000) “Faking the Grade.” Boston Magazine.

May, Meredith (October 4, 1999) “State Fears Cheating by Teachers.” San Francisco Chronicle.

Neal, Derek and William Johnson (1996), “The role of premarket factors in black-white wage differentials.” *Journal of Political Economy*, 104: 869–95.

Neal, Derek (2006) “Why Has Black-White Skill Convergence Stopped?” *Handbook of Economics of Education*, edited by Eric Hanushek and Finis Welch, Elsevier.

Nelson, Charles (2000) “The Neurobiological Bases of Early Intervention.” In Jack Shonkoff and Deborah Phillips (Eds.), *Handbook of Early Childhood Intervention* (2nd edition). New York: Cambridge University Press.

Newport, Elissa (1990) “Maturational Constraints on Language Learning.” *Cognitive Science*, 14(1, Special Issue): 11–28.

Nye, Barbara, B. DeWayne Fulton, Jayne Boyd-Zaharias, and Van A. Cain (1995) “The Lasting Benefits Study, Eighth Grade Technical Report.” Nashville, TN: Center for Excellence for Research in Basic Skills, Tennessee State University.

Olds, David (2006) “The Nurse-Family Partnership.” In N.F. Watt, C. Ayoub, R.H. Bradley, J.E. Puma, and W.A. LeBoeuf (Eds.), The Crisis in Youth Mental Health: Early Intervention Programs and Policies (pps. 147-180). Westport, CT: Praeger.

Oreopoulos, Philip (2007) “Do dropouts drop out too soon? Wealth, health and happiness from compulsory schooling,” *Journal of Public Economics*, 91 (11-12): 2213-2229.

Pettit, Becky, and Bruce Western (2004) “Mass imprisonment and the life course: Race and class inequality in US incarceration,” *American Sociological Review*, 69: 151–169.

Phillips, Meredith, James Crouse, and John Ralph (1998) “Does the Black-White Test Score Gap Widen after Children Enter School?” (pp. 229–272), in Christopher Jencks and Meredith Phillips (Eds.), *The Black-White Test Score Gap*. Washington, DC: The Brookings Institute.

Pinker, Steven (1994) The Language Instinct: How the Mind Creates Language. New York: W. Morrow and Co.

Podgursky, Michael J., and Matthew G. Springer (2007) “Teacher Performance Pay: A Review.” *Journal of Policy Analysis and Management*, 26(4), 909-950.

Redd, Zakia, Stephanie Cochran, Elizabeth Hair, and Kristin Moore (2002) “Academic Achievement Programs and Youth Development: A Synthesis.” Child Trends, Washington D.C.

Rockoff, Jonah E. (2004) “The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data.” *American Economic Review*, 94(2): 247–252.

Rouse, Cecilia Elena (1998) “Private School Vouchers and Student Achievement: An Evaluation of the Milwaukee Parental Choice Program.” *Quarterly Journal of Economics*, 113(2): 553-602.

Rouse, Cecilia Elena (2006) “The economic consequences of inadequate education for black males: The effects on labor market income and tax revenue.” Working paper, Teachers College Equity Symposium.

Sanbonmatsu, Lisa, Jeffrey Kling, Greg Duncan and Jeanne Brooks-Gunn (2006) “Neighborhoods and Academic Achievement: Results from the Moving to Opportunity Experiment.” *Journal of Human Resources*, 41(4): 649 – 691.

Segal, Carmit (2008) “Motivation, Test Scores, and Economic Success.” UPF Working Paper 1124.

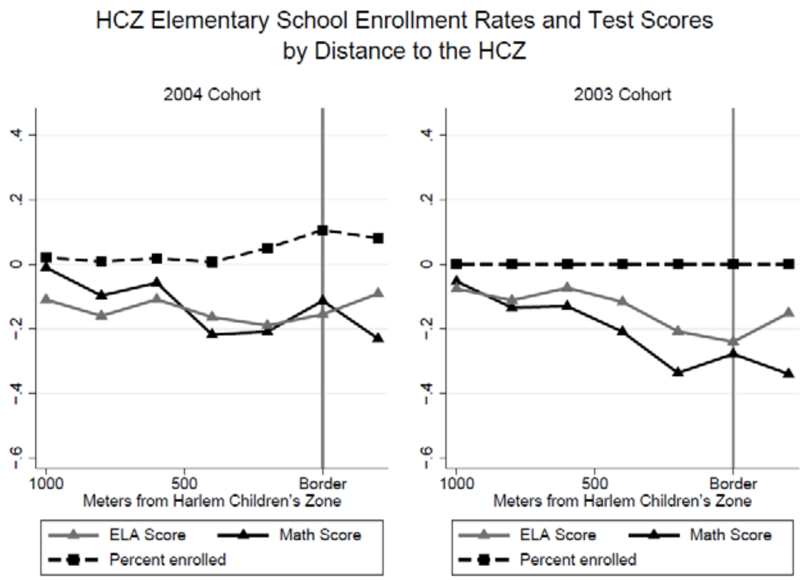
Tough, Paul (2008) *Whatever it Takes: Geoffrey Canada's Quest to Change Harlem and America*. New York City, NY: Houghton Mifflin Company.

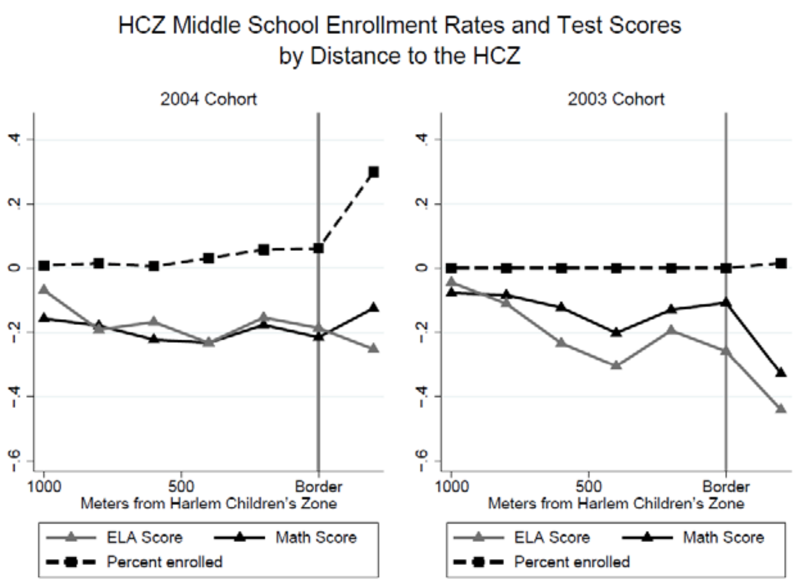
Tysome, Tony (August 19, 1994) “Cheating purge: Inspectors out.” New York Times Higher Education Supplement.

Zhou, Lel and Frank Johnson (2008) "Revenues and Expenditures for Public Elementary and Secondary School Districts: School Year 2005–06, Fiscal Year 2006 (NCES 2008–345)." National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Figure 1
The Harlem Children's Zone

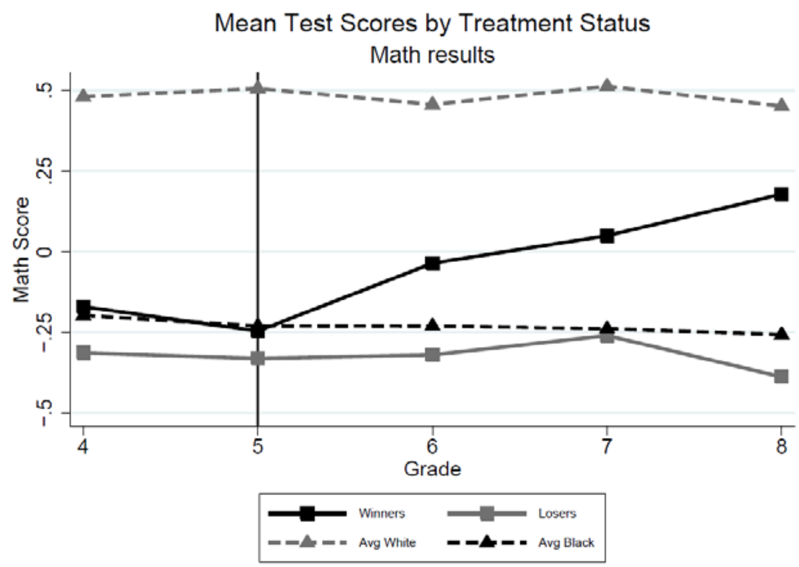
Figures 2A and 2B

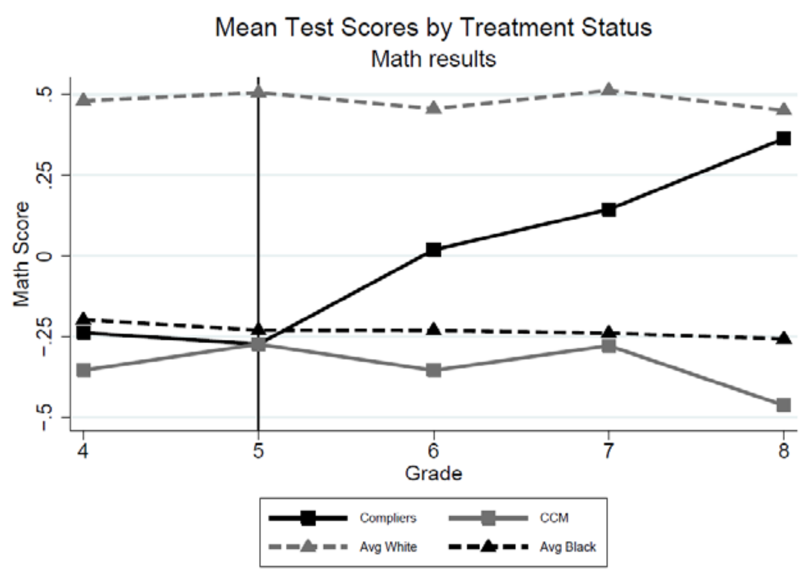




Notes: In our elementary-school sample a student's cohort is imputed from her initial kindergarten year. If we do not observe an elementary student in kindergarten, we impute the expected grade from the student's birth date. In our middle-school sample a student's cohort is imputed from her initial sixth-grade year.[1] If we do not observe a student in sixth grade, we impute an expected grade from the nearest observed grade. The HCZ is defined as the original 24-block Zone, ranging from 116th to 123rd Streets, 5th Avenue to 8th Avenue. Results are qualitatively similar if the Phase 2 or Phase 3 area is used.

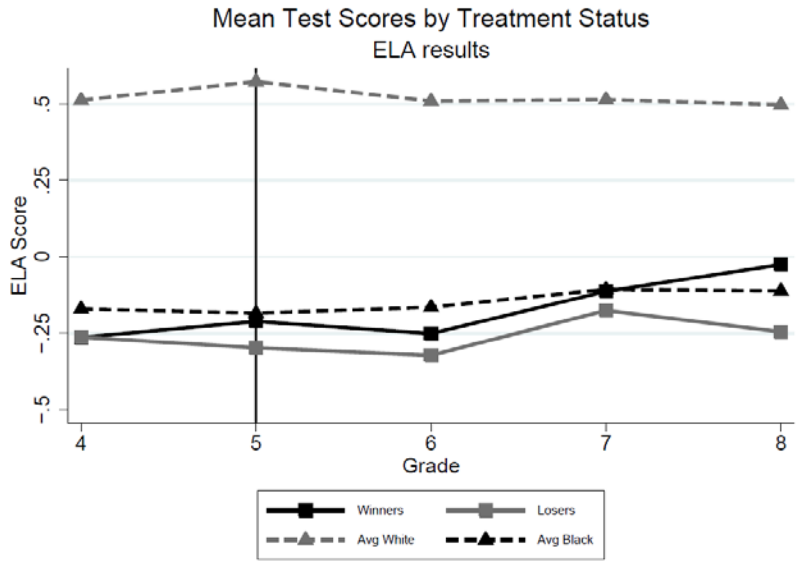
Figures 3A and 3B

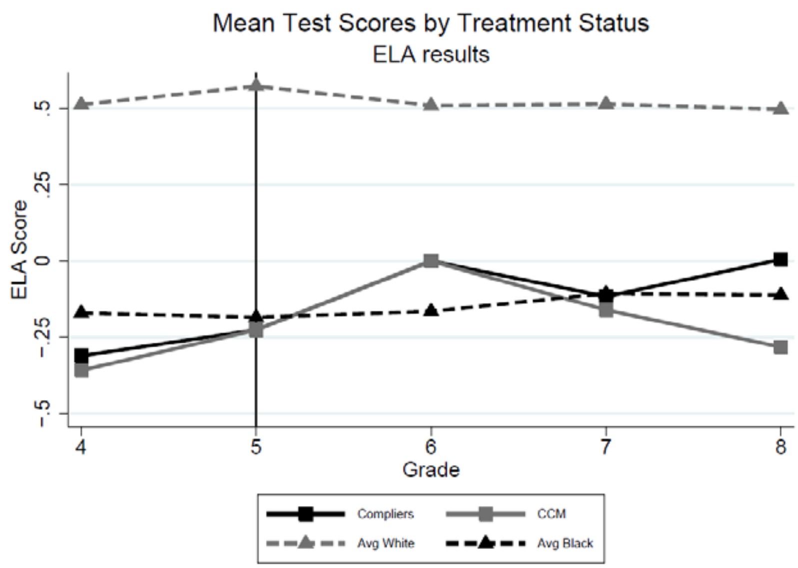




Notes: Lottery winners are students who receive a winning lottery number or who are in the top ten of the waitlist. Test scores are standardized by grade to have mean zero and standard deviation one in the entire New York City sample. The CCM is the estimated test score for those in the control group who would have complied if they had received a winning lottery number.

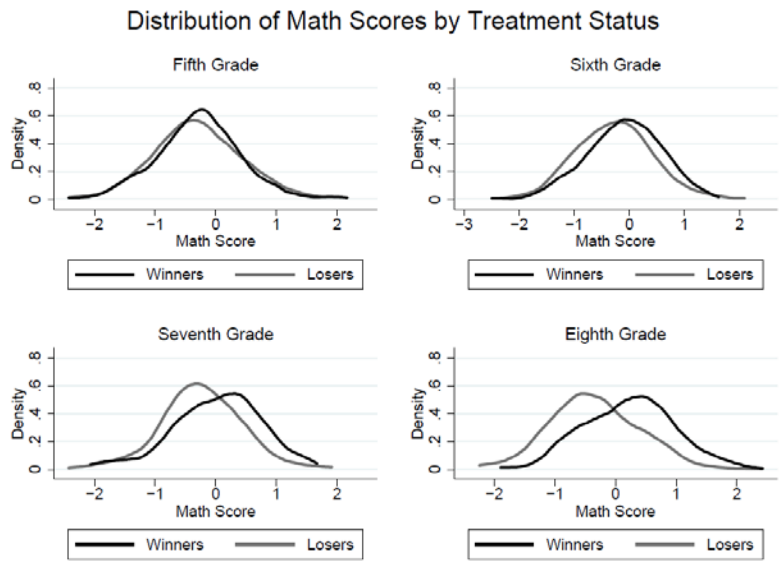
Figures 4A and 4B

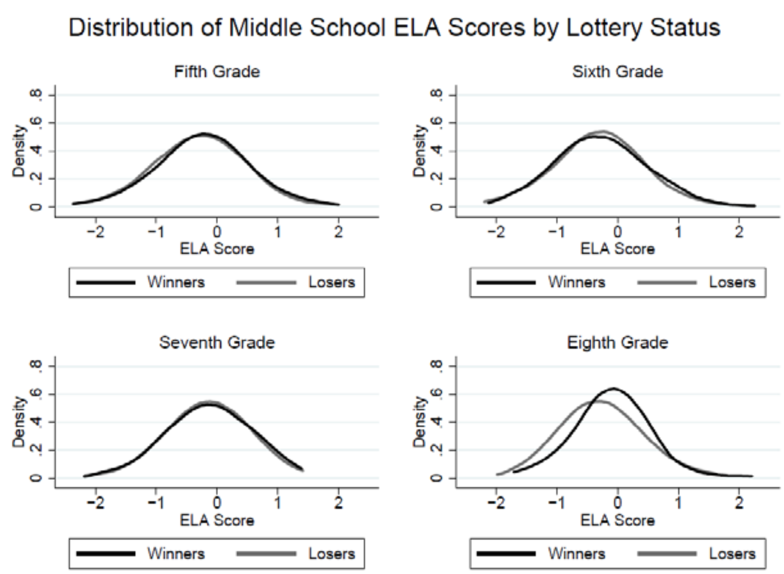




Notes: Lottery winners are students who receive a winning lottery number or who are in the top ten of the waitlist. Test scores are standardized by grade to have mean zero and standard deviation one in the entire New York City sample. The CCM is the estimated test score for those in the control group who would have complied if they had received a winning lottery number.

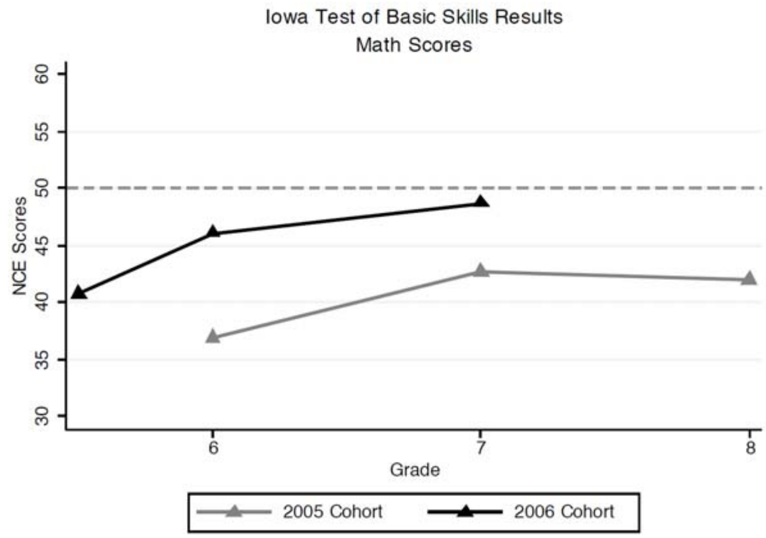
Figures 5A and 5B

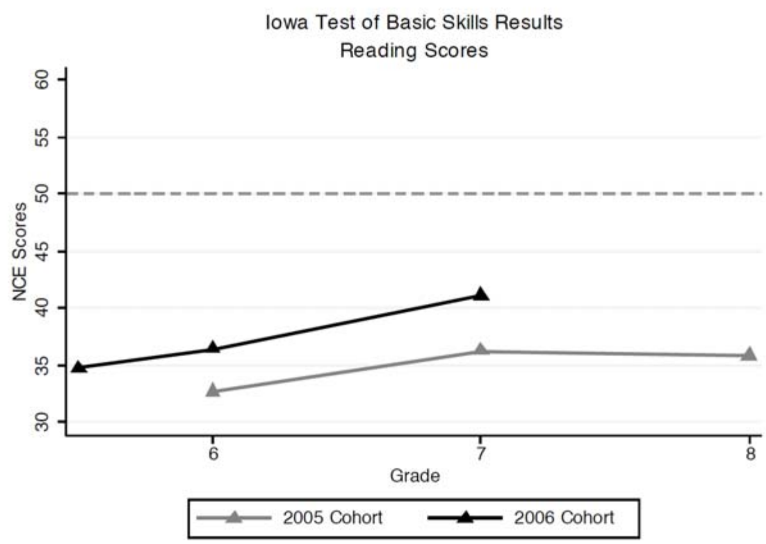




Notes: Lottery winners are students who receive a winning lottery number or who are in the top ten on the waitlist. Test scores are standardized by grade to have mean zero and standard deviation one in the entire New York City sample.

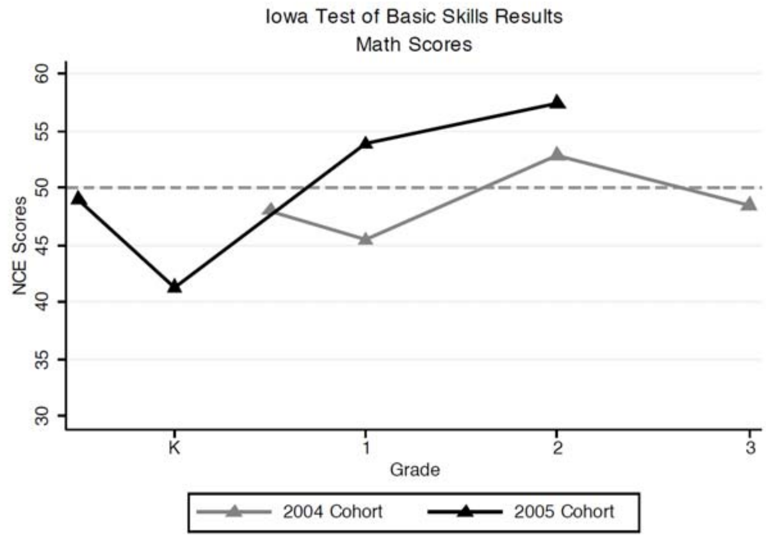
Figures 6A and 6B

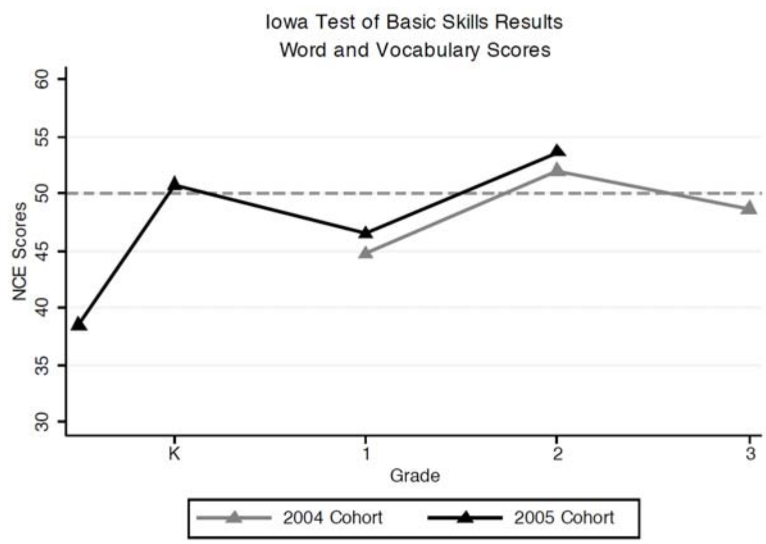




Notes: Sample is students who have non-missing test scores for all grades. We report National Percentile Rank scores, which are standardized to a national sample of students who are in the same grade and who were tested at the same time of year (fall or spring).

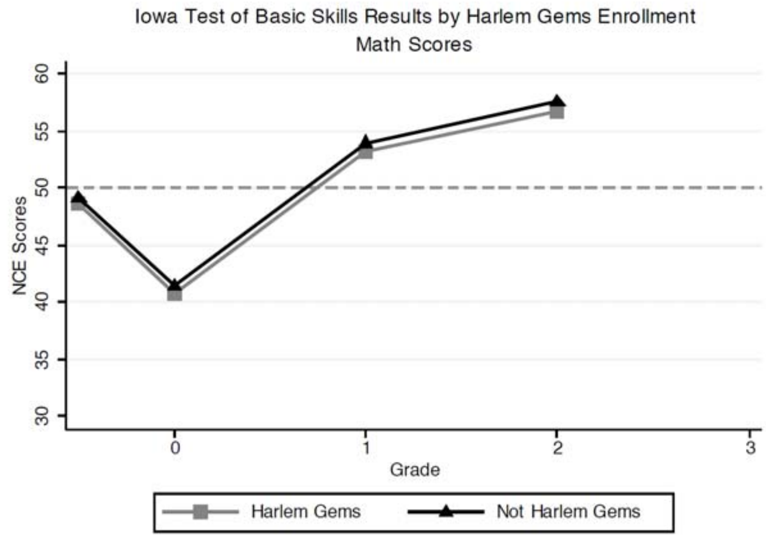
Figures 7A and 7B

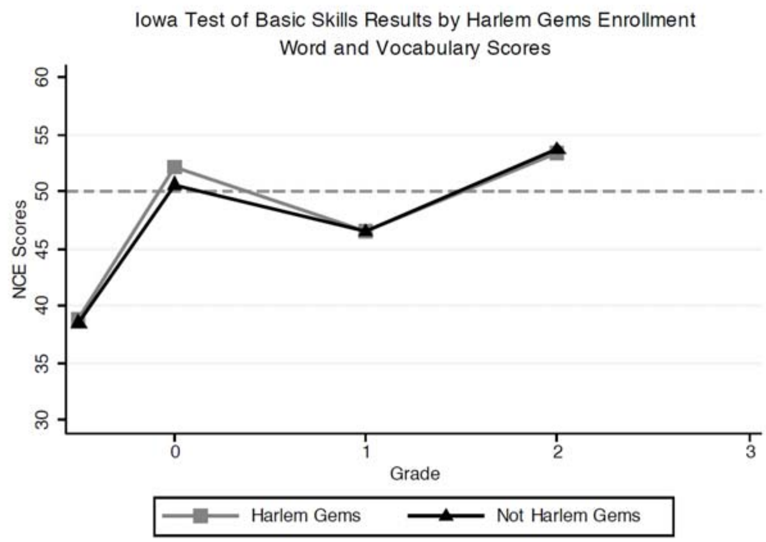




Notes: Sample is students who have non-missing test scores for all grades. We report National Percentile Rank scores, which are standardized to a national sample of students who are in the same grade and who were tested at the same time of year (fall or spring).

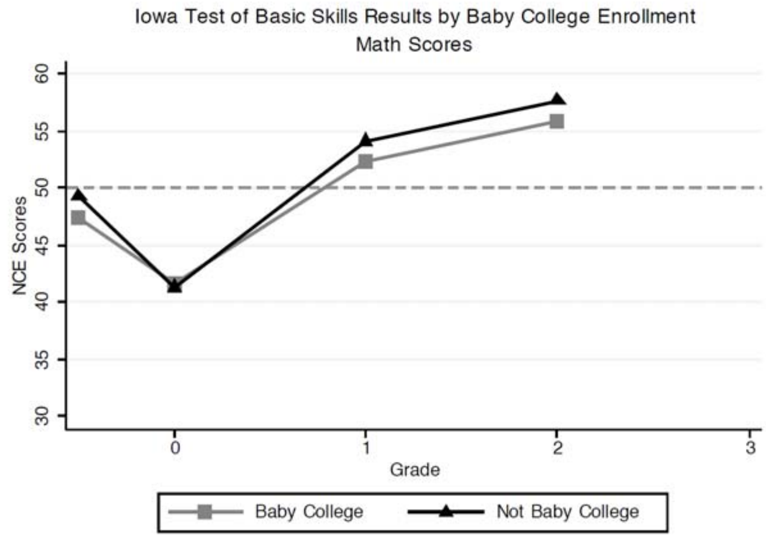
Figures 8A and 8B

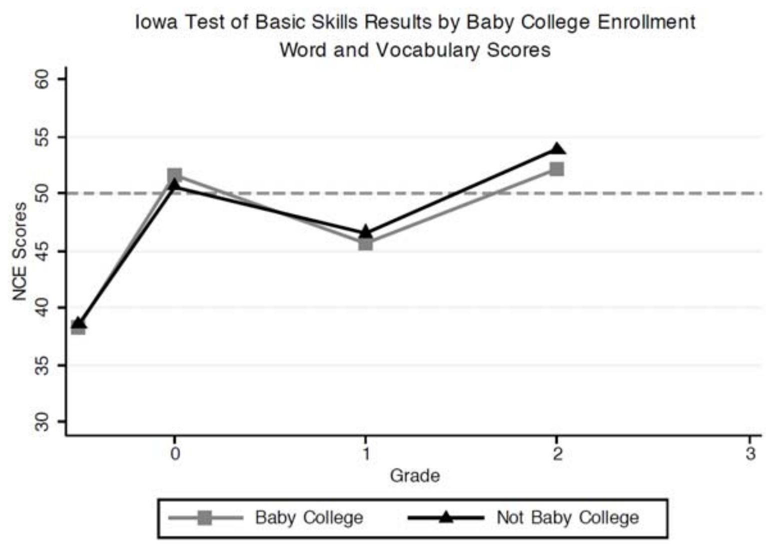




Notes: Sample is students who have non-missing test scores for all grades. We report National Percentile Rank scores, which are standardized to a national sample of students who are in the same grade and who were tested at the same time of year (fall or spring). Scores are adjusted by gender and lunch status. Results for 2004 cohort are excluded for simplicity, but are qualitatively similar and available on request.

Figures 9A and 9B





Notes: Sample is students who have non-missing test scores for all grades. We report National Percentile Rank scores, which are standardized to a national sample of students who are in the same grade and who were tested at the same time of year (fall or spring). Scores are adjusted by gender and lunch status. Results for 2004 cohort are excluded for simplicity, but are qualitatively similar and available on request.

Table 1
Summary Statistics of Lottery Sample

	Students in Lottery						Students living in designated areas			
	K – 4			6 – 8			K – 4		6 – 8	
Youth characteristics	Winners	Losers	p-value	Winners	Losers	p-value	HCZ	NYC	HCZ	NYC
Male	0.522	0.528	0.907	0.516	0.438	0.095	0.511	0.522	0.509	0.512
White	0.000	0.006	0.313	0.005	0.003	0.723	0.038	0.163	0.022	0.13
Black	0.861	0.813	0.215	0.846	0.833	0.700	0.719	0.287	0.741	0.308
Hispanic	0.139	0.170	0.411	0.143	0.157	0.672	0.207	0.390	0.222	0.381
Free Lunch	0.678	0.688	0.844	0.714	0.666	0.266	0.694	0.620	0.657	0.651
Reduced Lunch	0.078	0.102	0.421	0.115	0.127	0.705	0.086	0.071	0.096	0.095
Treatment Received	0.594	0.238	0.000	0.692	0.000	0.000	0.054	0.001	0.019	0.000
5th Grade Math	--	--	--	-0.246	-0.330	0.206	--	--	-0.425	0.004
5th Grade ELA	--	--	--	-0.211	-0.297	0.245	--	--	-0.306	-0.003
School characteristics										
Male	0.507	0.509		0.503	0.499		0.505	0.520	0.516	0.511
White	0.025	0.039		0.017	0.038		0.044	0.161	0.080	0.132
Black	0.745	0.669		0.764	0.511		0.640	0.290	0.535	0.306
Hispanic	0.194	0.245		0.188	0.404		0.272	0.390	0.334	0.380
Free Lunch	0.68	0.685		0.682	0.677		0.685	0.622	0.627	0.647
Reduced Lunch	0.108	0.107		0.114	0.091		0.083	0.073	0.094	0.094
5th Grade Math	--	--		-0.364	-0.257		--	--	-0.261	-0.009
5th Grade ELA	--	--		-0.301	-0.243		--	--	-0.190	-0.018
Charter	0.644	0.534		0.685	0.158		0.241	0.032	0.137	0.022
Observations	173	157		178	290		761	388923	516	233912

Notes: HCZ refers to the original 24-block area of the Harlem Children's Zone, ranging from 116th to 123rd Streets, 5th Avenue to 8th Avenue. NYC refers to the universe of New York City public-school students. Lottery winners are students who receive a winning lottery number or who are placed very high on the waitlist.

Table 2
Middle-School Lottery Results

	Pooled Sample		2005 Cohort		2006 Cohort	
Dependent Variable	ITT	TOT	ITT	TOT	ITT	TOT
6th Grade Math Score	0.249***	0.348***	0.215***	0.286***	0.280***	0.402***
	(0.053)	(0.073)	(0.069)	(0.091)	(0.085)	(0.120)
7th Grade Math Score	0.275***	0.394***	0.269***	0.372***	0.278***	0.406***
	(0.054)	(0.076)	(0.079)	(0.109)	(0.073)	(0.101)
8th Grade Math Score	0.546***	0.765***	0.682***	0.933***	0.403***	0.566***
	(0.076)	(0.100)	(0.103)	(0.135)	(0.103)	(0.140)
6th Grade Math on Level	0.116***	0.162***	0.096	0.128	0.141**	0.203**
	(0.043)	(0.059)	(0.059)	(0.079)	(0.061)	(0.088)
7th Grade Math on Level	0.178***	0.256***	0.211***	0.293***	0.140**	0.204**
	(0.043)	(0.062)	(0.062)	(0.085)	(0.060)	(0.087)
8th Grade Math on Level	0.275***	0.385***	0.369***	0.505***	0.183***	0.257***
	(0.043)	(0.056)	(0.059)	(0.078)	(0.065)	(0.088)
6th Grade ELA Score	0.033	0.047	0.033	0.044	0.022	0.032
	(0.053)	(0.075)	(0.074)	(0.099)	(0.079)	(0.114)
7th Grade ELA Score	0.029	0.041	-0.018	-0.024	0.045	0.066
	(0.050)	(0.072)	(0.071)	(0.097)	(0.074)	(0.110)
8th Grade ELA Score	0.190***	0.266***	0.210***	0.288***	0.138*	0.192*
	(0.058)	(0.081)	(0.080)	(0.109)	(0.079)	(0.111)
6th Grade ELA on Level	0.029	0.040	0.057	0.077	-0.015	-0.022
	(0.039)	(0.055)	(0.052)	(0.071)	(0.058)	(0.084)
7th Grade ELA on Level	-0.019	-0.027	-0.024	-0.033	-0.042	-0.062
	(0.042)	(0.060)	(0.055)	(0.076)	(0.061)	(0.091)
8th Grade ELA on Level	0.151***	0.211***	0.121**	0.166**	0.174***	0.243***
	(0.042)	(0.060)	(0.056)	(0.077)	(0.064)	(0.091)
6th Grade Absences	-2.984***	-4.187***	-1.628*	-2.176*	-4.383***	-6.400***
	(0.704)	(0.985)	(0.976)	(1.297)	(0.990)	(1.418)
7th Grade Absences	-2.202*	-3.192*	-4.067***	-5.618***	-0.758	-1.136
	(1.290)	(1.848)	(1.507)	(2.078)	(1.919)	(2.864)
8th Grade Absences	-3.960***	-5.563***	-7.629***	-10.431***	0.234	0.333
	(1.323)	(1.862)	(2.100)	(2.889)	(2.080)	(2.954)
On-Time at 6th Grade	--	--	--	--	--	--
On-Time at 7th Grade	-0.004	-0.006	-0.027	-0.037	0.016	0.023
	(0.023)	(0.034)	(0.040)	(0.055)	(0.025)	(0.037)
On-Time at 8th Grade	0.006	0.009	0.010	0.014	-0.010	-0.014
	(0.023)	(0.032)	(0.043)	(0.059)	(0.014)	(0.020)
Observations: Treated	178	178	83	83	95	95
Control	289	289	167	167	122	122

Notes: Each regression controls for the gender, race, lunch status, and predetermined values of the dependent variable. Standard errors are reported in parentheses. We cluster the standard errors at the individual level for the pooled regression as they inlcude students that applied in both 2005 and 2006. Regressions for the individual year samples are not clustered. The regression is run for all students entered in the indicated HCZ lottery who have valid scores through 2008 - 2009. Treatment is defined as either having a winning lottery number or being high on the waitlist. Test scores are standardized to have mean zero and standard deviation one by grade in the full New York City sample. *** = significant at 1% level, ** = significant at 5% level, * = significant at 10% level.

Table 3
Middle-School IV Results

Dependent Variable	800 Meters	1600 Meters	2400 Meters
6^{th} Grade Math Score	0.217	0.066	0.141
	(0.568)	(0.562)	(0.501)
7^{th} Grade Math Score	1.208^{**}	1.106^{**}	0.902^{**}
	(0.405)	(0.347)	(0.299)
8^{th} Grade Math Score	1.248^{***}	1.298^{***}	1.130^{***}
	(0.236)	(0.191)	(0.153)
6^{th} Grade Math on Level	-0.129	-0.192	-0.161
	(0.123)	(0.147)	(0.128)
7^{th} Grade Math on Level	-0.041	-0.120	-0.179
	(0.331)	(0.272)	(0.223)
8^{th} Grade Math on Level	0.636^{***}	0.629^{***}	0.512^{***}
	(0.127)	(0.130)	(0.124)
6^{th} Grade ELA Score	-0.010	0.015	0.052
	(0.402)	(0.354)	(0.311)
7^{th} Grade ELA Score	0.209	0.249	0.318
	(0.461)	(0.408)	(0.385)
8^{th} Grade ELA Score	0.136	0.273	0.260
	(0.303)	(0.257)	(0.254)
6^{th} Grade ELA on Level	0.026	-0.012	0.018
	(0.240)	(0.233)	(0.238)
7^{th} Grade ELA on Level	-0.297	-0.307	-0.290
	(0.213)	(0.185)	(0.181)
8^{th} Grade ELA on Level	0.191	0.161	0.175
	(0.217)	(0.201)	(0.201)
6^{th} Grade Absences	-5.840^{*}	-5.800^{*}	-5.514
	(2.322)	(2.494)	(2.996)
7^{th} Grade Absences	-11.434^{*}	-10.420^{*}	-10.100
	(4.723)	(5.088)	(5.574)
8^{th} Grade Absences	-4.550	-4.393	-2.217
	(3.918)	(3.182)	(2.534)
On-Time at 6^{th} Grade	--	--	--
On-Time at 7^{th} Grade	0.069	0.013	0.001
	(0.140)	(0.127)	(0.126)
On-Time at 8^{th} Grade	-0.115	-0.143	-0.100
Observations	9155	19871	29228

Notes: Each regression controls for the gender, race, lunch status, predetermined values of the dependent variable (when available), and whether the student lives within the 24-block Children's Zone. Enrollment in the HCZ charter school at sixth grade is the reported coefficient, which is instrumented for using interactions between the student's cohort and whether she lives within the 24-block Children's Zone. Standard errors corrected for clustering at the cohort level are reported in parentheses. The sample is all students who live within the indicated distance of the Zone. *** = significant at 1% level, ** = significant at 5% level, * = significant at 10% level.

Table 4
Middle-School Lottery Results by Subsample

Dependent Variable	Pooled	Boys	Girls	p-value
6th Grade Math Score	0.348***	0.333***	0.360***	0.854
	(0.073)	(0.100)	(0.105)	
7th Grade Math Score	0.394***	0.336***	0.446***	0.467
	(0.076)	(0.111)	(0.103)	
8th Grade Math Score	0.765***	0.865***	0.672***	0.314
	(0.100)	(0.136)	(0.141)	
6th Grade ELA Score	0.047	-0.008	0.097	0.485
	(0.075)	(0.108)	(0.104)	
7th Grade ELA Score	0.041	-0.004	0.081	0.567
	(0.072)	(0.113)	(0.094)	
8th Grade ELA Score	0.266***	0.208**	0.320**	0.493
	(0.081)	(0.096)	(0.130)	
Observations	486	228	258	

Dependent Variable	Pooled	Free lunch	Not on free lunch	p-value
6th Grade Math Score	0.374***	0.333***	0.381***	0.754
	(0.078)	(0.088)	(0.128)	
7th Grade Math Score	0.423***	0.441***	0.287**	0.333
	(0.081)	(0.091)	(0.133)	
8th Grade Math Score	0.827***	0.793***	0.704***	0.646
	(0.108)	(0.126)	(0.153)	
6th Grade ELA Score	0.050	0.048	0.045	0.984
	(0.081)	(0.093)	(0.126)	
7th Grade ELA Score	0.044	-0.028	0.206	0.163
	(0.076)	(0.083)	(0.145)	
8th Grade ELA Score	0.287***	0.254***	0.293*	0.837
	(0.087)	(0.091)	(0.165)	
Observations	486	375	111	

Dependent Variable	Pooled	Close to HCZ	Not close to HCZ	p-value
6th Grade Math Score	0.374***	0.264**	0.385***	0.303
	(0.078)	(0.109)	(0.080)	
7th Grade Math Score	0.423***	0.342***	0.418***	0.549
	(0.081)	(0.114)	(0.087)	
8th Grade Math Score	0.827***	0.782***	0.758***	0.872
	(0.108)	(0.139)	(0.114)	
6th Grade ELA Score	0.050	0.163	-0.005	0.193
	(0.081)	(0.122)	(0.081)	
7th Grade ELA Score	0.044	0.159	-0.013	0.114
	(0.076)	(0.098)	(0.082)	
8th Grade ELA Score	0.287***	0.214**	0.289***	0.566
	(0.087)	(0.104)	(0.099)	
Observations	486	154	333	

Notes: Each regression controls for the gender, race, lunch status, and predetermined values of the dependent variable. Standard errors corrected for clustering at the student level are reported in parentheses. The regression is run for all students entered in the indicated HCZ lottery who have valid scores through 2008 - 2009. The reported coefficient is the effect of being enrolled in the HCZ Promise Academy in sixth grade interacted with the defining subsample variable, instrumented for using treatment status interacted with the defining subsample variable. Scores are standardized to have mean zero and standard deviation one by grade in the full New York City sample. *** = significant at 1% level, ** = significant at 5% level, * = significant at 10% level.

Table 5
Elementary-School IV Results

Dependent Variable	800 Meters	1600 Meters	2400 Meters
3rd Grade Math Score	1.117	1.296**	1.245**
	(0.581)	(0.417)	(0.405)
3rd Grade Math on Level	0.556	0.631	0.535
	(0.437)	(0.406)	(0.375)
3rd Grade ELA Score	1.389*	1.524**	1.570**
	(0.601)	(0.489)	(0.521)
3rd Grade ELA on Level	0.815*	0.823**	0.701*
	(0.371)	(0.293)	(0.311)
4th Grade Math Score	0.446	0.669**	0.613**
	(0.242)	(0.197)	(0.224)
4th Grade Math on Level	0.708**	0.741***	0.545**
	(0.191)	(0.122)	(0.135)
4th Grade ELA Score	1.052*	1.479**	1.416**
	(0.418)	(0.376)	(0.366)
4th Grade ELA on Level	0.721**	1.131***	1.069***
	(0.249)	(0.196)	(0.201)
1st Grade Absences	-4.236	1.318	2.115
	(9.324)	(7.741)	(8.878)
2nd Grade Absences	-0.883	-4.633	-7.804
	(5.992)	(4.923)	(4.432)
3rd Grade Absences	4.583	-3.036	-4.537
	(7.232)	(4.679)	(4.120)
4th Grade Absences	5.797	5.379*	3.519
	(4.180)	(2.660)	(2.834)
On-time at 1st Grade	-0.425**	-0.411***	-0.386***
	(0.137)	(0.100)	(0.082)
On-time at 2nd Grade	-0.518	-0.433*	-0.440*
	(0.283)	(0.194)	(0.208)
On-time at 3rd Grade	-0.788**	-0.610**	-0.576*
	(0.226)	(0.197)	(0.241)
On-time at 4th Grade	-0.331***	-0.206***	-0.210***
	(0.038)	(0.014)	(0.029)
Observations	6156	13540	20138

Notes: Each regression controls for the gender, race, lunch status, predetermined values of the dependent variable (when available), and whether the student lives within the 24-block Children's Zone. Enrollment in the HCZ charter school at kindergarten is the reported coefficient, which is instrumented for using the interaction between the student's cohort and whether she lives within the 24-block Children's Zone. Standard errors corrected for clustering at the cohort level are reported in parentheses. The sample is all students living within the indicated distance of the Zone. *** = significant at 1% level, ** = significant at 5% level, * = significant at 10% level.

Table 6
Elementary-School Lottery Results

Dependent Variable	Pooled	2004 Cohort	2005 Cohort
3rd Grade Math Score	0.644	0.823**	0.470
	(0.436)	(0.375)	(0.814)
3rd Grade Math on Level	0.152	0.148	0.147
	(0.109)	(0.110)	(0.191)
3rd Grade ELA Score	0.994**	0.844**	1.254
	(0.410)	(0.349)	(0.778)
3rd Grade ELA on Level	0.453**	0.304	0.631*
	(0.208)	(0.237)	(0.350)
4th Grade Math Score		1.032*	
		(0.569)	
4th Grade Math on Level		0.437*	
		(0.244)	
4th Grade ELA Score		0.476	
		(0.514)	
4th Grade ELA on Level		0.099	
		(0.289)	
1st Grade Absences	-8.131*	-9.728	-7.168
	(4.617)	(5.994)	(6.898)
2nd Grade Absences	-10.002**	-13.664***	-6.105
	(4.334)	(4.550)	(6.883)
3rd Grade Absences	-6.915	-7.505	-5.237
	(4.486)	(4.559)	(7.663)
4th Grade Absences		-4.539	
		(9.444)	
On-time at 1st Grade	0.085	0.063	0.081
	(0.067)	(0.070)	(0.110)
On-time at 2nd Grade	0.229*	0.106	0.332
	(0.128)	(0.099)	(0.244)
On-time at 3rd Grade	0.211*	0.326*	0.075
	(0.113)	(0.173)	(0.118)
On-time at 4th Grade		0.326*	
		(0.173)	
Observations	323	107	216

Notes: Each regression controls for the gender, race, and lunch status. Enrollment in the HCZ charter school at kindergarten is the reported coefficient, which is instrumented for using the student's position in the lottery and waitlist. Standard errors corrected for clustering at the cohort level are reported in parentheses. The sample is all students entered in the HCZ lottery in the indicated year. Test scores are standardized to have mean zero and standard deviation one by grade in the full New York City sample. *** = significant at 1% level, ** = significant at 5% level, * = significant at 10% level.

Table 7
Harlem Gems and Baby College Results

Panel A - All Students in NYC

	Math Score in 3rd Grade			ELA Score in 3rd Grade			Total Absences in 3rd Grade			On-Time at 3rd Grade		
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Promise Academy	0.404***	0.411***	0.415***	0.476***	0.478***	0.480***	-2.954***	-2.845***	-2.911***	0.109***	0.109***	0.107***
	(0.090)	(0.091)	(0.091)	(0.089)	(0.090)	(0.090)	(1.003)	(1.017)	(1.019)	(0.033)	(0.033)	(0.034)
Harlem Gems		-0.066	-0.047		-0.019	-0.008		-0.993	-1.336		0.006	-0.001
		(0.133)	(0.137)		(0.132)	(0.135)		(1.487)	(1.523)		(0.049)	(0.050)
Baby College			-0.057			-0.033			1.039			0.021
			(0.091)			(0.091)			(1.000)			(0.033)
Observations	139585	139585	139585	136835	136835	136835	142793	142793	142793	143103	143103	143103

Panel B - Students Living Close to or in the Harlem Children Zone

	Math Score in 3rd Grade			ELA Score in 3rd Grade			Total Absences in 3rd Grade			On-Time at 3rd Grade		
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Promise Academy	0.621***	0.611***	0.635***	0.636***	0.654***	0.672***	-2.598	-2.348	-2.501	0.136**	0.144**	0.148**
	(0.156)	(0.161)	(0.162)	(0.147)	(0.152)	(0.152)	(1.983)	(2.046)	(2.057)	(0.068)	(0.070)	(0.070)
Harlem Gems		0.044	0.116		-0.080	-0.025		-1.135	-1.588		-0.037	-0.025
		(0.179)	(0.185)		(0.168)	(0.174)		(2.265)	(2.350)		(0.077)	(0.080)
Baby College			-0.223			-0.172			1.413			-0.039
			(0.156)			(0.148)			(1.950)			(0.066)
Observations	2809	2809	2809	2749	2749	2749	2883	2883	2883	2894	2894	2894

Panel C - Kindergarten Lottery Sample

	Math Score in 3rd Grade			ELA Score in 3rd Grade			Total Absences in 3rd Grade			On-Time at 3rd Grade		
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Promise Academy	0.486***	0.492***	0.494***	0.576***	0.575***	0.579***	-2.313**	-2.198*	-2.257**	-0.008	-0.010	-0.011
	(0.114)	(0.115)	(0.115)	(0.105)	(0.105)	(0.106)	(1.130)	(1.135)	(1.136)	(0.027)	(0.027)	(0.027)
Harlem Gems		-0.081	-0.066		0.003	0.036		-1.474	-2.043		0.026	0.014
		(0.140)	(0.150)		(0.129)	(0.138)		(1.389)	(1.479)		(0.033)	(0.035)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Baby College			-0.049			-0.108			1.856			0.039
			(0.168)			(0.155)			(1.663)			(0.040)
Observations	322	322	322	322	322	322	322	322	322	322	322	322

Notes: Each regression controls for the gender, race, and lunch status of the student. Columns (1) – (3) use 3rd grade math score as the dependent variable. Columns (4) – (6) use ELA score. Columns (7) – (9) use total absences in 3rd grade, and columns (10) – (12) use a student's on-time status at 3rd grade. Test scores are standardized to have mean zero and standard deviation one by grade in the full New York City sample. *** = significant at 1% level, ** = significant at 5% level, * = significant at 10% level.

Table 8
Baby College Results in the Harlem Gems Sample

Variable	(1)	(2)	(3)	(5)	(6)
Baby College	0.045	0.028	0.034	0.005	0.009
	(0.111)	(0.110)	(0.111)	(0.111)	(0.111)
Controls for					
Gender	No	Yes	Yes	Yes	Yes
Race	No	Yes	Yes	Yes	Yes
Age	No	No	Yes	Yes	Yes
Parent's Education	No	No	No	Yes	Yes
Parent's Income	No	No	No	No	Yes
Observations	451	451	451	451	451

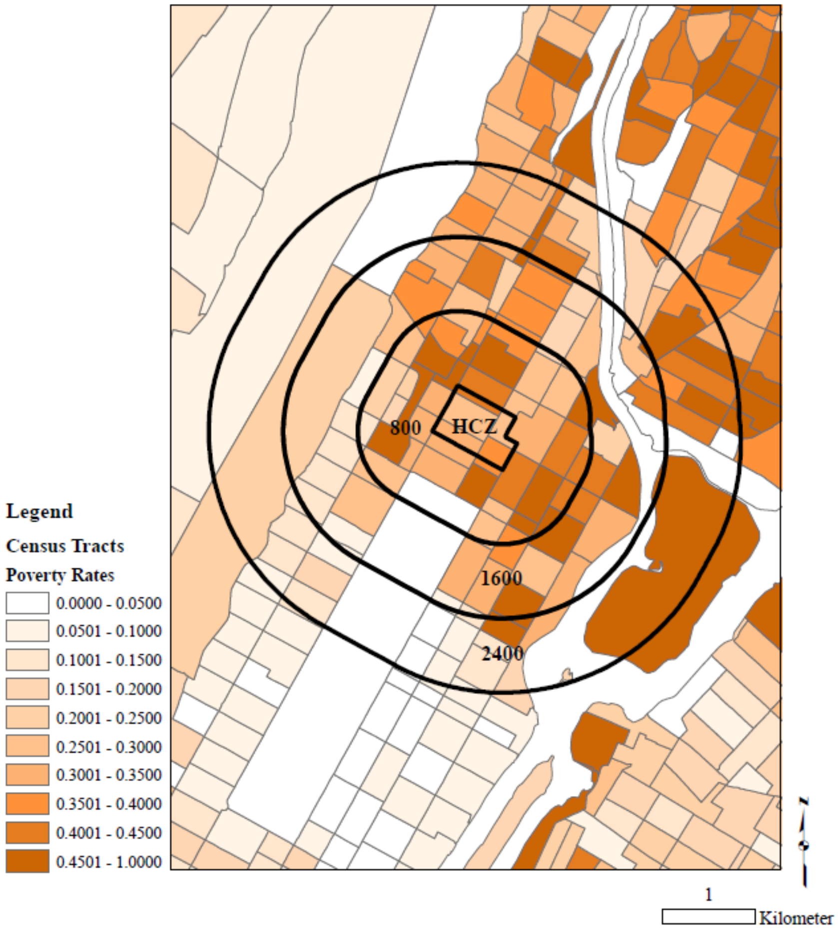
Notes: The sample is made up of all students ever enrolled in the original Harlem Gems program, the Uptown Harlem Gems Program, or the Head Start Gems Program, between 2001-02 and 2007-08. Note that after the 2005-06 school year, the original Harlem Gems program became part of the HCZ Promise Academies. The dependent variable for all regressions is the Bracken scale test score, which is administered before the Harlem Gems begins. Scores are standardized to have mean zero and standard deviation one. The omitted race category is non-black, which make up less than 15 percent of the sample and is comprised mostly of Hispanics and African students. The omitted parent income is less than $10,000 per year. The omitted education category is less than a high school diploma. *** = significant at 1% level, ** = significant at 5% level, * = significant at 10% level.

Table 9
Siblings Results in the Middle-School Lottery Sample

Dependent Variable	
Math Score#	0.053
	(0.337)
ELA Score#	-0.125
	(0.275)
Absences	-7.961*
	(5.062)
On-time	-0.093
	(0.075)
Observations	280

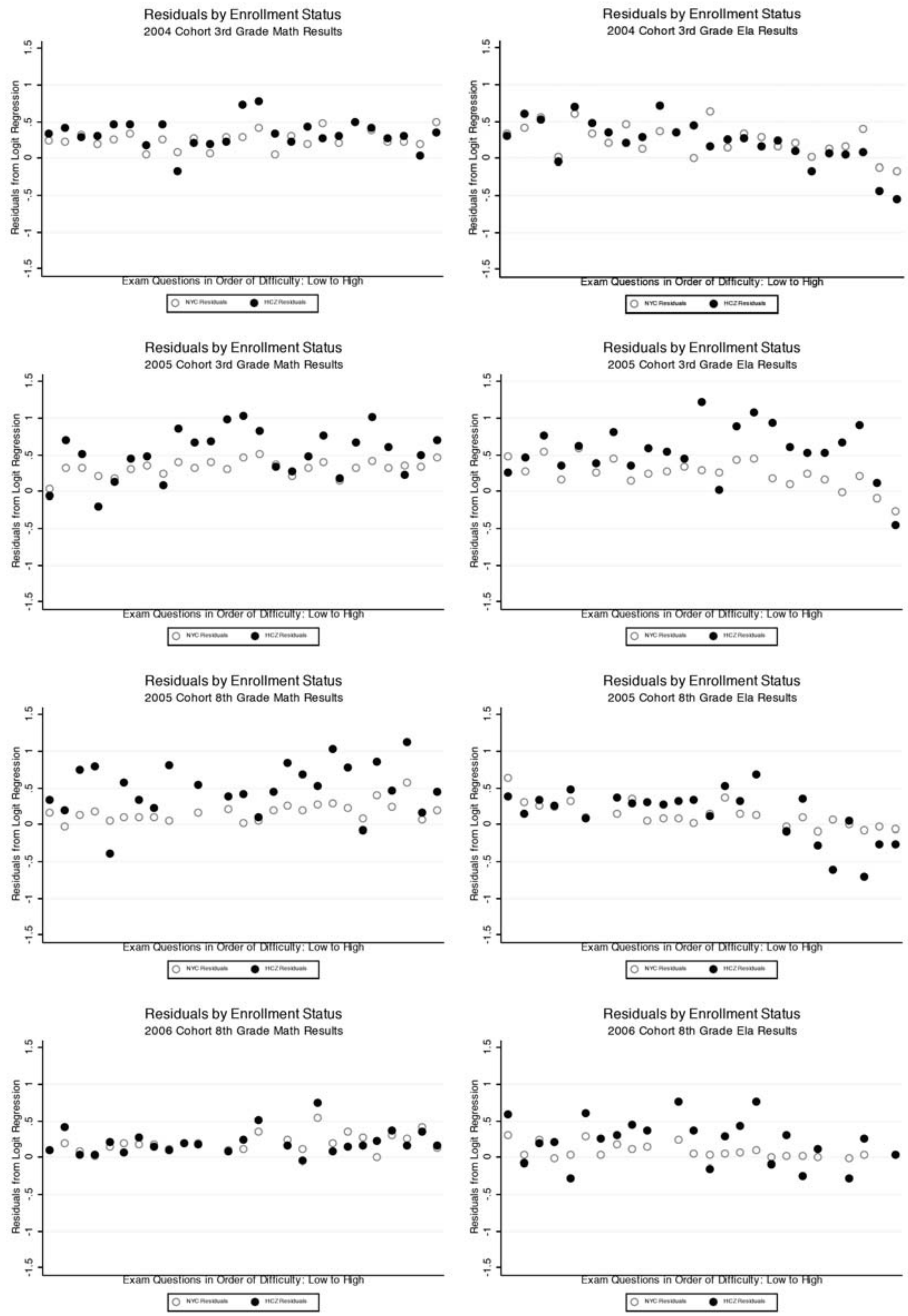
Notes: Each regression controls for sibling gender, race, lunch status, predetermined values of the dependent variable. The reported coefficient is an indicator for whether lottery student was enrolled at the HCZ Promise Academy Middle School in sixth grade, instrumented for using the lottery student's lottery status. Standard errors corrected for clustering at the family level are reported in parentheses. Sample is restricted to those siblings that did not attend the HCZ Promise Academy themselves. Results are qualitatively similar if all siblings are used. Test scores are standardized to have mean zero and standard deviation one by grade in the full New York City sample *** = significant at 1% level, ** = significant at 5% level, * = significant at 10% level. #Test score data are only available for grades 3 – 8, limiting the sample size to 139.

Appendix Figure 1
Poverty Rates in and around the Harlem Children's Zone



Notes: HCZ refers to the original 24-block Children's Zone. Circles correspond to the distance in meters from the nearest border of the original Children's Zone. Poverty rates by block group are from the 2000 Public Use Census data.

Appendix Figures 2A through 2H
Residuals from Logit Regressions on Exam Question Responses



Notes: Each regression controls for gender, race, lunch status, and test score. The dependent variable is whether a student answered a specific exam question correctly. We estimate this model for each question and separately by grade, subject and year. We then construct the unexplained residual for each student on each item, and sum the residuals for each item. These summed residuals are plotted above, with the questions ordered from easiest (most students answering correctly) to hardest (fewest students answering correctly). *** = significant at 1% level, ** = significant at 5% level, * = significant at 10% level.

Appendix Table 1
Middle School IV First Stage Results

Instruments	Elementary School	Middle School
2001 Cohort x Zone Interaction	0.0006**	--
	(0.000)	
2002 Cohort x Zone Interaction	0.0009**	-0.0004
	(0.000)	(0.000)
2003 Cohort x Zone Interaction	0.0009**	0.0042***
	(0.000)	(0.000)
2004 Cohort x Zone Interaction	0.0566***	0.1225***
	(0.000)	(0.000)
2005 Cohort x Zone Interaction	0.0226***	0.0511***
	(0.001)	(0.000)
2006 Cohort x Zone Interaction	0.0141***	0.0282***
	(0.000)	(0.000)
2007 Cohort x Zone Interaction	0.0276***	-0.0004
	(0.000)	(0.000)
2008 Cohort x Zone Interaction	0.0339***	-0.0003
	(0.001)	(0.000)
Select Control Variables		
Living in the Zone	-0.0065**	-0.0022*
	(0.002)	(0.001)
2001 Cohort Dummy	-0.0005**	--
	(0.000)	
2002 Cohort Dummy	-0.0008**	0.0000**
	(0.000)	0.000
2003 Cohort Dummy	-0.0001	0.0003***
	(0.000)	0.000
2004 Cohort Dummy	0.0117***	0.0131***
	(0.000)	0.000
2005 Cohort Dummy	0.0135***	0.0130***
	(0.000)	(0.000)
2006 Cohort Dummy	0.0166***	0.0077***
	(0.000)	(0.000)
2007 Cohort Dummy	0.0121***	0.0002
	(0.000)	(0.000)
Constant	-0.0023	-0.0038**
	(0.002)	(0.001)
Observations	35218	35790
F-test on instruments (p-value)	0.000	0.000

Notes: Each regression controls also for the gender, race, and lunch status of the student. Both samples include all students in the relevant cohorts that live within 1600 meters of the Zone. The dependent variable is enrollment at the HCZ charter school. Standard errors corrected for clustering at the cohort level are in parentheses. Results are comparable for other distances, and if other control variables are included. Standard errors are clustered at the cohort level. *** = significant at 1% level, ** = significant at 5% level, * = significant at 10% level.

Appendix Table 2
Sensitivity of Middle-School Lottery Results

	Pooled Sample		2005 Cohort		2006 Cohort	
Dependent Variable	ITT	TOT	ITT	TOT	ITT	TOT
6th Grade Math Score	0.217***	0.339***	0.178**	0.274**	0.280***	0.419***
	(0.059)	(0.086)	(0.071)	(0.106)	(0.085)	(0.125)
7th Grade Math Score	0.227***	0.346***	0.199**	0.305**	0.280***	0.414***
	(0.071)	(0.108)	(0.084)	(0.128)	(0.073)	(0.103)
8th Grade Math Score	0.432***	0.683***	0.543***	0.893***	0.350***	0.505***
	(0.116)	(0.156)	(0.120)	(0.168)	(0.107)	(0.150)
6th Grade ELA Score	-0.008	-0.013	-0.027	-0.044	0.022	0.033
	(0.064)	(0.101)	(0.079)	(0.126)	(0.079)	(0.118)
7th Grade ELA Score	0.014	0.021	-0.04	-0.06	0.043	0.065
	(0.051)	(0.077)	(0.069)	(0.104)	(0.075)	(0.112)
8th Grade ELA Score	0.193***	0.304***	0.194**	0.321**	0.157**	0.226**
	(0.060)	(0.093)	(0.078)	(0.128)	(0.078)	(0.113)
Observations: Treated	177	177	82	82	95	95
Control	306	306	173	173	133	133

Notes: This table explores the sensitivity of our results to the difference in match rates between the Middle School lottery winners and losers by creating pseudo-lottery losers to equalize the match rates for each cohort. These pseudo-students are assigned the highest test score observed in the control group for that year (about two standard deviations above the mean). The number of observations refers to the number of real, not pseudo students. Each regression controls for the gender, race, lunch status, and predetermined values of the dependent variable. Treatment is defined as either having a winning lottery number or being high on the waitlist. Standard errors corrected for clustering at the student level are in parentheses. Test scores are standardized to have mean zero and standard deviation one by grade in the full New York City sample. *** = significant at 1% level, ** = significant at 5% level, * = significant at 10% level.

Appendix Table 3
Summary Statistics of IV Sample

	Elementary School Sample					Middle School Sample				
	In the Zone		Out of the Zone			In the Zone		Out of the Zone		
Youth characteristics	Eligible	Ineligible	Eligible	Ineligible	Difference	Eligible	Ineligible	Eligible	Ineligible	Difference
Male	0.516	0.522	0.519	0.526	0.028 (0.016)	0.513	0.496	0.517	0.521	0.021 (0.017)
White	0.031	0.032	0.161	0.171	-0.018*** (0.005)	0.016	0.021	0.150	0.149	-0.006 (0.006)
Black	0.696	0.688	0.304	0.309	-0.032* (0.017)	0.754	0.750	0.310	0.327	0.020 (0.016)
Hispanic	0.236	0.241	0.377	0.371	0.029 (0.016)	0.216	0.199	0.372	0.382	0.027 (0.032)
Free Lunch	0.609	0.595	0.625	0.535	0.026 (0.030)	0.627	0.634	0.577	0.584	0.000 (0.022)
Reduced Lunch	0.081	0.079	0.090	0.070	0.013 (0.011)	0.132	0.091	0.089	0.084	0.036** (0.013)
5th Grade Math	--	--	--	--	--	-0.381	-0.465	0.025	-0.045	0.014 (0.096)
5th Grade ELA	--	--	--	--	--	-0.316	-0.289	0.020	-0.041	-0.088 (0.057)
Observations	4649	4649	379555	1228051	--	491	916	302510	501962	--

Notes: In the Zone refers to the original 24-block area of the Harlem Children's Zone, ranging from 116th to 123rd Streets, 5th Avenue to 8th Avenue. Out of the Zone refers to all addresses within 1600 meters of the original Zone. Columns 5 and 10 have p-values from a test of equality between the difference of eligible and ineligible cohorts in and out of the Zone: ($Eligible_{in}$-$Ineligible_{in}$= $Eligible_{out}$-$Ineligible_{out}$).

Appendix A: Complete List of Harlem Children's Zone Programs

COMMUNITY INVESTMENTS

Early Childhood Programs

The Baby College offers nine-week parenting workshops to expectant parents and those raising a child up to three years old.

The Three Year Old Journey works with parents of children who have won the HCZ Promise Academy charter school lottery. Held on Saturdays over several months, it teaches parents about their child's development, building language skills and parenting skills.

Get Ready for Pre-K involves children in small and large group activities that are designed to increase socialization skills, build routines, and provide exposure to their first classroom experience. The program has a particular focus on the development of pre-literacy skills. In order to prepare children for September entry into Harlem Gems Universal Pre-K or Head Start, four-year-old children attend Get Ready for Pre-K from 8 AM to 6 PM every day for six weeks during the preceding summer.

Harlem Gems Universal Pre-K is an all-day pre-kindergarten program that gets children ready to enter kindergarten. Classes have a 4:1 child-to-adult ratio, teach English, Spanish and French, and run from 8 a.m. to 6 p.m. HCZ runs three pre-kindergarten sites, serving over 250 children.

The *Harlem Gems Head Start* follows the same model as the Universal Pre-K but has a few important differences: 1) children can enter the Head Start at three or four years of age, thus a sizable proportion of the participants receive two years of instruction in the program, (2) because of income guidelines, students in Head Start tend to come from families with lower socioeconomic status than that of Universal Pre-K participants, and (3) while lead teachers at the UPK have master's degrees, Head Start teachers have bachelor's degrees.

Public Elementary School Programs

Harlem Peacemakers funded in part by AmeriCorps, trains young people who are committed to making their neighborhoods safe for children and families. The agency has Peacemakers working as teaching assistants in seven public schools and the HCZ Promise Academy Charter School.

Public Middle School Programs

The *TRUCE Fitness and Nutrition Center* offers free classes to children in karate, fitness and dance. Participants also learn about health and nutrition, as well as receiving regular academic assistance. The program is focused on developing middle-school youth, grades 5-8.

A Cut Above is an after-school program that helps students in the critical-but-difficult middle-school years. Supporting students who are not in the HCZ Promise Academy charter school, it provides academic help, leadership development, as well as high-school and college preparation.

Public High School Programs

TRUCE Arts & Media (The Renaissance University for Community Education) does youth development through the arts and media, working with youth in grades 9-12 on academic growth, career readiness as well as fostering media literacy and artistic ability.

The *Employment and Technology* center teaches computer and job-related skills to teens and adults.

Learn to Earn is an after-school program that helps high-school juniors and seniors improve their academic skills, as well as preparing them for college and the job market.

College Programs

The *College Success Office* supports students who have graduated from high school and HCZ programs. It helps them get into the most appropriate college, then assists them throughout their college years.

Family, Community and Health Programs

Community Pride organizes tenant and block associations, helping many hundreds of tenants convert their city-owned buildings into tenant-owned co-ops.

Single Stop offers access to a wide variety of services - from counseling to financial advice to legal consultations - at several locations each week.

The *HCZ Asthma Initiative* works closely with asthmatic children and their families so they can learn to manage the disease and lessen its effects.

The *Healthy Living Initiative* (formerly the Obesity Initiative) is a multi-pronged program to help children and their families reverse the alarming trend toward obesity and its health effects.

The Beacon Center Program

The *Beacon programs* turn school buildings into community centers, offering programs during the afternoon, evening and weekend. They offer programs for youth and adults from education to the arts to recreation. Each summer, they offer all-day camp so children have a safe, enriching place to spend their time instead of hanging out on the street.

Foster Care Prevention Services

The HCZ Foster Care Prevention programs work to stabilize and strengthen families so that their children are not placed in foster care. They include:

The *Family Development Program*, which serves 120 families and specializes in access to mental-health professionals who collaborate with caseworkers to support therapeutic interventions.

The *Family Support Center*, which serves 90 families, and specializes in providing crisis-intervention services, referrals, advocacy, as well as groups on parenting and anger management.

The *Midtown Family Place*, which has 45 families and is based in Hell's Kitchen. It provides counseling, referrals and advocacy, as well as an after-school and summer program for children ages 5-12, a literacy program, and a food pantry.

Project CLASS (Clean Living and Staying Sober), which serves as many as 50 families. It specializes in providing referrals to drug- and alcohol-abuse programs, as well as creating, implementing and monitoring drug- treatment service plans. It also includes the Babies Initiative, which is offered to 20 families with children ages five and under who are at immediate risk of being put in foster care. This intensive program works to get family members whatever services they need in order to stabilize.

Truancy-Prevention, which has 90 families with at-risk children, and conducts groups on domestic violence, groups on parenting called the Parenting Journey, as well as a group for teen-agers.

SCHOOL INVESTMENTS

HCZ Promise Academy Charter Schools

Now in its fifth year, HCZ Promise Academy 1 provides a comprehensive college-preparatory educational program, with an extended school day and school year. HCZ Promise Academy 2 is in its fourth year. Both HCZ Promise Academy and HCZ Promise Academy 2 will eventually serve children from kindergarten through twelfth grade, bringing a strong focus on literacy and mathematics (over two hours of literacy instruction and over 90 minutes of mathematics instruction each day) within a safe, structured and personalized environment. Each of the academies will be divided into four smaller "schools" (primary, elementary, middle, and high school) that will emphasize personalized relationships between students, teachers and families.

The academic day runs from 8 AM until 4 PM, approximately 20 percent longer than the vast majority of surrounding traditional public schools. Students also have the opportunity to participate in after-school programming from 4 PM – 6 PM. The academic year consists of 210 days of school, an increase over the 180 days required by law, which includes a 25-day mandatory summer program.

In 2006, a health clinic opened in the HCZ Promise Academy 1 middle-school building so the students could get free medical, dental and mental-health services. The Harlem Children's Health Project is a partnership of the Children's Health Fund, the Mailman School of Public Health at Columbia University, New York-Presbyterian Hospital and HCZ. In addition, the clinic works with the elementary schools to identify children's unmet health needs and to facilitate necessary care.

Appendix B: Variable Construction

<u>Attendance</u>

We use a student's annual total number of absences as reported by the New York Department of Education (NYCDOE), as a metric for student attendance. The NYCDOE collects student-level attendance data on a monthly basis, which we collapse for a total of an academic year (September-June). Duplicate values for students across a given month were included if the duplicate was from a different school, to incorporate students who changed schools mid-month. However, all duplicates across student ID, school ID, and month were dropped from the dataset.

<u>Close to HCZ</u>

We define those students who reside within 400 meters of the Zone's border and within the zone itself to be "Close to HCZ."

<u>Cohort Group</u>

We impute a student's cohort group by her initial kindergarten or sixth-grade year for students in the elementary-school and middle-school sample, respectively. If we do not observe a student in kindergarten, we impute her kindergarten cohort using her expected grade, which compares her date of birth with state-mandated entry dates. If we do not observe a student in sixth grade, we impute her sixth grade cohort using the first time we observe her in the years above or below sixth grade, from fourth to eighth. If we still cannot impute her sixth grade cohort, we follow the same methodology as with the kindergarten sample, and use her expected grade to determine her cohort.

<u>Age</u>

Using data provided by the HCZ for the Harlem Gems sample, we control for student age, in months in the Harlem Gems regressions.

<u>Distance to the Zone</u>

Using the ArcGIS geocoding utility, we gecode and chart student home addresses provided by the NYCDOE. We then calculate each addresses' distance to the nearest boundary of the Harlem Children's Zone using the *near* function in ArcGIS to calculate the distance from each point to the zone itself. We then identify those addresses within the zone using the ArcGIS *select by location* function, flagged them as inside the zone, and set their respective distances equal to zero. We then manually adjust our definition of the zone to include those students residing within 10 meters of the zone's boundary. The zone we consider includes the original Phase 1 of the zone, which encompasses the area from 116th street to 123rd street between 5th Avenue and 8th Avenue.

<u>Lottery Treatment</u>

Students in the lottery sample are considered a part of the "treatment" if they receive a winning lottery number or are placed high enough on the waitlist to be offered admission to the charter school. For the 2004 kindergarten lottery, the first ten students on the waitlist were considered a part of the treatment group. For the 2005 cohort, the first 18 and ten students on the waitlist are considered a part of the treatment group for kindergarten and middle school lotteries, respectively. For the 2006 middle school lottery, the first eight students on the waitlist were added to the treatment group. For the 2005 kindergarten lottery, we only consider the Promise Academy 1 lottery.

<u>Lunch Status</u>

As provided by the NYCDOE, students are defined as being eligible for free lunch, reduced-price lunch, full-price lunch, or as having an unknown lunch status. We use the most recent data available for this variable.

On Time

To determine whether a student is at or above grade level we first create a variable that identifies a student's expected grade level. For the elementary school cohort, we generate a student's expected grade by comparing her date of birth with the state-mandated school entry dates for each academic year. For the middle school cohort, we use the first year that the student was enrolled in fifth grade to compute expected grade. In both cases, we compare the student's expected grade with her reported grade level, defining a student as "on time" if her reported grade level is at or above her expected grade. We pool those students in kindergarten or pre-kindergarten, giving each student in either grade a value of zero for her reported grade level.

Race

While the dataset contains a number of race variables and we generate variables for all available races (Black, White, Hispanic, Asian, and Multi –Race) we define students as being Black Hispanic or Other (which includes missing values) in our regression. As with lunch status, we use the most recent data available from the NYCDOE to create this variable

Siblings

Sibling groups were formed using reported last names and address data for each year as provided by the NYCDOE's enrollment files.

Test Scores

Math and ELA test scores and test grade level are reported by the NYCDOE. Test scores are standardized across each grade and year to have a mean of zero and standard deviation of one. If a student is tested across multiple years for a given grade, she is assigned the standardized score from the first year that she is tested in that grade. Student percentiles for each test in each year were also generated using a student's scaled test score ranking and dividing it by the total number of scores present in a given year.

Appendix C: Statistical Tests of Cheating of at HCZ

This appendix investigates whether teacher or administrator cheating drives our results.

There have been documented cases of cheating in California (May, 1999), Massachusetts (Marcus, 2000), New York (Loughran and Comiskey, 1999), Texas (Kolker, 1999), Great Britain (Hofkins, 1995; Tysome, 1994) and Chicago (Jacob and Levitt, 2003). While these studies generally rely examination of erasure patterns and the controlled retesting of students, Jacob and Levitt (2003) develop a new method for statistically detecting cheating. Their approach is guided by the intuition that teacher cheating, especially if done in an unsophisticated manner, is likely to leave blocks of identical answers, unusual patterns of correlations across student answers within the classroom, or unusual response patterns within a student's exam.

Following Jacob and Levitt's (2003) algorithm, we use two strategies to investigate the possibility of cheating at HCZ. First, we examine blocks of consecutive questions to see if HCZ students are more likely to have an unusual string of identical answers. Second, we examine the performance of HCZ students on both "easy" and "hard" questions. If students at HCZ systematically miss the easy questions while correctly answering the hard questions, this may be an indication of cheating.[55]

We should note that there are more subtle ways teachers can cheat, such as by providing extra time to students or changing answers in a random way, that our algorithm is unlikely to detect. Even when cheating is done naively our approach is not likely to detect every instance of cheating (see Jacob and Levitt (2003) for details and calibration exercises). Our results should be interpreted with these caveats in mind.

Suspicious Answer Strings

The quickest and easiest way for a teacher to cheat is to change the same block of consecutive questions for a subset of students in his or her class. In this section we compare the most unlikely block of identical answers given on consecutive questions at HCZ to the most unlikely block of answers at other New York City schools.

To find the most unlikely string of answers we first predict the likelihood that each student will give the answer they did on each question using a multinomial logit. We estimate this model separately for each question in grade, subject and year, controlling for test score performance in that year and

[55]Jacob and Levitt (2003) also search for large, unexpected increases in test scores one year, followed by very small test score gains (or even declines) the following year. Their identification strategy exploits the fact that these two types of indicators are very weakly correlated in classrooms unlikely to have cheated, but very highly correlated in situations where cheating likely occurred. We cannot use this second measure, as administrative cheating at HCZ would imply large gains that are maintained as long as students attend.

background characteristics such as gender and free lunch eligibility.[56][57] A student's predicted probability of choosing any particular response is therefore identified by the likelihood that other students (in the same year, grade and subject) with similar background characteristics and test scores choose that response.

Using the estimates from this model, we calculate the predicted probability that each student would answer each item in the way that he or she in fact did. This provides us with one measure per student per item. Taking the product over items within each student, we calculate the probability that a student would have answered a string of consecutive questions from item *m* to item *n* as he or she did. We then take the product across all students in the classroom who had identical responses in the string. We repeat this calculation for all possible consecutive strings of length three to seven, and take the minimum of the predicted block probability for each school. This measure captures the least likely block of identical answers given on consecutive questions in the school.

The results of this procedure show that HCZ students are no more likely to have a suspicious block of answers than students at other New York City schools. On the 3rd and 8th grade ELA and Math exams (calculated separately for each cohort), HCZ's most suspicious answer string is between the 42nd and 72nd percentiles. Adjusting for school size to account for the number of strings we search over, HCZ is between the 43rd and 60th percentiles.

Performance by Question Difficulty

The typical student will answer most of the easy questions correctly but get most of the hard questions wrong (where "easy" and "hard" are based on how well students of similar ability do on the question). If students in a school systematically miss easy questions while correctly answering hard questions, this may be an indication of cheating. Our second test therefore examines the performance of HCZ students on easy and hard exam questions.

To find the likelihood that each student will get an question correct we estimate a logit for each item on the exam. We estimate this model separately for each grade, subject and year and control for test score performance in that year and background characteristics such as gender and free lunch eligibility. We then construct the residual for each student on each item and sum the residuals by item for each

---

[56] This procedure implicitly assumes that a given student's answers are conditionally uncorrelated across questions on the exam, and that answers are uncorrelated across students. While this assumption is unlikely to be true in practice, because all of our comparisons rely on the relative unusualness of the answers given in different schools, this simplifying assumption is not likely to bias our results unless the correlation within and across students varies by school.

[57] By estimating the probability of selecting each possible response, rather than simply estimating the probability of choosing the correct response, we take advantage of any additional information that is provided by particular response patterns in a classroom.

school. If students do relatively better (worse) on a question than expected, the residual will be positive (negative). If students at a school do as well on an item as their total score and background would predict, the residual will be around zero.

In Appendix Figures 2A – 2H we plot residuals for HCZ students and for students citywide. We order the questions from easiest to hardest (as determined by the percentage of students who answered that question correctly). If HCZ students are systematically cheating on hard questions, we would expect them to underperform on easy questions (to have lower residuals than the rest of the city) and to overperform on hard questions (to have higher residuals than the rest of the city). We find no evidence of this type of cheating. Residuals for HCZ students follow the same general pattern as other schools in New York City, and there is no clear evidence of "over-performance" on hard questions.

Appendix D: Cost-Benefit Calculation

This appendix describes the assumptions underlying our cost-benefit calculation.

Costs

The total per-pupil costs of the HCZ public charter schools can be calculated with relative ease. The New York Department of Education provided every charter school, including the Promise Academy, $12,443 per pupil in 2008-2009. HCZ estimates that they added an additional $4,657 per-pupil for in-school costs and approximately $2,172 per pupil for after-school and "wrap-around" programs. This implies that HCZ spends $19,272 per pupil. To put this in perspective, the median school district in New York State spent $16,171 per pupil in 2006, and the district at the 95th percentile cutpoint spent $33,521 per pupil (Zhou and Johnson, 2008).

Benefits of Achievement

There are relatively few studies relating test scores to later life outcomes. Currie and Thomas (1999) find that a one standard deviation increase in reading scores at age seven is associated with an 8.0 percent increase in wages at 33, and a one standard deviation increase in math scores is associated with 7.8 percent higher wages. Using our middle school lottery estimates from Table 2 - 0.827 in math and 0.427 in ELA - suggests a (0.827*7.6 + 0.427*8.0) = 9.7 percent increase in wages for HCZ students. Neal and Johnson (1996) find that a one standard deviation increase in AFQT scores at ages 15 – 18 is associated with a 20 percent increase in wages at ages 26 to 29. Taking an average of the estimated HCZ effect across math and reading, this corresponds to a (0.627 * 0.2) = 12.5 percent increase in wages.[58] Levitt and Lochner (2001) find that a one quartile increase in AFQT scores is associated with a 3 to 4 percent decrease in self-reported property and violent crime participation. Assuming normality and using the average effect across both math and reading, this implies a (0.627/0.67) * (3 to 4) = 2.8 to 3.7 percent decrease in criminal participation. Auld and Sidhu (2005) find that a one standard deviation increase in AFQT scores is associated with a 20 to 30 percent decrease in the probability of reporting a health limitation, implying a (0.627) * (20 to 30) = 12.5 to 18.8 percent decrease for HCZ students. Elias (2005) and Kaestner (2009) report similar findings using self-reported health status.

[58] Kruger (2003) suggests three reasons why Neal and Johnson (1996) find a larger effect of test scores on wages than Currie and Thomas (1996). First, students were older when they took the AFQT exam, and there is evidence of mean regression in test scores. Second, Currie and Thomas simultaneously enter the highly correlated reading and mathematics scores in the wage regression, whereas Neal and Johnson (1996) use just a single test score. Finally, the British and American labour markets are different in ways that may change the correlation between test scores and wages.

The above calculations assume that the relationship between test scores and future wages is causal. We can also calculate the future benefit of HCZ using the rule of thumb that a one standard deviation increase in test scores is approximately equal to an additional year of schooling. Under this scenario (and assuming that there are no signaling effects of education), our middle estimates suggest a 4.7 to 6.7 percent increase in earnings (Angrist and Krueger, 1991; Card, 1999), a 0.23 percent decrease in the probability of incarceration (Lochner and Moretti, 2004), a 2.6 percent reduction in 10-year mortality (Lleras-Muney, 2005) and a 2.9 percent lower probability of having a teen birth (Black, Devereux, and Salvanes, 2004).

Attainment

If HCZ affects educational attainment as dramatically as achievement, the implied benefits are enormous (Angrist and Krueger, 1991; Grossman and Kaestner, 1997; Card, 1999; Lantz et al., 1998; Card, 2001; Lochner and Moretti, 2004; Pettit and Western, 2004; Lleras-Muney, 2005; Belfield, 2006; Cutler and Lleras-Muney, 2006; Rouse, 2006; Levin et al., 2007; Oreopoulos, 2007; Kaestner, 2009). The public benefits alone would more than justify the costs. Using 2003 and 2004 Current Population Survey (CPS) data and the NBER TAXSIM, Rouse (2006) finds that present value lifetime earnings at age 20 of black male high school dropouts are $292,200 versus $601,800 for high school graduates--this means that the average black male dropout contributes $118,000 in income taxes over his lifetime versus $222,400 for a high school graduate. Accounting for property and sales taxes increases these figures by 5 percent. Overall, each additional black male high school graduate would produce a present value at age 20 of $167,600 in additional tax revenue. Using data from the 2002 Medical Expenditure Panel Survey (MEPS) combined with enrollment costs from the National Health Accounts (NHA), Levin et al. (2007) estimate that over the lifetime, each additional high school graduate would result in savings in public health costs with a net present value of $33,500 at age 20. Using data from the Bureau of Justice Statistics as well as FBI Uniform Crime Rate data, Belfield (2006) estimates that converting a black male high school dropout to a graduate is associated with criminal justice cost savings of $55,500. Taken together, this implies a public benefit of approximately $256,700 per new high school graduate.