American Economic Review 2016, 106(7): 1878–1920
http://dx.doi.org/10.1257/aer.20150479

Charters without Lotteries: Testing Takeovers in New Orleans and Boston[†]

By Atila Abdulkadiroğlu, Joshua D. Angrist, Peter D. Hull, and Parag A. Pathak*

Charter takeovers are traditional public schools restarted as charter schools. We develop a grandfathering instrument for takeover attendance that compares students at schools designated for takeover with a matched sample of students attending similar schools not yet taken over. Grandfathering estimates from New Orleans show substantial gains from takeover enrollment. In Boston, grandfathered students see achievement gains at least as large as the gains for students assigned charter seats in lotteries. A non-charter Boston turnaround intervention that had much in common with the takeover strategy generated gains as large as those seen for takeovers, while other more modest turnaround interventions yielded smaller effects. (*JEL* D44, H75, I21, I28)

No child's chance in life should be determined by the luck of a lottery.
— President Barack Obama

The National Alliance for Public Charter Schools (NACPS) reports a net increase of 1,092 charter schools between fall 2011 and fall 2015, with an enrollment gain of 43.6 percent. Charter growth has been especially strong in large urban districts where many students are poor and most are nonwhite. The schools in these districts are often described as low-performing, with low standardized test scores and

*Abdulkadiroğlu: Duke University, 419 Chapel Drive, Durham, NC 27708 (e-mail: aa88@duke.edu); Angrist: Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, and NBER (e-mail: angrist@mit.edu); Hull: Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139 (e-mail: hull@mit.edu); Pathak: Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, and NBER (e-mail: ppathak@mit.edu). Our thanks to Raymond Cwiertniewicz, Alvin David, Gabriela Fighetti, and Jill Zimmerman from the Recovery School District; to Kamal Chavda and the Boston Public Schools; to Scott Given, Ryan Knight, and the staff at UP Education Network; and to Andrew Bott for graciously sharing data and answering our many questions. We're grateful to Alonso Bucarey, Stephanie Cheng, Mayara Felix, Ye Ji Kee, Olivia Kim, Elizabeth Setren, Daisy Sun, and Danielle Wedde for exceptional research assistance, and to MIT SEII program directors Annice Correia Gabel and Eryn Heying for invaluable administrative support. Data from the Recovery School District were made available to us through the Institute for Innovation in Public School Choice. We gratefully acknowledge financial support from the Institute for Education Sciences (under Award R305A120269), from the National Science Foundation (under award SES-1426541), and from the Laura and John Arnold Foundation. Thanks also go to seminar participants at the Federal Reserve Bank of New York, the Stanford Graduate School of Business, the Fall 2014 NBER education meeting, and the 2015 SOLE/EALE annual meeting for helpful comments. Joshua Angrist's daughter teaches at UP Academy Boston. The views expressed here are those of the authors alone. The Obama quote appears in Harmon (2011).

[†]Go to http://dx.doi.org/10.1257/aer.20150479 to visit the article page for additional materials and author disclosure statement(s).

high truancy and dropout rates.[1] Studies using randomized admissions lotteries to evaluate urban charter schools have repeatedly and convincingly shown remarkable achievement gains for urban charter lottery winners. The external validity of these estimates is less clear, however.

In the 2014–2015 school year, the New Orleans Recovery School District (RSD) became America's first all-charter public school district. This unique transformation offers the opportunity to explore the predictive value of lottery-based charter effects. RSD emerged from a 2003 effort to improve underperforming public schools in New Orleans, home to some of the worst schools in the country. State legislation allowed the Louisiana Department of Education (LDE) to take control of, manage, and delegate the operation of low-performing schools to outside operators. New Orleans public schools that came under state control became part of RSD, while other schools remained under the authority of the Orleans Parish School Board (OPSB).[2]

Hurricane Katrina decimated New Orleans' schools in August 2005. In the aftermath of the storm, RSD took control of 114 low-performing New Orleans schools, leaving OPSB with authority over only 17 of the schools it ran before Katrina. At the same time, both RSD and OPSB converted increasing numbers of low-performing schools to charters. By fall 2008, when combined RSD and OPSB enrollment had reached 36,000 (just over half of pre-Katrina OPSB enrollment), the RSD charter share hit 49 percent. Since 2008, RSD charter enrollment growth has accelerated further: September 2014 saw the closure of the few remaining direct-run traditional public schools in RSD (OPSB continues to operate a mix of traditional and charter schools).

A distinctive feature of New Orleans' charter expansion is the fact that most of the RSD charter schools that have opened since 2008 are *takeovers*. A charter takeover occurs when an existing public school, including its facilities and staff, comes under charter management. Importantly, takeovers guarantee seats for incumbent students, "grandfathering" these students into the new school. By contrast, most charter schools in other districts open as *startups*, that is, as new schools (sometimes in existing school buildings), with no seats guaranteed by virtue of previous enrollment and an active enrollment process that uses a lottery when schools are oversubscribed.

Boston's experiment with charter takeovers has unfolded with less urgency than New Orleans', but some of the forces behind it are similar. At the end of the 2010–2011 school year, nine schools in the Boston Public School (BPS) district were closed as a consequence of their persistently low performance. Two of these schools were replaced by charters: UP Academy Boston replaced the former Gavin middle school and Boston Green Academy (BGA) replaced the former Odyssey high school. These *in-district charter schools*, known in the state bureaucracy as Type-III Horace Mann schools, mark a new approach to charter authorization and school

[1] Charter schools are publicly funded private schools that operate outside the public sector. See the National Center for Education Statistics (NCES 2013) for national enrollment statistics by sector and NACPS (2013, 2014a, and 2015) for statistics on charter growth and market share. The Center for Research on Education Outcomes (CREDO 2013a) compares the demographic characteristics of traditional public and charter school students; NACPS (2014b) gives statistics on charter shares by district.

[2] Cowen Institute (2011) outlines the history of RSD.

autonomy in Massachusetts. The Boston School Committee authorizes in-district charter schools and funds them through the BPS general budget. In-district charter teachers are also members of the Boston Teachers Union. Outside of pay and benefits, however, terms of the relevant collective bargaining agreements are waived and these schools are free to operate according to their charters. Boston's in-district charters opened with new school leaders and new teaching staff, employed on an essentially at-will basis, while guaranteeing seats to students formerly at Gavin and Odyssey ("legacy schools," in our vernacular).

This paper evaluates the causal effects of RSD and Boston takeover schools on their students' achievement using an instrumental variables (IV) strategy that exploits the grandfathering provisions used initially to fill takeover seats. By offering a tool for the evaluation of the rapidly proliferating charter takeover model, grandfathering provides the opportunity to answer new questions about urban school reform. The growing set of econometric estimates exploiting charter admissions lotteries consistently show large gains for students at urban charters, but these estimates necessarily capture causal effects only for charter applicants—a self-selected population that may be especially likely to see gains from the charter treatment.[3] By contrast, grandfathered enrollment in charter takeovers is passive: an existing population is guaranteed seats in the new school. Takeover experiments therefore identify causal effects for students who haven't actively sought a charter seat.

In addition to contributing to the long-running charter debate, our empirical results are of immediate policy interest. The proliferation of charter takeovers reflects a federal push to encourage states to "require significant changes in schools that are chronically underperforming and aren't getting better" (Duncan 2010). The FY2011 federal budget addressed this challenge with a dramatic increase in funding for School Improvement Grants (SIGs). Federal SIGs, which offer up to two million dollars annually per qualifying school, support three restructuring models; the takeover charters studied here qualify for federal support under the "restart" model (US Department of Education 2009). Large urban districts besides Boston and New Orleans have also begun experimenting with takeovers. Tennessee's Achievement School District and Michigan's Education Achievement Authority are modeled on RSD, each with a large share of charter takeovers. Philadelphia's Renaissance Initiative has likewise turned many low-performing schools over to charter management. A British takeover model has also flowered in the form of England's Academies, conversions of state-run schools that remain publicly-funded but operate with charter-like autonomy (Eyles and Machin 2015).

Our results suggest takeover enrollment boosts achievement by as much or more than the gains seen for urban charter lottery applicants. In addition to a detailed analysis of takeover treatment effects in Boston and New Orleans, we also look briefly at an alternative school restructuring model in Boston, known as a "turnaround." One turnaround intervention was charter-like, replacing most staff with young outsiders

[3] Lottery estimates are reported in, e.g., Abdulkadiroğlu et al. (2011); Angrist et al. (2012); Angrist, Pathak, and Walters (2013); Dobbie and Fryer (2011); Dobbie and Fryer (2013); Hoxby, Murarka, and Kang (2009); and Tuttle et al. (2013). Ravitch (2010, pp. 141–144) and Rothstein (2011) challenge the external validity of charter treatment effects estimated using lotteries. See also Rothstein's account of high scores at KIPP: "They select from the top of the ability distribution those lower-class children with innate intelligence, well-motivated parents, or their own personal drives, and give these children educations they can use to succeed in life" (Rothstein 2004, p. 82).

much like those employed at UP and emphasizing data-driven instruction and student discipline and comportment. Two other middle school turnarounds were more modest, involving limited reforms and less staff turnover. The first intervention appears to have generated gains as large as those seen at Boston's in-district charter middle school (subsidized, in part, by greater SIG funding), while the other turnarounds yielded less impressive effects.

I. Background

A. *Why Lottery and Grandfathering Estimates Might Differ*

A stylized sample selection model shows why the effects of charter enrollment induced by grandfathering might differ from charter gains identified by admissions lotteries.[4] Suppose students face a normally-distributed unobserved net cost of charter application, denoted η, applying when $\eta < A$ for some constant threshold A. We write the gains from charter attendance in potential outcomes notation as $Y_1 - Y_0$, also assumed to be normally distributed.

Lottery-generated admission offers, indicated by Z_L, are randomly assigned conditional on application and therefore conditionally independent of η and potential outcomes. For the purposes of this theoretical discussion we ignore ex post noncompliance with offers, assuming that any applicant offered a charter seat takes it. This implies that lottery-based comparisons of applicants identify the average causal effect for lottery applicants,

$$E[Y|Z_L = 1, \eta < A] - E[Y|Z_L = 0, \eta < A] = E[Y_1 - Y_0|\eta < A].$$

With joint normality of outcomes and costs, the average effect of charter enrollment on lottery applicants can be written

$$E[Y_1 - Y_0|\eta < A] = E[Y_1 - Y_0] - \rho(Y_1 - Y_0, \eta)\lambda(A),$$

where $\rho(Y_1 - Y_0, \eta)$ is the coefficient from a regression of gains on costs and $\lambda(A)$ is a positive Mills ratio term.

The selection-on-net-costs model suggests $\rho(Y_1 - Y_0, \eta) < 0$, since η equals costs minus benefits. This in turn implies that the average causal effect for lottery applicants exceeds the population average charter attendance gain, $E[Y_1 - Y_0]$. In other words, as in a simple Roy (1951) model, applicants selected on gains see larger causal effects than would be seen in a random sample.

In the grandfathering scenario, school districts select takeovers from a set of candidate schools judged to be underperforming. Suppose that takeover candidates have $Y^b < L$, where Y^b is a standardized baseline score, assumed here to be constant within schools, and L is a performance cutoff (e.g., a "Level 4" designation

[4] Oreopoulos (2006) similarly compares the causal returns to schooling parameters identified by alternative compulsory schooling instruments.

in Massachusetts). Suppose also that takeover events, indicated by Z_T, are as good as randomly assigned among the set of low-performing candidates (an assumption supported by the covariate balance tests discussed below). Conditional on candidacy, comparisons by grandfathering eligibility, that is, by Z_T, identify the average causal effect for students with low baseline scores,

$$E\left[Y|Z_T = 1,\, Y^b < L\right] - E\left[Y|Z_T = 0,\, Y^b < L\right] = E\left[Y_1 - Y_0|\, Y^b < L\right].$$

Again using normality, we have

$$E\left[Y_1 - Y_0|\, Y^b < L\right] = E[Y_1 - Y_0] - \rho\left(Y_1 - Y_0,\, Y^b\right)\lambda(L),$$

where $\rho\left(Y_1 - Y_0,\, Y^b\right)$ is the coefficient from a regression of gains on baseline scores and $\lambda(L)$ is a Mills ratio term. Here too, we ignore ex post noncompliance so as to focus on the takeover decision.

In this model, the correlation between baseline scores and the gains from charter enrollment determines the average causal effect identified by grandfathering. We've seen elsewhere that applicants with low baseline scores often seem to reap especially high gains from charter enrollment (e.g., Angrist et al. 2012). Most importantly, however, conditional on the baseline score used to gauge low performance, the grandfathering instrument identifies a population average treatment effect.

This discussion shows why the grandfathering identification strategy might generate a more representative average causal effect than lottery-based identification strategies, at least for populations with similar baseline scores. In practice, however, lottery applicants need not be selected on gains (indeed, Walters 2014 finds evidence for a kind of "reverse Roy" selection pattern). Ultimately, the relative magnitude of lottery- and grandfathering-based estimates is an empirical question, resolved in part by the analysis that follows.

B. *Takeovers in New Orleans RSD*

The 2008 school year marked the beginning of a period of relative stability in RSD enrollment, leadership, and finances, along with district-wide improvements in test scores. RSD achievement gains in both direct-run and charter schools are described by Figure 1, which compares post-2008 math achievement trends in RSD and OPSB with all schools in Louisiana. Average achievement for traditional and RSD charter students runs mostly below the statewide and OPSB averages, but the RSD shortfall was much reduced by 2014.

Among the RSD charter schools opened since fall 2008 and operating in spring 2014 (excluding alternative schools that serve special populations), 21 are takeovers and 13 are startups.[5] Even by the standards of the heated debate over school reform, the proliferation of charter takeovers in New Orleans has proven to be especially controversial (see, for example, Darling-Hammond 2012).

[5] See Figure B1 in the online Appendix.

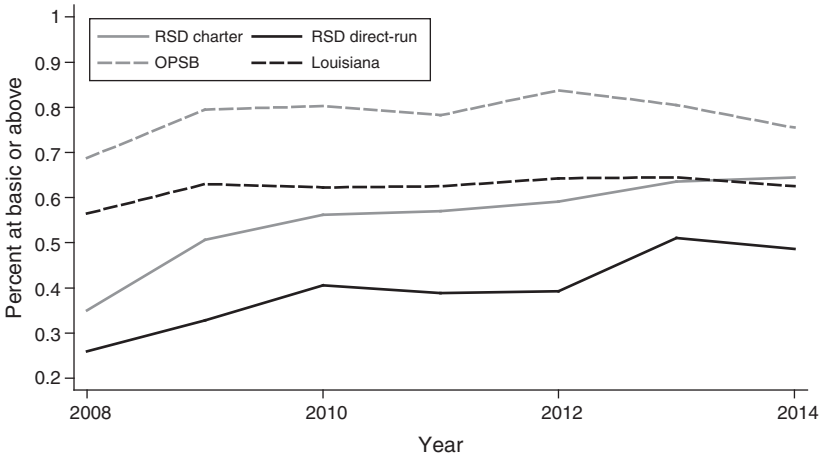

FIGURE 1. MATH SCORES IN RSD AND ELSEWHERE

Notes: This figure plots the average percentage of RSD, OPSB, and Louisiana students that achieve basic or above status on LEAP/iLEAP math exams in fifth–eighth grades. Scores for OPSB and Louisiana students are from the Louisiana Department of Education, https://www.louisianabelieves.com/resources/library/test-results (accessed October 14, 2014).

Appendix Table A1 lists the 18 New Orleans RSD schools that experienced what the district calls a *full charter takeover* between fall 2008 and fall 2013. Full takeovers convert all grades in the legacy school at the same time; the takeover school grandfathers legacy students in the relevant grades, and typically opens in the legacy school building. Alternatives to the full takeover model include principal-led conversions, phased-in takeovers, and school mergers. We focus on full takeovers because these are broad and well-defined transformations, with a clearly identified grandfathering cohort at the relevant legacy school. Our takeover analysis omits charter-to-charter takeovers, for which we were unable to construct a credible control group (though these play a role in a supplementary analysis that allows for non-takeover charter effects). The two high schools in the table are also omitted; our analysis focuses on schools with middle school grades (in RSD, these are almost all K–8 schools) because this is where takeovers are most common and because the legacy school scores used in our IV strategy are unavailable for high schoolers.[6]

The decision to effect a full takeover at a low-performing RSD school was driven in part by average test scores and in part by the availability of an interested and acceptable charter operator. Operators typically applied for a charter early in the legacy year, with some indicating a preference for specific schools. Takeover decisions were usually announced no earlier than December of the legacy year, with the charter operator selected between January and May. Low test scores figured importantly in takeover decisions, but legacy schools have not usually been the very

[6]Louisiana issues five types of charters, according to whether the charter is authorized by the local school board or the LDE, whether the school is new or a conversion, and whether it's in RSD. RSD's Type 5 charter schools, the focus of our study, are authorized and overseen by the LDE.

lowest-performers in the district. The matching strategy detailed below exploits the fact that many similarly low-performing direct-run schools continued to operate alongside legacy schools after the latter were converted to charters.

Table A1 shows that the 11 legacy schools in our study were taken over by 6 charter management organizations (CMOs), with the Crescent City and ReNEW CMOs each operating multiple schools. In two 2013 takeovers, two legacy schools were merged into a single takeover school. Table A1 also shows that seven out of nine study takeovers were operated by CMOs that identify with No Excuses pedagogy.[7] The No Excuses model for urban education—sometimes also called "high expectations"—is characterized by extensive use of tutoring and targeted remedial support, reliance on data and teacher feedback, a focus on basic skills, high expectations from students and staff, and an emphasis on discipline and comportment. New Orleans Parents' Guide school brochures suggest that almost all takeovers enacted policies associated with No Excuses, including an extended school day, student uniforms, and parent involvement groups; many also extended the school year and added weekend classes. Angrist, Pathak, and Walters (2013) and Dobbie and Fryer (2013) present evidence suggesting that No Excuses practices explain the success of urban charters in Massachusetts and New York.

RSD's charter schools function outside the collective bargaining agreement between OPSB and the United Teachers of New Orleans union that represents teachers at non-charter OPSB schools and a few OPSB charters (Cowen Institute 2011). Appendix Table A2 compares teacher characteristics, expenditure, and class size at RSD direct-run and charter schools. Teachers at RSD charters tend to have less experience and earn lower base salaries than those at direct-run schools. Class sizes at takeover and legacy schools are similar and close to those seen at other charter and direct-run public schools. Per-pupil expenditure is somewhat lower at RSD charter schools, though this may reflect differences in the student body and the teacher experience distribution. The per-pupil expenditure contrast between takeover and legacy schools shows only a small gap.

C. *UP from Gavin Middle School*

We supplement the RSD analysis with estimates of attendance effects at UP Academy, Boston's first in-district charter middle school. The UP Education Network is rapidly expanding, having recently assumed responsibility for management of two schools in Boston's Dorchester neighborhood (one elementary and one K–8), and opened two (non-charter) middle schools in Lawrence, Massachusetts. Our middle school focus necessarily excludes BGA, Boston's in-district charter high school. In this context, it's worth noting that BGA is more of an in-district conversion than a charter takeover, since it was initially staffed by BPS teachers and administrators previously employed elsewhere in the district.[8]

[7] Table B1 in the online Appendix lists sources for this classification.

[8] Concerns about low achievement and other problems led the state to put BGA on probation in October 2014.

Boston's in-district model is one of a number of policy experiments initiated at low-achieving schools in 2010.[9] As in RSD, the birth of an in-district charter reflects both the district's desire to address poor performance and the interest expressed by a qualified charter operator: UP Education Network was selected partly because UP was ready to grandfather all Gavin students (Toness 2010). Gavin students were automatically admitted to UP Boston, though a simple application form was required (UP staff visited Gavin students' homes to encourage application).[10]

Unlike other charter schools in Boston, which operate as independent districts and are funded by inter-district transfers, UP's spending appears in the BPS budget. Former Gavin teachers were free to apply for positions at UP, and a handful did so, but their positions were not grandfathered and, according to school officials, none were ultimately hired. UP administrators and staff are part of the collective bargaining units representing other BPS workers, but the school functions in a looser framework established in memoranda between UP and the district. UP is required to pay collectively bargained salaries, but school leaders and UP administrators make personnel decisions freely, as in a nonunion workplace.

As can be seen in column 8 of Table A2, which also compares teacher characteristics at the Boston schools in our study, UP's teachers are much younger than the Gavin staff they replaced: 60 percent of the UP teachers in our sample are no older than 28. This is unusually youthful even by the standards of Boston's other charter schools. UP class sizes are smaller and UP's per-pupil expenditure is somewhat lower than at the Gavin school. Like most of our RSD takeovers, the UP charter aligns the school with No Excuses principles.[11] The UP school day is two hours longer than the Gavin day had been and UP teachers are expected to report for work on August 1.

D. *Related Research on Takeovers and Turnarounds*

Dee (2012) uses the test proficiency cutoffs that determine SIG qualification to evaluate SIG participation in a regression discontinuity design. Dee's estimates suggest that SIG-funded interventions improve performance for students at treated schools. A companion difference-in-differences analysis points to the intermediate federal turnaround model as the most effective, while estimates for the remaining two SIG strategies, including restarts, are not significantly different from zero. It's worth

[9] Gavin and Odyssey were among BPS's lowest-performing schools in 2010, though not the lowest. The state categorized these schools as "Level 3," meaning they were found in the bottom 20 percent of the relevant grade-specific performance distribution. In response to our queries, BPS administrators emphasized that in-district charter conversion was one of several strategies available to the district for schools in Level 3. Lower-ranked Level 4 schools were not eligible for in-district conversion.

[10] Some high needs special education students at Gavin were grandfathered into the Richard Murphy school, which operates a satellite program in the former Gavin building (BPS 2013, p. 6, p. 146). These cases notwithstanding, the overall UP enrollment take-up rate for grandfathered special education students is close to that for other grandfathered students. Estimates conditional on baseline special education status are also similar to those from the full sample.

[11] Specifically, UP's charter application states "all stakeholders should not make or accept excuses for anything less than excellence," and describes key No Excuses practices as part of their educational programming (UP Academy 2010). More recent school documents emphasize a culture of "high expectations." (http://www.upeducationnetwork.org/uploads/documents/15-1015-UPEN-frequently-used-terms-vf.pdf, accessed May 5, 2016.)

noting, however, that very few California schools opted for the more radical restart intervention, and Dee's (2012) estimates for the restart treatment are imprecise.

Houston's pioneering Apollo 20 program revamped educational practices along No Excuses' lines in 20 of Houston's lowest performing schools, while also replacing most school leaders and half of the teaching staff in these schools; a similar effort was undertaken on a smaller scale in Denver. The insertion of charter school best practices in existing public schools provides a natural alternative to the takeover model studied here, and qualifies for the same sort of federal support. Fryer's (2014) analysis of Apollo using randomized and quasi-experimental research designs shows statistically significant gains in math of between one-fifth and one-sixth of a standard deviation, with little effect on reading. In the spirit of our grandfathering strategy, Fryer's quasi-experimental analysis uses baseline enrollment zones to construct instruments.[12]

CREDO (2013b) evaluates the effects of attending three RSD takeover charters. The CREDO study contrasts students based on baseline and post-takeover enrollment status, comparing, for example, students who move into and who exit from schools slated for charter conversion. The potential for selection bias would seem to make these sorts of comparisons hard to interpret. In related work, CREDO (2013c) reports modest gains from the New Orleans charter sector as a whole in a national matched-pair study of overall urban charter school effectiveness. Along the same lines, McEachin, Welsh, and Brewer (2014) offer a regression-controlled value-added style analysis of New Orleans school sectors that does not isolate effects on takeover students.

II. Grandfathering Identification

A. *The RSD Comparison Group*

Our grandfathering research design uses a combination of matching and regression to mitigate omitted-variables bias in comparisons of grandfathering-eligible and ineligible students. To see how the matched comparison group is constructed, consider the set of sixth graders enrolling at an RSD school slated for takeover at year's end: sixth grade legacy school enrollment entitles this group to seventh grade seats in the takeover charter. Since legacy and takeover schools in RSD typically enroll grades K–8, there are few non-legacy sixth graders who share a fifth grade school with the grandfathering-eligible group. We therefore look for a comparison group in the population of sixth graders not enrolled at the legacy school, but who attended schools similar to those attended by legacy school students in fifth grade (we refer to these as *baseline schools*). Specifically, baseline schools are considered matched when they have School Performance Scores (SPS) in the

[12] Unlike Fryer (2014), our grandfathering strategy matches on baseline school characteristics to eliminate covariate differences associated with the grandfathering instrument and allows for violations of the exclusion restriction that may compromise naïve matched comparisons. In a methodologically related study, Jacob (2004) also uses an initial condition—whether a student resides in a public housing building later slated for demolition—as an instrument for the effect of public housing on children's achievement.

Table 1—Grade Progression in the Grandfathering Research Design

	Baseline grade	Legacy grade	Takeover grades				Legacy enrollment years (No. of schools)
			First	Second	Third	Fourth	
Panel A. RSD							
Study takeovers	3	4	5	6	7	8	2009–2010 (5)
	4	5	6	7	8		2010–2011 (1)
	5	6	7	8			2011–2012 (1)
	6	7	8				2012–2013 (4)
Panel B. Boston							
UP	5	6	7	8			2010–2011 (1)
	5	7	8				
Dearborn/Harbor	5	6	7	8			2009–2010 (2)
	5	7	8				
Orchard Gardens	5	6	7	8			2009–2010 (1)
	6	7	8				

Notes: This table summarizes grade-based timing for the selection of baseline schools, grandfathering eligibility, and takeover outcomes in the RSD and Boston analysis samples. Grandfathering eligibility is determined by enrollment in the fall of the legacy enrollment year, while matching uses information from the baseline grade. Outcomes are from the spring of the corresponding school year for each takeover grade. The number of legacy schools in each academic year appears in parentheses.

same five-point bin.[13] In addition to baseline schools, the RSD comparison sample matches grandfathering-eligible and ineligible students on race, sex, baseline year, baseline special education status, and baseline subsidized lunch eligibility.

In practice, the RSD grandfathering experiment involves multiple grades, schools, and years. The relationship between legacy grades, baseline grades, and takeover grades is described in Table 1. Because the earliest available baseline information is from third grade, the RSD sample covers legacy school enrollment in grades four through seven and takeover charter enrollment in grades five through eight. Potential takeover exposure thus ranges from one year for students in seventh grade in the legacy year to four years for students in fourth grade in the legacy year (or more if grades are repeated). Students may have been eligible for grandfathering into multiple takeover charters; the grandfathering instrument indicates eligibility at any of the takeover schools we study. When pooling across grades, we retain students in the first year they become grandfathering-eligible or are matched to a grandfathering-eligible student. The number of grandfathering-eligible students enrolled in a legacy school in the fall of the year prior to takeover averages roughly 70 students per school and is about one-third the size of the matched comparison group (Table B2 in the online Appendix reports the number of observations contributed by each RSD legacy school).

The primary RSD outcomes are math and English Language Arts (ELA) scores from the Louisiana Educational Assessment Program (LEAP) in fourth and eighth grade and the Integrated Louisiana Educational Assessment Program (iLEAP) in

[13] SPS scores are used for accountability purposes within RSD. In the period relevant to our study, SPS scores ranged from 0 to 200. Our results are virtually unchanged when smaller bins are used; bins wider than about 10 points generate a coarse match with many low-scoring schools grouped together. Instrument balance, documented in Table 2 and discussed below, is driven mainly by matching on SPS bins.

grades five through seven. Scores are from spring 2011, the first post-takeover test date for the schools in our sample, through 2014.[14] A data Appendix details the construction of our analysis files from raw student enrollment, demographic, and outcome data. For the purposes of statistical analyses, scores are standardized to the population of RSD test-takers in the relevant subject, grade, and year.

As can be seen in the first two columns of Table 2, almost all RSD and RSD charter-bound students (those enrolled in an RSD charter school in the grades following baseline) are black, and most are poor enough to qualify for a subsidized lunch. RSD charter-bound students have baseline scores near the overall district mean (zero, by construction). By contrast, students who enroll in takeover charters and those eligible for grandfathering have much lower baseline test scores. For example, the average baseline math score of grandfathering-eligible students in our analysis sample is around 0.27σ below the corresponding RSD population average. This marks an important contrast with samples of lottery applicants at oversubscribed charter schools, a group that's often positively selected on baseline characteristics.[15] Columns 3–5 of Table 2 compare characteristics of takeover charter students and grandfathering-eligible students with those of the grandfathering compliers for whom grandfathering instruments identify average causal effects. The latter group is defined as the set of students who enroll in an in-district charter when grandfathered but not otherwise.[16] Compliers' baseline scores are not as low as the scores in the population of students at risk for grandfathering, but they still fall around 0.1–0.15σ below the district average.

The RSD comparison group appears to be well-matched to the RSD grandfathering-eligible sample. This is documented in column 6 of Table 2, which reports regression-adjusted differences in variables that were not used for matching. The balance coefficients in column 6 come from a model that includes a full set of matching-cell fixed effects, with no further controls. These estimates show no statistically significant differences in limited English proficiency rates or in baseline scores.

Table B3 in the online Appendix reports follow-up rates and measures of differential attrition in the RSD analysis sample. Follow-up scores are available for almost three-quarters of students in the first two post-takeover years. The follow-up rate declines in years three and four, reflecting RSD's highly mobile low-income population. Importantly, however, the likelihood an RSD student contributes an outcome score to the analysis sample is unrelated to his or her grandfathering eligibility within matching cells. Column 7 of Table 2 similarly shows that baseline covariates are balanced in the subsample for which we can measure outcomes.

[14] LEAP and iLEAP include multiple-choice and open-answer questions. LEAP scores are used for determining grade progression according to Louisiana state guidelines. The iLEAP test combines a test of academic standards and (through 2013) a norm-referenced component from the Iowa Test of Basic Skills (ITBS) through 2012–2013.

[15] In the middle school sample analyzed in Abdulkadiroğlu et al. (2011), for example, the baseline math gap between charter applicants and Boston students is around 0.36σ.

[16] Following Abadie (2003), complier means are computed as weighted averages, weighting by $\kappa = 1 - \frac{D(1-Z)}{1-E[Z|\mathbf{X}]} - \frac{(1-D)Z}{E[Z|\mathbf{X}]}$, where D denotes takeover enrollment in the first exposure year and Z denotes grandfathering eligibility. For this purpose, $E[Z|\mathbf{X}]$ is estimated by a saturated regression of the grandfathering instrument on matching-cell fixed effects.

Table 2—RSD Descriptive Statistics and Grandfathering Balance

	Sample means					Balance coefficients	
	RSD		Analysis sample				
	All students (1)	Charter-bound students (2)	Takeover charter students (3)	Grandfathering-eligible students (4)	Grandfathering compliers (5)	Analysis sample (6)	First exposure year (7)
Hispanic	0.026	0.024	0.018	0.029	0.008	—	—
Black	0.964	0.971	0.994	0.982	0.997	—	—
White	0.019	0.016	0.008	0.016	0.004	—	—
Asian	0.008	0.008	0.001	0.009	0.000	—	—
Female	0.475	0.473	0.489	0.501	0.496	—	—
Special education	0.069	0.066	0.071	0.093	0.048	—	—
Free/reduced price lunch	0.912	0.926	0.955	0.919	0.963	—	—
Limited English proficient	0.017	0.016	0.013	0.020	0.007	0.000 (0.001)	−0.001 (0.001)
N	14,575	11,381	1,040	763	2,061	3,503	2,572
Baseline math score	−0.001	0.019	−0.320	−0.266	−0.151	−0.019 (0.048)	−0.042 (0.052)
N	12,960	10,565	1,038	760	2,059	3,500	2,570
Baseline math gain	−0.099	−0.084	−0.261	−0.254	−0.281	0.007 (0.069)	0.009 (0.081)
N	4,871	4,099	330	241	819	1,235	993
Baseline ELA score	0.000	0.022	−0.303	−0.261	−0.112	−0.009 (0.048)	−0.032 (0.055)
N	12,967	10,572	1,040	762	2,061	3,502	2,572
Baseline ELA gain	−0.105	−0.097	−0.181	−0.182	−0.141	−0.015 (0.072)	0.001 (0.079)
N	4,879	4,105	330	241	819	1,235	993

Notes: This table reports sample means and coefficients from regressions of the variable in each row on a grandfathering eligibility dummy indicating enrollment in an RSD takeover legacy school in the fall of the academic year prior to takeover. Baseline test score gains are relative to the previous pre-baseline grade, constructed only for the subsample with pre-baseline information. All regressions include matching cell fixed effects (cells are defined by race, sex, special education status, subsidized lunch eligibility, baseline grade and year, and baseline school SPS scores in five-point bins). The sample in columns 3–7 is restricted to students enrolled in an RSD direct-run school at baseline. Column 1 reports means for a sample of RSD students enrolled in the same baseline years as the analysis sample, while column 2 is restricted to those students that enroll in an RSD charter school in grades following the baseline grade. Column 3 reports means for students that enroll in a takeover charter in potential takeover grades, while column 4 describes students enrolled in a legacy school. Column 5 reports means and counts of grandfathering compliers. Robust standard errors are reported in parentheses.

B. *RSD Grandfathering Graphics*

We motivate the grandfathering identification strategy for RSD with a graphical comparison of achievement trends in the grandfathering-eligible and matched comparison samples. Provided scores in the eligible cohort and the comparison group

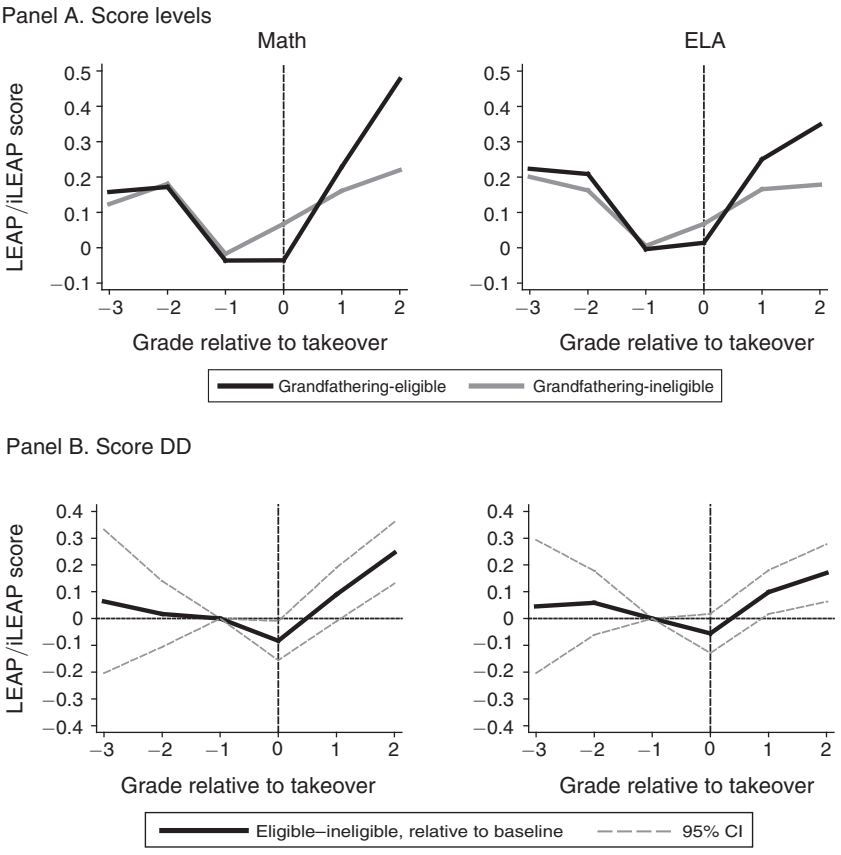

Figure 2. Test Scores in the RSD Grandfathering Sample

Notes: Panel A plots average LEAP/iLEAP math and ELA scores of students in the RSD legacy middle school matched sample. Panel B plots achievement growth relative to the baseline (−1) grade. Estimates in both panels control for matching cell fixed effects. Scores are standardized to those of students at direct-run schools in New Orleans RSD, by grade and year. Grade 0 is the last grade of legacy school enrollment.

move in parallel in the pre-takeover period, differences in score growth between eligible and ineligible students in the post-takeover period offer compelling evidence of a takeover treatment effect. The scores plotted here are standardized to samples of students at RSD's direct-run schools, so achievement trends are cast relative to this group (the statistical analysis uses scores standardized to the district, as in Table 2).

The upper panels of Figure 2 show remarkably similar pre-takeover trajectories for the math and ELA scores of grandfathering-eligible students and their matches (as for the balance regression estimates reported in Table 2, Figure 2 compares residuals from a regression on matching-cell fixed effects with no other controls). Consistent with RSD's goal of transforming low-performing schools, relative achievement in the grandfathering-eligible group declines in the grade before takeover. Importantly, the pre-takeover dip (reminiscent of the pre-treatment earnings

dip documented by Ashenfelter (1978) for applicants to job training programs) is mirrored in the matched comparison group. The comparison in Figure 2 does not adjust for baseline student achievement, so parallel trends are not guaranteed, but rather reflects the success of the matching strategy in producing similar treatment and control groups.

Matching effectively eliminates baseline differences by grandfathering status, so simple post-treatment comparisons seem likely to reveal causal effects. Difference-in-differences (DD) style comparisons of achievement growth appear in the lower panel of Figure 2, which plots achievement growth in the grandfathering-eligible and ineligible subsamples relative to the baseline grade. Pre-baseline growth differences by grandfathering status are centered around zero, while achievement contrasts after the legacy year strongly favor the grandfathered cohort. Since around 66 percent of students are caused to matriculate at a takeover charter when grandfathered (a figure reported in Table 3, panel C), this pattern suggests takeover enrollment significantly boosts achievement.

Figure 2 shows parallel pre-takeover trends in years up to, but not including, the last grade of legacy school enrollment (grade 0 in the figure). The negative and significant (for math) DD contrast in the last legacy grade signals a possible causal effect of legacy enrollment per se, regardless of whether legacy attendance leads to subsequent enrollment in the takeover charter. This is an unsurprising but potentially important finding: legacy schools were slated for closure in part because of extraordinarily low and even declining achievement. Moreover, closure itself might be disruptive, with lasting consequences for legacy students. Our grandfathering IV strategy therefore allows for direct effects of legacy school attendance when using legacy school enrollment to instrument takeover attendance.

C. *Econometric Framework*

Consider a group of legacy school students and their matched comparison counterparts with covariate values falling in a single matching stratum. Achievement for each student is observed in two grades: at the end of the legacy grade, immediately prior to the takeover (grade l), and after the takeover (grade g). The grandfathering-eligible group is mostly enrolled in the takeover school in grade g, while few in the comparison group are. A dummy variable Z—the grandfathering instrument—indicates legacy school enrollment in grade l (observed at the start of the school year) while the variable D indicates takeover school enrollment at any time in grade g. Achievement in the two grades is denoted Y^l and Y^g, observed at the conclusion of the school year.

Legacy school enrollment in grade l potentially affects grade g achievement through two channels: by increasing the likelihood of takeover attendance in grade g and by adding a year's exposure to the legacy school in grade l, an event that may have lasting consequences. Potential outcomes in grade g are therefore double-indexed. Specifically, we write Y^g_{zd} to indicate the grade g outcome that would be observed when $Z = z$ and when $D = d$. Potential outcomes in grade l, written Y^l_z, are indexed against Z alone, since grade l predates takeover exposure. Using the potential treatments notation of Imbens and Angrist (1994), legacy enrollment shifts takeover

exposure from D_0 to D_1. In this setup, observed outcomes are determined by potential outcomes and grandfathering as follows:

$$\begin{aligned}
Y^l &= Y_0^l + Z\big(Y_1^l - Y_0^l\big),\\
D &= D_0 + Z(D_1 - D_0),\\
Y^g &= Y_{00}^g + Z\big(Y_{10}^g - Y_{00}^g\big) + D\Big(Y_{01}^g - Y_{00}^g + Z\Big(Y_{11}^g - Y_{10}^g - \big(Y_{01}^g - Y_{00}^g\big)\Big)\Big)\\
&= Y_{00}^g + Z\big(Y_{10}^g - Y_{00}^g\big)\\
&\quad + (D_0 + Z(D_1 - D_0))\Big(Y_{01}^g - Y_{00}^g + Z\Big(Y_{11}^g - Y_{10}^g - \big(Y_{01}^g - Y_{00}^g\big)\Big)\Big),
\end{aligned}$$

where the last line uses the expression for D to obtain a representation for observed Y^g as a function of potential outcomes, potential treatments, and the instrument.

We assume potential outcomes and treatments satisfy the following assumptions:

ASSUMPTION 1 (Independence): $\big(Y_0^l, Y_1^l, Y_{00}^g, Y_{01}^g, Y_{10}^g, Y_{11}^g, D_0, D_1\big) \perp\!\!\!\perp Z$.

ASSUMPTION 2 (Monotonicity): $\Pr(D_1 \geq D_0) = 1$.

ASSUMPTION 3 (First stage): $\Pr(D_1 > D_0) > 0$.

Assumption 1—Independence—asserts that the grandfathering instrument is as good as randomly assigned, that is, independent of potential outcomes and potential treatment take-up (implicitly, within matching strata). Table 2 and Figure 2, which show that matching eliminates covariate and baseline score differences in our RSD analysis sample, support this. Monotonicity says that legacy enrollment either induces takeover enrollment or has no effect for everyone in the analysis sample. Assumption 3 requires legacy enrollment to induce takeover enrollment, at least for some.

As in the Angrist, Imbens, and Rubin (1996) framework for identification of local average treatment effects (LATEs) with possible violations of the exclusion restriction, Assumptions 1–3 allow for direct effects of legacy exposure on grade g outcomes. Such effects arise if $Y_{1d}^g \neq Y_{0d}^g$ when D is fixed at d. In other words, maintaining the assumption that legacy enrollment is as good as randomly assigned, we've allowed for violations of the exclusion restriction associated with use of Z as an instrument for D. This is motivated by the possibility that an additional year of exposure to a low-performing school has lasting effects.

Rather than defend a conventional exclusion restriction in this setting, we replace it with a weaker restriction on potential achievement *gains*. This allows for direct additive effects of legacy enrollment that are free to vary within the LATE subpopulations of always-takers (those with $D_1 = D_0 = 1$), never-takers (those with $D_1 = D_0 = 0$), and compliers (those with $D_1 > D_0$):

ASSUMPTION 4 (Gains Exclusion): $E\left[Y_{1d}^g - Y_1^l \,|\, T\right] = E\left[Y_{0d}^g - Y_0^l \,|\, T\right]$, *where* $T = aD_0 + n(1 - D_1) + c(D_1 - D_0)$ *identifies always-takers* (a), *never-takers* (n), *and compliers* (c).

Assumption 4 requires that expected potential achievement gains be the same for those who do and don't attend the legacy school in grade l, once takeover enrollment is fixed. This allows for $Y^g_{1d} \neq Y^g_{0d}$, while also weakening the canonical exclusion restriction applied to gains, which says that $Y^g_{1d} - Y^l_1 = Y^g_{0d} - Y^l_0$ for everyone rather than just on average. Balance in pre-baseline to baseline score gains by grandfathering eligibility status—documented for the RSD matched sample in Table 2—serves as an indirect test of this assumption.

Assumption 4 is justified by an additive structure for expected potential outcomes in each grade:

$$E\left[Y^l_z | T = s\right] = \alpha_{1s} + z\gamma_s \tag{1}$$

$$E\left[Y^g_{zd} | T = s\right] = \alpha_{2s} + z\gamma_s + d\beta_s. \tag{2}$$

The parameters α_{1s} and α_{2s} in these expressions are subgroup-specific potential outcome means with both the legacy- and takeover-enrollment indicators switched off; γ_s is an additive legacy school enrollment effect common to grades l and g, and β_s is the causal effect of takeover attendance for LATE subgroup s. This additive model rules out interactions between legacy and takeover attendance effects, while allowing legacy effects to persist across grades.

The theorem below (proved in the Appendix) shows that under Assumptions 1–4, a Wald-type IV estimator applied to achievement gains captures the average causal effects of takeover attendance on grandfathering compliers' achievement:

THEOREM 1: *Under Assumptions* 1–4,

$$\begin{aligned}\frac{E\left[Y^g - Y^l | Z = 1\right] - E\left[Y^g - Y^l | Z = 0\right]}{E[D | Z = 1] - E[D | Z = 0]} &= E\left[Y^g_{11} - Y^g_{10} | D_1 > D_0\right] \\ &= E\left[Y^g_{01} - Y^g_{00} | D_1 > D_0\right].\end{aligned}$$

In the notation of equations (1) and (2), this theorem establishes identification of β_c for a model where legacy enrollment has direct effects.

We use Theorem 1 in two ways: to capture causal effects of Bernoulli takeover enrollment in the year following a takeover and to capture causal effects of years of takeover exposure on outcomes in later years. The latter is supported by an extension of Theorem 1 detailed in the Appendix, which shows how the IV estimand for an ordered treatment can be interpreted as a convex combination of incremental average causal effects. The Appendix also discusses results from a model that relaxes Assumption 4.

These econometric considerations motivate a two-stage least squares (2SLS) estimator with second-stage estimating equation that can be written

$$Y^g_{it} - Y^l_i = \alpha' \mathbf{X}_{it} + \sum_j \kappa_j d_{ij} + \beta D_{it} + \eta_{it}, \tag{3}$$

where Y^g_{it} is student i's score in year t in grade g and Y^l_i is i's score in the last grade in which he or she was potentially enrolled at the legacy school. The treatment variable

here, D_{it}, counts the number of years student i spent at the takeover school as of year t, up to and including the grade enrolled in that year (D_{it} is Bernoulli for tests taken in the first year of takeover operation).

The first stage equation that accompanies (3) is

$$D_{it} = \delta' \mathbf{X}_{it} + \sum_j \mu_j d_{ij} + \pi Z_i + \nu_{it}, \tag{4}$$

where Z_i is the excluded instrument, an indicator of legacy enrollment in the fall of the legacy school's final year in operation, and π is the associated first stage coefficient. As with the models used to investigate covariate balance, equations (3) and (4) control for matching cell fixed effects. In particular, because the comparison sample is generated by an exact match on race, sex, baseline special education status, baseline subsidized lunch eligibility, baseline school SPS bins, baseline year, and the legacy grade, equations (3) and (4) include dummies for each of these cells, denoted d_{ij} for cell j. The empirical first- and second-stage models also include dummies for English proficiency and year-of-test (denoted by the vector $\mathbf{X}_{it}$, with coefficients α and δ). Finally, although baseline scores appear to be uncorrelated with grandfathering exposure in RSD, $\mathbf{X}_{it}$ includes these controls to boost precision.

III. Charters without Lotteries in New Orleans RSD

A. *Grandfathering Results*

Each year of enrollment in an RSD takeover charter increases math and ELA scores by an average of 0.21σ and 0.14σ, respectively (the associated standard errors are on the order of 0.04). These IV estimates, reported in the last column of Table 3, are generated by a first stage of about 1.1 years of additional takeover exposure for grandfathered students (first stage estimates are reported in column 3 of the table).[17]

Analyses that disaggregate by outcome grade and by years of potential takeover exposure show that takeover effects are larger in seventh and eighth grade than earlier, and are larger in the first two years of takeover exposure than later. The first stage effect of grandfathering eligibility on enrollment in the first exposure year, reported at the top of panel C, reveals that grandfathering boosted initial takeover enrollment rates by around 66 percentage points.

The IV estimates generated by the grandfathering design exceed (and, in many cases, are significantly different from) the corresponding OLS estimates reported in column 2 of Table 3. This suggests that uninstrumented comparisons by takeover enrollment status, such as those reported in CREDO (2013b), may be biased downward.

IV estimates that fail to adjust for possible effects of pre-takeover legacy school enrollment also appear to be too small. Fitting versions of equations (3) and (4) to post-treatment levels rather than gains generates math and ELA effects of 0.16σ

[17] In January 2016, ReNEW SciTech Academy was accused of breaking state testing rules in the 2014–2015 school year (Dreilinger 2016). This year is not in our data. Estimates of takeover effects are changed little by the omission of SciTech and ReNEW schools.

Table 3—Grandfathering IV Estimates of RSD Takeover Effects

				2SLS	
		Comparison group mean (1)	OLS (2)	First stage (3)	Enrollment effect (4)
Panel A. All grades					
(Fifth through eighth)	Math (N = 5,625)	−0.089	0.123	1.073	0.212
			(0.020)	(0.052)	(0.038)
	ELA (N = 5,621)	−0.092	0.082	1.075	0.143
			(0.018)	(0.052)	(0.039)
Panel B. By grade					
Fifth and sixth grades	Math (N = 2,579)	−0.091	0.099	0.738	0.165
			(0.035)	(0.041)	(0.068)
	ELA (N = 2,579)	−0.116	0.023	0.745	0.101
			(0.033)	(0.042)	(0.070)
Seventh and eighth grades	Math (N = 3,046)	−0.086	0.133	1.355	0.231
			(0.020)	(0.070)	(0.037)
	ELA (N = 3,042)	−0.071	0.104	1.352	0.171
			(0.019)	(0.070)	(0.036)
Panel C. By potential exposure					
First exposure year (fifth through eighth grades)	Math (N = 2,553)	−0.105	0.200	0.659	0.230
			(0.044)	(0.023)	(0.069)
	ELA (N = 2,553)	−0.103	0.099	0.659	0.197
			(0.043)	(0.023)	(0.068)
Second exposure year (sixth through eighth grades)	Math (N = 1,664)	−0.151	0.168	1.148	0.332
			(0.031)	(0.061)	(0.058)
	ELA (N = 1,664)	−0.124	0.101	1.158	0.158
			(0.028)	(0.061)	(0.051)
Third and fourth exposure year (seventh and eighth grades)	Math (N = 1,408)	0.015	0.097	1.698	0.117
			(0.022)	(0.131)	(0.042)
	ELA (N = 1,404)	−0.033	0.077	1.698	0.094
			(0.020)	(0.132)	(0.043)

Notes: This table reports OLS and 2SLS estimates of the effects of RSD takeover charter enrollment on fifth through eighth grade LEAP/iLEAP math and ELA test scores using the grandfathering eligibility instrument. The sample in columns 2–4 includes RSD direct-run school students matched to legacy school students. The endogenous regressor counts the number of years enrolled at a takeover charter prior to testing. All models control for matching strata, limited English proficiency, baseline test scores, and year/grade effects. Robust standard errors, clustered by student, are reported in parentheses. Means in column 1 are for grandfathering-ineligible matched students.

and 0.11σ, respectively. Differences between these estimates and those for gains are consistent with the negative legacy-year treatment effects suggested by Figure 2. Appendix Table A3 reports legacy year treatment effects and estimates of models that weaken Assumption 4 to allow for partial pass-through of legacy effects—these estimates differ little from the estimates reported in Table 3.[18]

B. *Interpreting RSD Takeover Effects*

The RSD grandfathering identification strategy compares students that mostly attend takeover charters with a grandfathering-ineligible comparison group that went to various sorts of schools. Most students in the comparison group began

[18] Estimates of effects on science and social science are reported in Table B4 of the online Appendix. Estimates by sex and baseline achievement appear in online Appendix Table B5.

Table 4—The Changing School Sector Distribution in RSD

	All students		Compliers	
	$Z = 0$ (1)	$Z = 1$ (2)	$Z = 0$ (3)	$Z = 1$ (4)
Panel A. First exposure year				
Study takeover	0.089	0.780	0.000	1.000
Other RSD takeover	0.072	0.032	0.087	—
Non-takeover RSD charter	0.331	0.154	0.316	—
RSD direct-run	0.508	0.034	0.597	—
N	2,027	531		
Panel B. Second exposure year				
Study takeover	0.206	0.714	0.000	1.000
Other RSD takeover	0.105	0.016	0.142	—
Non-takeover RSD charter	0.395	0.223	0.393	—
RSD direct-run	0.294	0.047	0.465	—
N	1,349	318		
Panel C. Third exposure year				
Study takeover	0.277	0.754	0.000	1.000
Other RSD takeover	0.112	0.031	0.218	—
Non-takeover RSD charter	0.450	0.188	0.517	—
RSD direct-run	0.161	0.026	0.265	—
N	795	191		
Panel D. Fourth exposure year				
Study takeover	0.316	0.646	0.000	1.000
Other RSD takeover	0.167	0.051	0.030	—
Non-takeover RSD charter	0.485	0.291	0.864	—
RSD direct-run	0.032	0.013	0.106	—
N	342	79		

Notes: Columns 1 and 2 contrast the school sector distribution for grandfathering eligible ($Z = 1$) and ineligible ($Z = 0$) students; columns 3 and 4 show the same for grandfathering compliers. The category labeled "other RSD takeover" includes charter-to-charter conversions, principal-led conversions, and mergers. Non-takeover RSD charters include startup charters created since the 2008–2009 academic year, and charters operating as of 2007–2008.

middle school at one of RSD's direct-run public schools. But the distribution of takeover alternatives evolved as RSD closed its direct-run schools and as students changed schools for reasons other than closure. Estimates of RSD takeover effects therefore reflect a growing share of charter-to-charter comparisons. If non-takeover charters also boost achievement, the takeover effects reported in Table 3 might mask a higher charter-versus-traditional average causal effect.

Table 4 describes the counterfactual school sector distribution in detail, focusing on the distinction between the charters that define the takeover treatment for the purposes of Table 3 ("study takeovers"), other takeover schools (including charter-to-charter conversions), non-takeover RSD charters, and direct-run RSD schools. Specifically, the first two columns show the distribution of school types by grandfathering status, while column 3 describes the types of schools attended by untreated compliers. Complier counterfactuals are constructed by estimating causal effects of the takeover enrollment dummy, D, on a vector of school sector indicators, $\mathbf{W}$. Associated with each $\mathbf{W}$ are potential attendance outcomes, $\mathbf{W}_0$ and $\mathbf{W}_1$, describing school choices in non-treated and treated states (that is, potential school type

when $D = 0$ and $D = 1$). Column 3 of Table 4 reports estimates of $E[\mathbf{W}_0 \,|\, D_1 > D_0]$, the distribution of school types among compliers when they don't enroll in a takeover.[19] By definition, treated compliers enroll in a takeover school when they're grandfathering-eligible; column 4 in the table is included as a reminder of this fact.

The grandfathering first stage contrasts a 78 percent first-year takeover enrollment rate for those grandfathered (reported in column 2 of Table 4) with a 9 percent comparison group enrollment rate (reported in column 1). The first-year increase in study takeover enrollment reflects a substantial reduction in attendance at non-takeover charters (compare 0.33 with 0.15) and, especially, a sharp reduction in attendance at direct-run schools (compare 0.51 with 0.03). The counterfactual attendance distribution in column 3 shows that 32 percent of untreated compliers enrolled initially in a non-takeover charter school, while 60 percent attended a direct-run school.

Not surprisingly, both the takeover first stage and the proportion of the non-grandfathered comparison group enrolled in direct-run schools shrank over our sample period. The (study) takeover first stage in the third year of potential takeover exposure was around 0.48 (0.754–0.277), while the counterfactual direct-run enrollment share for compliers fell to about 0.27. The remainder of third-year non-treated complier enrollment was in other RSD charter schools. Reflecting RSDs accelerating charter transformation, the other-charter enrollment rate for compliers exceeded 86 percent after four years of exposure.

The growing share of the RSD comparison sample enrolled in charter schools dilutes estimated takeover effects if other charter schools generate similar achievement gains. This observation motivates a 2SLS model with two endogenous variables, one tracking attendance at study charters and one tracking attendance at other charters. Our model with two charter treatments is

$$Y_{it}^{g} - Y_{i}^{l} = \alpha' \mathbf{X}_{it} + \sum_{j} \kappa_j d_{ij} + \beta_D D_{it} + \beta_C C_{it} + \eta_{it}, \tag{5}$$

where C_{it} counts the number of years of attendance in charters other than those covered by D_{it}. Equation (5) is identified here by the addition of interactions between the grandfathering instrument and covariates to the instrument list (specifically, 22 interactions with dummies for baseline year, baseline grade, and special education status/SPS bins). These interactions contribute to the first stage for C_{it} because students with differing characteristics are more or less likely to wind up in non-takeover charters in the event they aren't grandfathered. This multiple-instruments strategy therefore identifies β_D and β_C in a constant-effects framework.

As can be seen in the contrast between the estimated takeover effects reported in columns 1 and 2 of Table 5, removing other charters from the counterfactual outcome distribution increases the estimated takeover effect on math scores by

[19] The counterfactual school type distribution is estimated using a weighting scheme similar to that used to construct complier characteristics means in Table 2. The weights in this case are given by $\kappa_0 = (1 - D)\dfrac{E[Z|\mathbf{X}] - Z}{E[Z|\mathbf{X}](1 - E[Z|\mathbf{X}])}$. $E[Z|\mathbf{X}]$ is modeled using a probit specification that includes the same controls as were used for Table 3, as well as matching cell fixed effects. Abdulkadiroğlu, Angrist, and Pathak (2014) similarly estimate the counterfactual school sector distribution for applicants to Boston and New York exam schools.

Table 5—Grandfathering IV Estimates of RSD Charter Effects

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Math						
Takeover charter	0.210	0.358	0.344			
	(0.037)	(0.079)	(0.138)			
	[72.3]	[33.4]	[194.9]			
Other RSD charter		0.342	0.308			
		(0.164)	(0.376)			
		[12.6]	[27.4]			
Any RSD charter				0.366	0.359	0.385
				(0.060)	(0.095)	(0.071)
				[41.7]	[237.9]	[223.4]
No. of instruments	23	23	2	23	2	1
N	5,625	5,625	2,553	5,625	2,553	5,625
Panel B. ELA						
Takeover charter	0.141	0.154	0.205			
	(0.038)	(0.077)	(0.140)			
	[71.4]	[33.8]	[192.8]			
Other RSD charter		0.031	0.022			
		(0.161)	(0.363)			
		[12.7]	[27.1]			
Any RSD charter				0.214	0.281	0.257
				(0.061)	(0.098)	(0.072)
				[43.0]	[238.0]	[228.0]
No. of instruments	23	23	2	23	2	1
N	5,621	5,621	2,553	5,621	2,553	5,621

Notes: This table reports 2SLS estimates of the effects of enrollment in study takeovers and other RSD charters. The endogenous regressors count the number of years enrolled in RSD charters prior to testing. The grandfathering instrument is used to construct the estimates in column 6. Columns 1, 2, and 4 report 2SLS estimates from models that add interactions between grandfathering eligibility and baseline special education/SPS bin cells, baseline year, and baseline grade to the instrument list. The estimates in columns 3 and 5 add only interactions with a dummy for baseline year after 2009 and limit the sample to the first outcome grade. All models control for the covariates used to construct the estimates in Table 3. Robust standard errors, clustered by student, are reported in parentheses. First-stage *F*-statistics are reported in brackets (in models with two endogenous variables, these are as described in Angrist and Pischke 2008).

70 percent. Column 1 reports an over-identified estimate of the takeover effect analogous to the just-identified estimates of takeover effects reported in Table 3, while column 2 reports 2SLS estimates of β_D and β_C. The takeover estimate for math in the latter specification rises to 0.36σ, while the other RSD charter effect is a less precisely estimated 0.34σ. These results are similar to the estimates of math effects for Boston charter lottery applicants reported in Abdulkadiroğlu et al. (2011), and much larger than the observational estimates for New Orleans charters found by CREDO (2013c). At the same time, the other-charter ELA effect in column 2 is close to zero. Consequently, the takeover effect on ELA scores remains near 0.14σ when estimated with or without a second endogenous variable.

Although motivated by a constant-effects model, 2SLS estimates of β_D and β_C in equation (5) have a LATE interpretation when treatment is Bernoulli and average causal effects are mean-independent of the stratification variables used to generate instruments (Hull 2015). It is therefore of interest to consider models identified with

fewer interactions, for which the associated homogeneity assumptions are weaker. Column 3 of Table 5 shows estimates from the first outcome grade (so both D_{it} and C_{it} are Bernoulli), computed with instrument interactions limited to an indicator for post-2009 baseline year. The homogeneous effects assumption here requires only charter effect stability across cohorts. Though imprecise, these estimates are close to those reported in column 2.[20]

The estimates in columns 2 and 3 of Table 5 suggest takeover and other RSD charters have similar effects on math scores. Assuming effects are the same in the two types of schools, we can estimate a common charter effect by fitting a version of equation (5) that replaces $\beta_D D_{it} + \beta_C C_{it}$ with $\beta_A A_{it}$, where the variable $A_{it} = D_{it} + C_{it}$ counts years of attendance at any RSD charter. The resulting estimates of β_A, reported in columns 4 and 5 of Table 5 for the heavily over-identified and single-interaction specifications, indeed show a precision gain, with standard errors falling from 0.079 in column 2 to 0.06 in column 4 and from 0.138 in column 3 to 0.095 in column 5. As can be seen in column 6, the pooled math estimate of around 0.37 changes little when the instrument list is further reduced to a single grandfathering dummy with no interaction terms. The pooled specification for ELA yields a similar precision gain and somewhat larger estimates of the takeover effect than are generated by equation (5). The estimates of β_D and β_C for ELA, though much farther apart than the corresponding estimates for math, are imprecise enough to be compatible with the common effects specification.[21]

IV. Measuring UP in Boston

Estimates from RSD suggest takeover charters increase middle school achievement sharply, with treatment effects as large or larger than estimates for urban charter lottery applicants (compare, for example, math gains of around 0.37–0.39σ in Table 5 with lottery effects of 0.32σ reported in Table 4 of Angrist, Pathak, and Walters (2013) for a sample of Massachusetts urban charters). This weighs against the theoretical model sketched in Section I, in which Roy-type selection on potential gains makes lottery estimates misleadingly large. At the same time, RSD's rapid transformation to an all-charter district is unusual; for this and other reasons, takeover gains in historically struggling New Orleans schools need not be typical. The 2011 takeover of Boston's Gavin middle school affords another opportunity to measure charter takeover effects using the grandfathering research design, in this case against a more stable and higher-performing urban backdrop. UP is only one school, of course, but as Boston's first in-district middle school charter, UP has been focal in the debate over Boston charter policy. UP data also allow a head-to-head comparison of results for lottery applicants and grandfathered students enrolled at the same school. Finally, our Boston analysis compares takeover effects with those of other in-district reforms undertaken around the same time.

[20] A dummy for post-2009 baseline splits the sample roughly in half. Table B6 in the online Appendix explores other identification strategies.

[21] Figure B3 in the online Appendix presents visual IV-type plots that measure the goodness of fit of the common effects model. The fit appears better for math than ELA.

A. *The UP Comparison Group*

As in the RSD analysis, we use a combination of regression and matching to reduce omitted-variables bias in Boston grandfathering comparisons. Middle schoolers eligible for UP grandfathering were enrolled at Gavin in sixth or seventh grade in the fall of 2010. Because both Gavin and UP serve grades six through eight, we can match each grandfathered student to non-Gavin students who attended the same school in fifth grade. The Gavin comparison group consists of non-Gavin students matched on this baseline school, as well as race, sex, baseline special education status, and subsidized lunch eligibility (Table 1 describes the timing of the UP grandfathering research design). On-track sixth and seventh graders at Gavin transitioned to seventh and eighth grade when UP opened in fall 2011. Achievement outcomes therefore come from seventh and eighth grade Massachusetts Comprehensive Assessment System (MCAS) tests from spring 2012–2014, standardized to the population of Boston students.

Most BPS fifth graders are black or Hispanic, a fact documented in the first two columns of Table 6, which reports descriptive statistics for the Boston analysis. As can be seen in columns 2–4 of the table, blacks are overrepresented and Hispanics underrepresented in the charter-bound, UP-enrolled, and grandfathering-eligible groups. Almost all UP and grandfathering-eligible Gavin students qualify for a subsidized lunch. In contrast with the positive selection seen in the wider sample of charter-bound students, the set of students eligible for grandfathering into UP and those that subsequently enroll there have baseline scores well below those of the general BPS population. UP grandfathering compliers, described in column 5 of Table 6, are nearly two-thirds black, with baseline scores that aren't as low as those in the grandfathering-eligible sample, but still lower than the BPS average.

The extent to which matching produces balanced grandfathering comparisons is explored in the last three columns of Table 6. The balance coefficients in column 6 of the table are from models that control only for matching cells; these show statistically significant grandfathering gaps in baseline scores, suggesting the comparison group here is not as similar as that used for our analysis of RSD. Importantly, however, the difference in baseline scores can be eliminated by conditioning on further lagged scores. The power of lagged score controls to produce balanced comparisons is illustrated in column 7, which shows that the addition of fourth grade (pre-baseline) scores to the model used to estimate balance eliminates the grandfathering gap in fifth grade scores; in other words, lagged score controls neutralize differences in measured achievement in a subsequent pre-takeover grade. The estimates of UP effects that follow are therefore from models that include lagged (baseline) scores, as are the estimates of complier means in Table 6.[22]

B. *UP Estimates*

Achievement in the Gavin grandfathering cohort and in the matched comparison group moves largely in parallel in pre-takeover grades, diverging thereafter. This is documented in Figure 3, which plots achievement paths in the same format used

[22] Estimates that also match students on terciles of combined math and ELA baseline scores are similar; these appear in Table B7 in the online Appendix.

Table 6—UP Descriptive Statistics and Grandfathering Balance

	Sample means					Balance coefficients		
	Boston		Analysis sample			Analysis sample		First exposure year
	All students	Charter-bound students	UP students	Grandfathering-eligible students	Grandfathering compliers	No score controls	Lagged score controls	Lagged score controls
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Hispanic	0.346	0.275	0.240	0.241	0.228	—	—	—
Black	0.407	0.516	0.511	0.469	0.621	—	—	—
White	0.135	0.152	0.124	0.152	0.079	—	—	—
Asian	0.072	0.024	0.089	0.100	0.060	—	—	—
Female	0.483	0.502	0.489	0.483	0.545	—	—	—
Special education	0.226	0.186	0.267	0.317	0.163	—	—	—
Free/reduced price lunch	0.804	0.752	0.951	0.928	0.965	—	—	—
Limited English proficient	0.231	0.131	0.342	0.307	0.254	0.045 (0.032)	0.034 (0.033)	0.026 (0.033)
N	8,506	1,563	225	290	816	1,203	1,060	998
Baseline math score	0.006	0.171	−0.286	−0.253	−0.131	−0.117 (0.070)	−0.040 (0.053)	−0.032 (0.055)
N	8,054	1,530	210	258	821	1,142	1,037	983
Baseline math gain	0.023	0.087	0.083	0.058	0.046	0.006 (0.055)	−0.016 (0.060)	−0.005 (0.062)
N	7,468	1,355	195	239	821	1,059	975	925
Baseline ELA score	0.010	0.177	−0.273	−0.235	−0.113	−0.177 (0.065)	−0.012 (0.053)	−0.014 (0.054)
N	7,935	1,527	208	254	825	1,105	1,036	982
Baseline ELA gain	0.023	0.054	0.140	0.158	0.065	0.063 (0.059)	0.004 (0.060)	−0.005 (0.061)
N	7,373	1,356	193	235	825	1,038	973	923

Notes: This table reports sample means and coefficients from regressions of the variable in each row on a grandfathering eligibility dummy indicating enrollment in Gavin Middle School in sixth or seventh grade in the fall of 2010. Baseline test score gains are relative to the pre-baseline grade. All regressions include matching cell fixed effects (cells are defined by race, sex, special education status, subsidized lunch eligibility, and fifth grade school and year). Regressions in columns 7 and 8 also control for lagged MCAS scores (pre-baseline for baseline demographics and test scores, pre-pre-baseline for baseline score gains). The sample in columns 3–8 is restricted to students enrolled at a BPS school at baseline. Column 1 reports means for a sample of Boston students enrolled in the same baseline years as the analysis sample; column 2 is restricted to students from the Boston sample who enroll in a Boston charter school in grades six–eight. Column 3 reports means for students in the analysis sample who enroll at UP in grades seven and eight, and column 4 describes students enrolled at Gavin Middle School in the fall of 2010. Column 5 reports complier means, estimated using matching cell fixed effects and lagged scores as controls. Robust standard errors are reported in parentheses.

for RSD schools in Figure 2. The solid lines in the bottom panel compare score growth in the grandfathered and comparison groups, relative to scores from the year preceding the last year of legacy enrollment. These DD-style comparisons show marked and statistically significant differences in score growth in post-treatment years, with no significant differences earlier. In contrast with the RSD results, the Gavin experiment generates a positive DD estimate of effects on legacy-year math scores (of about one-tenth of a standard deviation). Though only marginally significant, this modest gain may reflect an effort by Gavin staff to improve outcomes in advance of—and perhaps in response to—the threat of school closure.

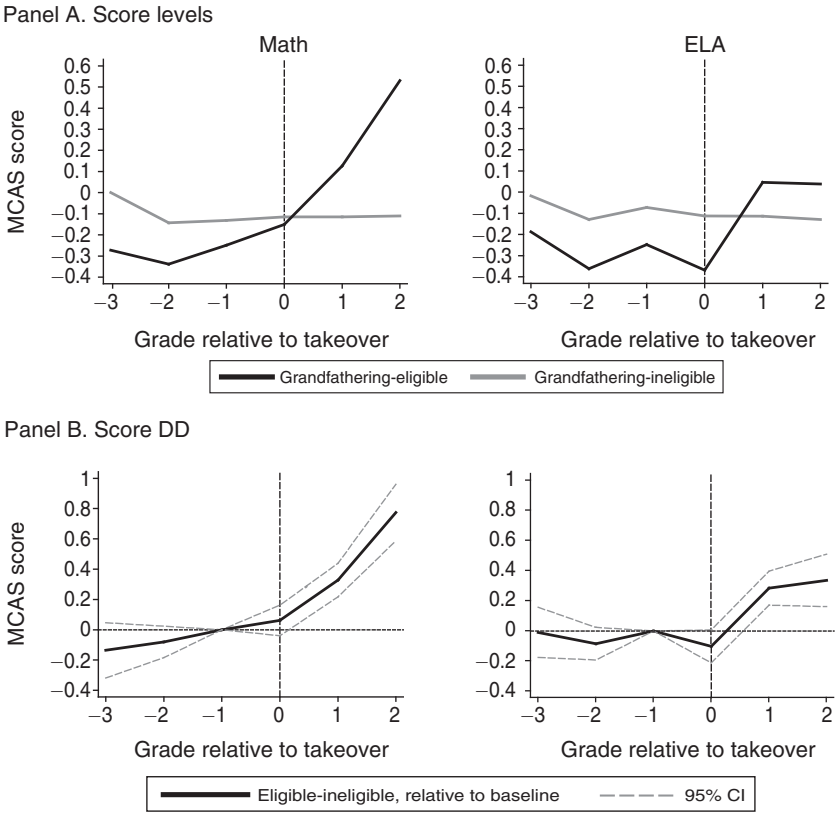

FIGURE 3. TEST SCORES IN THE UP GRANDFATHERING SAMPLE

Notes: Panel A plots average MCAS math and ELA scores of students in the Gavin Middle School matched sample. Panel B plots achievement growth relative to the baseline (-1) grade. Estimates in both panels control for matching cell fixed effects. Scores are standardized to those of BPS students, by grade and year. Grade 0 is the last grade of legacy school enrollment.

The increased UP enrollment generated by grandfathering boosted middle school math and ELA scores by an average of 0.3–0.4σ per year. This can be seen in the pooled 2SLS estimates of equation (3) reported in column 4 of panel A in Table 7. The first stage estimate for the first year of potential takeover exposure is reported in column 3 of panel B. This estimate reveals the proportion of grandfathered sixth graders who remained at UP, a little over 80 percent.

2SLS estimates of effects on math scores in the first and second years of exposure are indistinguishable, but the ELA estimate falls after the second year of exposure, from 0.5σ to 0.27σ. Given the exceptionally large first-year ELA impact this change seems unsurprising. This pattern is also consistent with Figure 3's difference-in-differences estimates for ELA, which show a post-takeover achievement jump, followed by a plateau. As with the estimates for RSD, the grandfathering estimates reported here are as large or larger than lottery-based estimates of urban charter middle school effects from Boston.[23]

[23] Compare, for example, lottery estimates of 0.25σ for math and 0.21σ for ELA reported in Table 5 of Abdulkadiroğlu et al. (2011). UP results without differencing post-takeover and legacy-grade scores are also similar to those reported in Table 7 (0.43σ for math and 0.24σ for ELA).

TABLE 7—GRANDFATHERING IV ESTIMATES OF UP EFFECTS

				2SLS	
		Comparison group mean (1)	OLS (2)	First stage (3)	Enrollment effect (4)
Panel A. All grades					
(Seventh through eighth)	Math (N = 1,543)	−0.233	0.400	1.051	0.321
			(0.032)	(0.040)	(0.039)
	ELA (N = 1,539)	−0.214	0.296	1.040	0.394
			(0.035)	(0.041)	(0.044)
Panel B. By potential exposure					
First exposure year (seventh and eighth grades)	Math (N = 1,028)	−0.214	0.365	0.822	0.325
			(0.047)	(0.025)	(0.048)
	ELA (N = 1,025)	−0.195	0.475	0.809	0.495
			(0.055)	(0.026)	(0.060)
Second exposure year (eighth grade)	Math (N = 515)	−0.272	0.408	1.541	0.324
			(0.038)	(0.087)	(0.044)
	ELA (N = 514)	−0.252	0.221	1.543	0.271
			(0.042)	(0.087)	(0.049)

Notes: This table reports OLS and 2SLS estimates of the effects of UP enrollment on seventh and eighth grade MCAS math and ELA test scores using a grandfathering instrument. The sample in columns 2–4 includes BPS students matched to a 2010–2011 sixth or seventh grade Gavin Middle School student. The endogenous regressor counts the number of years enrolled at UP prior to testing. All models control for matching strata, limited English proficiency, baseline test scores, and year/grade effects. Robust standard errors, clustered by student, are reported in parentheses. Means in column 1 are for grandfathering-ineligible matched students.

Since the fall of 2012, UP Academy, like other Boston charters, has filled its sixth grade seats through open lotteries, with priority going to current BPS students. Earlier, UP used lotteries to allocate seats not taken by grandfathering-eligible students. A natural benchmark for the Gavin grandfathering strategy is therefore the causal effect of charter attendance on UP students who participated in the lotteries used to fill the seventh grade seats not taken by former Gavin students in fall 2011, and to fill all sixth grade seats (few students apply for eighth grade seats at UP).

The UP lottery sample includes applicants who applied for sixth grade seats in the 2011–2012 and 2012–2013 school years, the first two years of UP operation. Also included are a smaller number of lottery applicants for seventh grade seats in 2011; lotteries for other entry grades through fall 2013 were not oversubscribed. Outcome data for the lottery analysis are from sixth–eighth grade tests, taken in spring 2012–2014. Table B8 in the online Appendix describes the UP lottery sample and documents baseline covariate balance and a lack of differential attrition by win/loss status. Black students are moderately overrepresented and Hispanic students somewhat underrepresented among UP lottery applicants, while poverty, special education status, and limited English proficiency rates are similar to those in the Boston population.[24]

Importantly, while UP lottery applicants' share many characteristics with other Boston students in the same grade, and their baseline scores are not very different

[24] Baseline scores for the lottery sample are from fifth grade for applicants for sixth grade seats and from sixth grade for applicants for seventh grade seats. As with the grandfathering estimates, the UP lottery sample is limited to students who attended a BPS elementary school in the baseline grade. Table B9 in the online Appendix details our lottery applicant data processing.

from the overall Boston mean, lottery applicants' baseline achievement exceeds that in the UP grandfathering sample, which has baseline scores roughly a quarter of a standard deviation below those for Boston. UP lottery applicants are also less likely than students in the grandfathered cohort to have been poor enough to qualify for a subsidized lunch.

The UP lottery estimation framework mirrors the grandfathering IV setup described by equations (3) and (4), with three modifications. First, there's no matched comparison sample. Rather, the estimation sample consists of all lottery applicants, while the empirical models adjust for year and grade of application (that is, for lottery "risk sets"), instead of matching cell fixed effects. Second, the dependent variable is the level of Y_{it}^{g} and not the gain relative to a legacy year, which is undefined in the lottery setting. Finally, as in previous lottery studies, we use two lottery instruments: an *initial offer* indicator, Z_{i1}, for students offered a seat immediately, and a *waitlist offer* indicator, Z_{i2}, for students high on the waiting list.[25] The lottery estimating equations can be written

$$Y_{it}^{g} = \alpha' \mathbf{X}_{it} + \sum_{j} \kappa_j d_{ij} + \beta D_{it} + \eta_{it} \tag{6}$$

$$D_{it} = \delta' \mathbf{X}_{it} + \sum_{j} \mu_j d_{ij} + \pi_1 Z_{i1} + \pi_2 Z_{i2} + \nu_{it}, \tag{7}$$

where dummies d_{ij} indicate lottery risk sets and $\mathbf{X}_{it}$ is a set of additional controls included to increase precision.[26] As in the grandfathering design, the endogenous variable D_{it} counts years enrolled at UP between the time of application and the outcome test date.

The first stage effect of an immediate lottery offer, close to 0.8 for the full sample, exceeds the first stage for waitlist lottery offers, which is just under 0.6. These estimates appear at the top of columns 3 and 4 in Table 8. UP lottery applicants offered a seat in sixth and seventh grade admissions lotteries earned higher math and ELA scores as a result. Pooled sixth through eighth grade 2SLS estimates, reported at the top of the last column of Table 8, show statistically significant average per-year score gains of 0.27σ in math and 0.12σ in ELA. Disaggregation by exposure time reveals larger average per-year effects after one year than after two.

The results in Tables 7 and 8 suggest that the achievement benefits of UP enrollment for those enrolled there by virtue of grandfathering are at least as large as the gains for UP students who won their seats in a lottery. For example, after one year, gains for the lottery cohort are estimated to be 0.37σ in math and 0.22σ in ELA, while gains after one year for those grandfathered into UP come to 0.33σ in math and 0.5σ in ELA. Per year gains for the grandfathered cohort after two years of potential exposure are estimated to be 0.32σ in math and 0.27σ in ELA. These estimates can be compared with estimated gains of 0.24σ in math and 0.08σ in ELA for similarly-exposed lottery cohorts. We therefore find little support for the theoretical argument that applicant

[25] Specifically, the waitlist instrument indicates applicants with lottery numbers below the highest number offered a seat in the relevant application cohort through September.

[26] As in the grandfathering analysis of UP, lottery estimates control for student race, sex, special education status, limited English proficiency, subsidized lunch status, baseline test scores, and outcome year and grade effects.

Table 8—Lottery IV Estimates of UP Effects

				2SLS		
				First stage		
		Comparison group mean (1)	OLS (2)	Immediate offer (3)	Waitlist offer (4)	Enrollment effect (5)
Panel A. All grades						
(Sixth through eighth)	Math (N = 2,202)	0.059	0.301	0.760	0.562	0.270
			(0.022)	(0.063)	(0.067)	(0.056)
	ELA (N = 2,205)	0.103	0.148	0.759	0.562	0.118
			(0.020)	(0.063)	(0.067)	(0.051)
Panel B. By potential exposure						
First exposure year (sixth and seventh grades)	Math (N = 881)	0.056	0.347	0.519	0.397	0.365
			(0.044)	(0.034)	(0.038)	(0.086)
	ELA (N = 882)	0.058	0.239	0.521	0.394	0.220
			(0.044)	(0.034)	(0.038)	(0.088)
Second and third exposure year (seventh and eighth grades)	Math (N = 1,321)	0.061	0.294	0.921	0.665	0.242
			(0.021)	(0.088)	(0.091)	(0.054)
	ELA (N = 1,323)	0.129	0.131	0.918	0.668	0.083
			(0.020)	(0.088)	(0.091)	(0.047)

Notes: This table reports OLS and 2SLS estimates of the effects of UP enrollment on sixth–eighth grade MCAS test scores using lottery offer instruments. The sample in columns 2–4 includes Boston students entering sixth grade in the 2011–2012 and 2012–2013 academic years and seventh grade in the 2011–2012 academic year. The endogenous regressor counts the number of years enrolled at UP prior to testing. The instruments are immediate and waitlist offer dummies. Immediate offer indicates applicants offered a seat immediately following the lottery in March, while waitlist offer indicates applicants offered seats later, up through the end of September. All models control for application cohort and for student race, sex, special education status, limited English proficiency, subsidized lunch status, baseline test scores, and year/grade effects. Robust standard errors, clustered by student, are reported in parentheses. Means in column 1 are for applicants not given an immediate or waitlist offer.

selection on potential achievement gains generates misleadingly large lottery estimates. On the other hand, the comparison of grandfathering and IV estimates for UP are broadly consistent with Walters' (2014) findings, which favor reverse Roy selection: students with the largest potential gains from charter attendance appear *less* likely to apply through lotteries, rather than the other way around.

Finally, as in the analysis of RSD takeover effects, an important consideration in this context is the type of school attended by the set of compliers who don't enroll at UP. Differences in counterfactual school choices might account for the smaller achievement gains seen for lottery compliers: perhaps an especially large fraction of those not offered seats in UP lotteries wound up at other high-performing Boston charters, thereby diluting lottery-generated treatment effects as in RSD. Table B10 in the online Appendix shows, however, that roughly 86 percent of untreated compliers in the grandfathering research design enrolled in a traditional BPS school, with 7 percent enrolled at another Boston charter. By way of comparison, the lottery design leaves 94 percent of untreated compliers in a traditional BPS school, with only 6 percent in other charters. This suggests that the relative magnitude of grandfathering and lottery estimates for UP is not explained by differences in non-charter enrollment.

C. *Turnarounds without Charters*

The 2010 reform that gave birth to UP sparked other Boston public school interventions as well. A dozen of the lowest-performing "Level 4" BPS schools were

restructured under either the federal transformation or turnaround models (BPS refers to all 12 as "turnaround schools"). These schools were given a longer day and assigned enhanced performance monitoring. Five school leaders were replaced, while teachers at seven schools had to reapply for their positions and many were not rehired. How do the gains from these non-charter interventions compare with the effects of a charter takeover? We use our grandfathering research design to evaluate and contrast UP attendance effects with those from non-charter, SIG-funded turnarounds at the three turnaround middle schools: Orchard Gardens, Henry Dearborn, and Harbor.[27]

The grandfathering IV strategy for Dearborn and Harbor, which serve grades six through eight, is similar to that for UP in that it compares sixth and seventh grade students enrolled in these schools in the fall of 2009 to students not eligible for grandfathering but who share a baseline (5th grade) school. For Orchard Gardens, a K–8 school, we replicate the RSD design by matching grandfathering-eligible sixth and seventh grade students to control students who attended similar schools in the previous baseline grade, where similarity is defined by the deciles of combined average math and ELA test scores (in place of RSD's SPS bins). As before, control students in both designs are also matched on baseline special education status, subsidized lunch eligibility, race, and sex. Table 1 again sketches the timing.

As with the Gavin cohort, most of the students eligible for grandfathering into the transformed Harbor and Dearborn schools are black. Most in the Orchard Gardens grandfathering cohort are Hispanic. These and other descriptive comparisons are reported in Table B11 in the online Appendix, which also shows that gaps in baseline scores by grandfathering status are eliminated by the addition of lagged (pre-baseline) score controls. Score growth from the pre-baseline to baseline year looks similar for grandfathered and matched control students, with or without further lagged score controls (these are pre-pre-baseline). Follow-up rates are similar in the grandfathering-eligible and comparison groups for all three schools, as can be seen in Table B3 in the online Appendix.

Grandfathering into the reconstituted Dearborn and Harbor schools appears to have had little effect on math scores, while increasing ELA scores by less than we've seen for students grandfathered into UP. By contrast, gains for students grandfathered into the Orchard Gardens turnaround school are similar to those enjoyed by the grandfathered cohort at UP. These findings emerge from the comparisons of score trajectories for grandfathering-eligible students and their ineligible matches in Figures 4 and 5, and are clear in the grandfathering IV estimates reported in Table 9. In particular, the IV estimates show that turnaround enrollment generates an estimated average yearly gain of 0.02σ in math and of 0.17σ in ELA for Dearborn and Harbor, while the estimated Orchard Gardens restructuring effects exceed 0.35σ, not far from those estimated for UP.

Why do the Orchard Garden effects look like those at UP, while two other turnarounds generated more modest results? As with UP, all three turnarounds benefited from an injection of federal funds, from an increased focus on teacher performance, and from a longer school day. But the experience of Orchard Gardens is notable

[27] Orchard Gardens and Harbor are pilot schools, a BPS model examined in Abdulkadiroğlu et al. (2011).

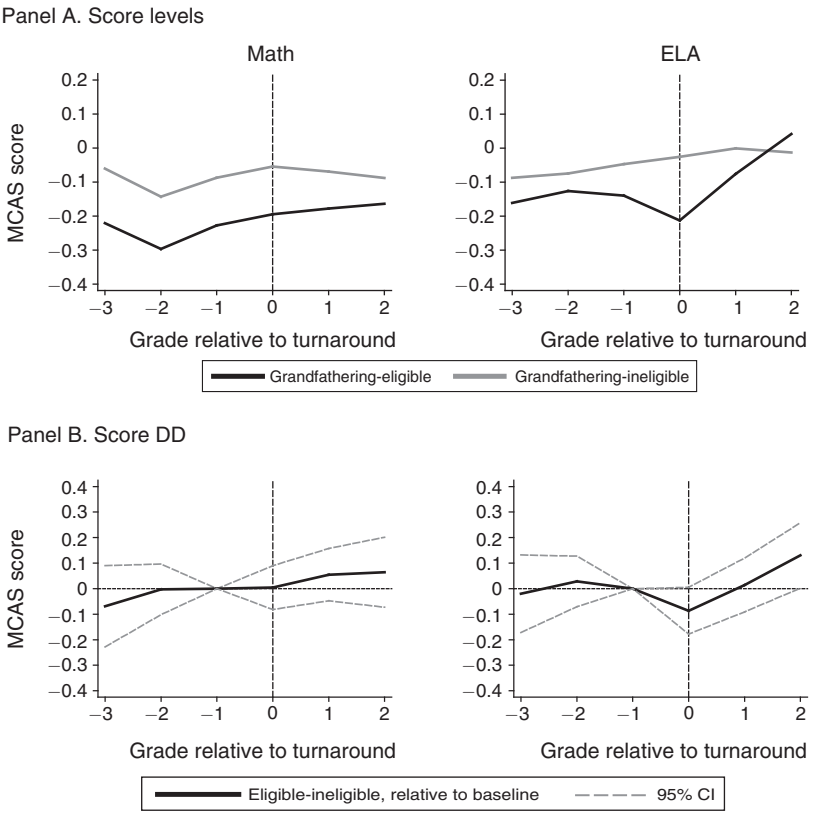

FIGURE 4. TEST SCORES IN THE DEARBORN/HARBOR GRANDFATHERING SAMPLE

Notes: Panel A plots average MCAS math and ELA scores of students in the Dearborn and Harbor legacy middle school matched sample. Panel B plots achievement growth relative to the baseline (-1) grade. Estimates in both panels control for matching cell fixed effects. Scores are standardized to the BPS population, by grade and year. Grade 0 is the last grade of legacy school enrollment.

for the intensity of its restructuring (Education Resource Strategies 2013). Orchard Gardens received almost four million dollars in SIG finding, roughly triple the SIG funding received by Dearborn, Harbor, and UP. Orchard Gardens also replaced over 80 percent of its pre-turnaround teaching staff and instituted a far longer school day than did the other two turnarounds. Orchard Gardens' extended day included homework time and tutoring sessions. This echoes changes at UP, which replaced all legacy school teachers and added two hours to the school day.[28]

In addition to more instruction time, the Orchard Gardens turnaround adopted practices similar to those used by effective urban charters. These include the hiring of a chief operating officer and a director of professional development and data, extensive use of performance monitoring software, a restructuring of curricula,

[28] See the Institute for Strategic Leadership and Learning (2013) for statistics on staff replacement. By the second turnaround year, Dearborn and Harbor had extended instruction time by 30 minutes a day, while sixth–eighth graders at Orchard Gardens saw as much as 3.5 hours added to their schedule on some days (National Center on Time and Learning 2013).

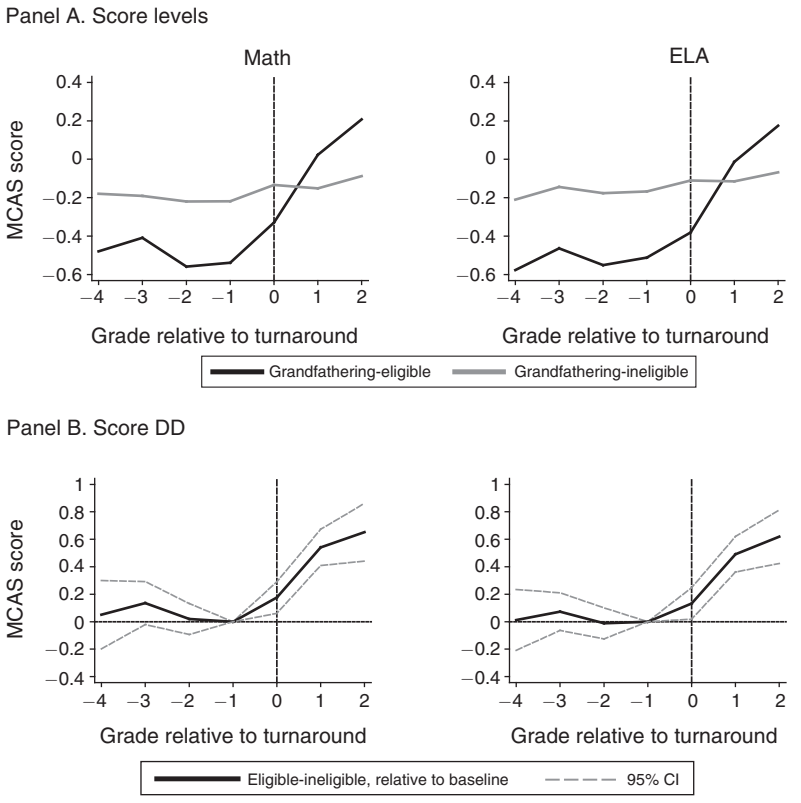

Figure 5. Test Scores in the Orchard Gardens Grandfathering Sample

Notes: Panel A plots average MCAS math and ELA scores of students in the Orchard Gardens legacy middle school matched sample. Panel B plots achievement growth relative to the baseline (-1) grade. Estimates in both panels control for matching cell fixed effects. Scores are standardized to the BPS population, by grade and year. Grade 0 is the last grade of legacy school enrollment.

an emphasis on student comportment and a climate of high expectations, and the recruitment of Teach for America and other interns (National Center on Time and Learning 2013; Education Resource Strategies 2013).[29] Table A2 shows an average teacher age at post-turnaround Orchard Gardens of around 30, a full decade younger than at Dearborn and Harbor and close to the UP average of 28. Roughly half of the Orchard Gardens post-turnaround staff were new to the district, compared with only 11 percent at Dearborn and Harbor (all of UP's initial teacher roster came from outside BPS). These statistics reinforce the view that, in addition to being unprecedented in scope and relatively resource-intensive, the Orchard Gardens turnaround had much in common with the approach taken by No Excuses charter management organizations in RSD and Boston.

[29] Many of the charter-like features of the Orchard Gardens turnaround, particularly its pedagogical similarities with a No Excuses model, were described to us in a March 2015 interview conducted with former principal Andrew Bott. In an interview with us around the same time, Massachusetts Teach For America (TFA) coordinator Josh Biber noted that roughly a quarter of Orchard Garden's post-turnaround staff were newly-placed TFA corps members.

Table 9—Grandfathering IV Estimates of BPS Turnaround Effects

	Comparison group mean (1)	OLS (2)	2SLS First stage (3)	2SLS Attendance effect (4)
Panel A. Dearborn/Harbor (UP design)				
Math (N = 1,915)	−0.149	0.019	0.971	0.022
		(0.028)	(0.043)	(0.035)
ELA (N = 1,921)	−0.063	0.089	0.981	0.174
		(0.033)	(0.043)	(0.040)
Panel B. Orchard Gardens (RSD design)				
Math (N = 2,246)	−0.234	0.307	1.113	0.367
		(0.042)	(0.046)	(0.048)
ELA (N = 2,256)	−0.179	0.352	1.114	0.397
		(0.044)	(0.046)	(0.052)

Notes: This table reports OLS and 2SLS estimates of the effects of BPS turnaround enrollment on seventh and eighth grade MCAS math and ELA test scores using the grandfathering eligibility instrument. The sample includes BPS students matched to a 2009–2010 sixth or seventh grade student at Dearborn, Harbor, or Orchard Gardens. The endogenous regressor counts the number of years enrolled at the turnaround prior to testing. All models control for matching strata, limited English proficiency, baseline test scores, and year/grade effects. Robust standard errors, clustered by student, are reported in parentheses. Column 1 reports outcome means for grandfathering-ineligible matched students.

Finally, it's worth highlighting the fact that the achievement gains generated by takeover charters and Orchard Gardens cannot be explained by changes in peer composition. This is documented in Table 10, which reports estimates of takeover and turnaround effects on peer characteristics. Specifically, the table shows 2SLS estimates of the effects of enrollment on the average baseline characteristics of peers in the same school, grade, and calendar year in the first outcome grade. These estimates reveal, for example, that students who enrolled in an RSD takeover by virtue of grandfathering were in classes with students who were slightly more likely to have limited English proficiency than would otherwise have been the case. Most importantly, students grandfathered into RSD takeovers were exposed to a marked reduction in peer achievement as a result. UP grandfathering likewise reduced peer achievement sharply, while increasing exposure to both poor and special needs peers.

Students grandfathered into the Orchard Gardens turnaround were exposed to exceptionally disadvantaged peers: turnaround effects on peer composition at this school show a 17 point increase in exposure to limited English students and a decline in peer achievement of almost half a standard deviation. The peer composition effects at Dearborn and Harbor are much more modest, as are those for UP lottery applicants. The estimates in Table 10 therefore weigh strongly against the view that peer effects are a primary determinant of education outcomes in this setting. These results also show that takeover gains in New Orleans and Boston cannot be explained by the argument that high-achieving charters push out or otherwise discourage enrollment by low-achievers. The net result of takeover and turnaround enrollment in these cities was to *increase* the share of low achieving students in affected students' classrooms.

Table 10—Effects on Peer Composition

Peer characteristic:	RSD takeovers (1)	UP (grandfathering) (2)	Orchard Gardens (3)	Dearborn/ Harbor (4)	UP (lottery) (5)
Special education	−0.001	0.042	−0.007	0.003	0.047
	(0.004)	(0.007)	(0.008)	(0.008)	(0.010)
Free/reduced price lunch	−0.007	0.113	−0.049	0.015	−0.006
	(0.004)	(0.013)	(0.011)	(0.013)	(0.019)
Limited English proficient	0.005	0.079	0.165	−0.031	−0.027
	(0.003)	(0.008)	(0.008)	(0.010)	(0.014)
Baseline math score	−0.189	−0.150	−0.456	−0.111	−0.003
	(0.017)	(0.033)	(0.036)	(0.034)	(0.040)
Baseline ELA score	−0.187	−0.146	−0.425	0.032	−0.028
	(0.019)	(0.030)	(0.029)	(0.032)	(0.037)
N	2,892	1,068	1,574	1,267	900

Notes: This table reports 2SLS estimates of the effects of enrollment at five groups of schools on the average baseline characteristics of other students enrolled in the same school, grade, and year in the first outcome grade. The samples, endogenous regressors, instruments, and controls are as in Tables 3, 7, 8, and 9. Robust standard errors, clustered by student, are reported in parentheses.

V. Summary and Conclusions

Charter school takeovers in the New Orleans Recovery School District appear to have generated substantial achievement gains for a highly disadvantaged student population that enrolled in charters passively. The New Orleans experience is undoubtedly unique in some ways. On the other hand, New Orleans schools before Katrina, while very likely among the nation's most troubled, were not uniquely low-performing—similarly low-performing districts include Atlanta, Baltimore, Chicago, Detroit, Philadelphia, and Washington, DC (NCES 2011). Of course, other districts were not called upon to weather the hurricane that eventually produced America's first all-charter public school district. It's especially noteworthy, therefore, that our analysis uncovers similarly large effects for students grandfathered into Boston's first in-district charter middle school; NCES (2011) ranks Boston as one of the better large urban districts.

Our econometric framework addresses important methodological problems that arise in the grandfathering research design. First, while legacy school enrollment provides a valuable source of exogenous variation in charter exposure, grandfathering IV strategies should adjust for possible violations of the exclusion restriction due to legacy grade exposure. Second, in an environment with schools of many types, charter treatment effects may be diluted by charter attendance in the control group; a simple 2SLS procedure allowing for multiple treatment channels yields easier-to-interpret effects. In practice, cleaning up the non-charter counterfactual substantially boosts estimates of RSD takeover effects on math, from about 0.21σ to about 0.36σ. A pooled any-charter model for RSD charters generates a common ELA effect of around 0.25σ.

The strong results for RSD and the comparison of estimates from grandfathering and lottery-based research designs for Boston's UP Academy weigh against the view that urban charter lottery applicants enjoy an unusually large and potentially

misleading gain from charter attendance because they're uniquely primed to benefit from the experience these schools offer. Boston and RSD takeovers generate gains for their passively enrolled students that are similar to, and in some specifications even larger than, the lottery estimates reported in Angrist, Pathak, and Walters (2013) for a sample of Massachusetts urban charters.

The achievement gains generated by takeover enrollment also exceed those seen for two of Boston's three turnaround middle schools. At the same time, Boston's Orchard Gardens turnaround appears to have generated gains as large as those estimated for the UP grandfathering cohort. The fact that Orchard Gardens is the most charter-like of Boston's non-charter turnarounds offers a possible explanation for this success. The lessons of the UP, Orchard Gardens, and other takeover and turnaround experiments appear to be influencing education policy in Boston (BPS 2014).

Finally, our findings highlight the question of charter access. In a pioneering effort to streamline charter admissions and broaden school choice, RSD runs a centralized match for schools in every sector. This match uses the tools of market design to reduce application costs and improve student-school matching (Abdulkadiroğlu et al. 2015). Denver, the District of Columbia, and Newark use similar unified enrollment systems (Ash 2013). Many other districts, however, have yet to integrate charter and direct-run assignment (including Boston and OPSB; see, for example, Dreilinger 2013). The results reported here suggest the possibility of gains from centralized school assignment schemes that facilitate charter attendance among students who might not otherwise choose to apply.

Appendix

THEOREM 1: *Under Assumptions* 1–4,

$$\frac{E\left[Y^g - Y^l | Z = 1\right] - E\left[Y^g - Y^l | Z = 0\right]}{E[D|Z = 1] - E[D|Z = 0]} = E\left[Y^g_{11} - Y^g_{10} | D_1 > D_0\right]$$

$$= E\left[Y^g_{01} - Y^g_{00} | D_1 > D_0\right].$$

PROOF:

Note first that the assumptions of the theorem imply

$$\text{(A1)} \quad E\left[Y^g - Y^l | Z = 1\right] - E\left[Y^g - Y^l | Z = 0\right]$$

$$= \left(E\left[Y^g_{11} - Y^l_1 | D_0 = 1\right] - E\left[Y^g_{01} - Y^1_0 | D_0 = 1\right]\right) P(D_0 = 1)$$

$$+ \left(E\left[Y^g_{10} - Y^l_1 | D_1 = 0\right] - E\left[Y^g_{00} - Y^l_0 | D_1 = 0\right]\right) P(D_1 = 0)$$

$$+ \left(E\left[Y^g_{11} - Y^l_1 | D_1 > D_0\right] - E\left[Y^g_{00} - Y^l_0 | D_1 > D_0\right]\right) P(D_1 > D_0).$$

Furthermore, as a consequence of Assumption 4, we have

$$E\left[Y_{11}^g - Y_1^l \mid D_0 = 1\right] = E\left[Y_{01}^g - Y_0^l \mid D_0 = 1\right] \tag{A2}$$

$$E\left[Y_{11}^g - Y_1^l \mid D_1 > D_0\right] = E\left[Y_{01}^g - Y_0^l \mid D_1 > D_0\right] \tag{A3}$$

and

$$E\left[Y_{00}^g - Y_0^l \mid D_1 = 0\right] = E\left[Y_{10}^g - Y_1^l \mid D_1 = 0\right] \tag{A4}$$

$$E\left[Y_{00}^g - Y_0^l \mid D_1 > D_0\right] = E\left[Y_{10}^g - Y_1^l \mid D_1 > D_0\right]. \tag{A5}$$

Equations (A2) and (A4) imply that the first two terms in (A1) equal zero. Equation (A5) and the fact that $E[D \mid Z = 1] - E[D \mid Z = 0] = P(D_1 > D_0)$ by independence and monotonicity imply further that

$$\begin{aligned}
&\frac{E\left[Y^g - Y^l \mid Z = 1\right] - E\left[Y^g - Y^l \mid Z = 0\right]}{E[D \mid Z = 1] - E[D \mid Z = 0]} \\
&\quad = E\left[Y_{11}^g - Y_1^l \mid D_1 > D_0\right] - E\left[Y_{10}^g - Y_1^l \mid D_1 > D_0\right] \\
&\quad = E\left[Y_{11}^g - Y_{10}^g \mid D_1 > D_0\right].
\end{aligned}$$

The proof is completed by noting that (A3) implies

$$\begin{aligned}
&\frac{E\left[Y^g - Y^l \mid Z = 1\right] - E\left[Y^g - Y^l \mid Z = 0\right]}{E[D \mid Z = 1] - E[D \mid Z = 0]} \\
&\quad = E\left[Y_{01}^g - Y_0^l \mid D_1 > D_0\right] - E\left[Y_{00}^g - Y_0^l \mid D_1 > D_0\right] \\
&\quad = E\left[Y_{01}^g - Y_{00}^g \mid D_1 > D_0\right]. \blacksquare
\end{aligned}$$

Our empirical work presumes that Assumptions 1–4 hold conditional on a set of mutually-exclusive and exhaustive matching cell dummies, d_j. These covariates add a layer of cross-cell averaging to the within-cell average-causal-effects interpretation of the 2SLS estimand. With matching-cell fixed effects as the only controls, the covariate parameterization is saturated. Therefore, as shown by Abadie (2003), a 2SLS regression of $Y^g - Y^l$ on D and $\{d_j\}$ that instruments D with Z identifies the treatment coefficient in a regression of $Y^g - Y^l$ on $\{d_j\}$ and D for compliers (this follows from the linearity of the propensity score in a saturated model). Angrist (1998) shows that such regressions identify variance-weighted averages of within-cell causal effects.

In practice, the grandfathering estimates reported here come from models that include additive controls for baseline covariates and year-of-test controls, as well as a full set of matching-cell fixed effects. Since the additional controls are independent of Z within cells, the weighted average interpretation of an IV estimand with fully interacted controls is unchanged, while we can expect estimates of models that include additional controls to be more precise.

A. *Extension of Theorem 1 to an Ordered Treatment*

Suppose treatment D takes on values in the set $\{0, 1, \ldots, \bar{d}\}$. Assumption 1 is modified to accommodate this ordered treatment below:

ASSUMPTION 1′: $\left(Y_0^l, Y_1^l, Y_{00}^g, \ldots, Y_{0\bar{d}}^g, Y_{10}^g, \ldots, Y_{1\bar{d}}^g, D_0, D_1\right) \perp\!\!\!\perp Z$.

We also adopt a stronger version of Assumption 4:

ASSUMPTION 4′: $P\left(Y_{1d}^g - Y_1^l = Y_{0d}^g - Y_0^l\right) = 1$ for $d \in \{0, 1, \ldots, \bar{d}\}$.

Under Assumptions 1′, 2, 3, and 4′

$$\frac{E\left[Y^g - Y^l | Z = 1\right] - E\left[Y^g - Y^l | Z = 0\right]}{E[D|Z = 1] - E[D|Z = 0]}$$

$$= \sum_{d=1}^{\bar{d}} \frac{E\left[\left(Y_{1d}^g - Y_1^l\right) - \left(Y_{1d-1}^g - Y_1^l\right) | D_1 \geq d > D_0\right] P(D_1 \geq d > D_0)}{\sum_{d=1}^{\bar{d}} P(D_1 \geq d > D_0)}$$

$$= \sum_{d=1}^{\bar{d}} \frac{E\left[Y_{1d}^g - Y_{1d-1}^g | D_1 \geq d > D_0\right] P(D_1 \geq d > D_0)}{\sum_{d=1}^{\bar{d}} P(D_1 \geq d > D_0)},$$

by Theorem 1 in Angrist and Imbens (1995). Likewise,

$$\frac{E\left[Y^g - Y^l | Z = 1\right] - E\left[Y^g - Y^l | Z = 0\right]}{E[D|Z = 1] - E[D|Z = 0]}$$

$$= \sum_{d=1}^{\bar{d}} \frac{E\left[\left(Y_{0d}^g - Y_0^l\right) - \left(Y_{0d-1}^g - Y_0^l\right) | D_1 \geq d > D_0\right] P(D_1 \geq d > D_0)}{\sum_{d=1}^{\bar{d}} P(D_1 \geq d > D_0)}$$

$$= \sum_{d=1}^{\bar{d}} \frac{E\left[Y_{0d}^g - Y_{0d-1}^g | D_1 \geq d > D_0\right] P(D_1 \geq d > D_0)}{\sum_{d=1}^{\bar{d}} P(D_1 \geq d > D_0)}.$$

The assumptions behind this interpretation of the ordered estimand are assumed to hold within matching cells, while the IV estimates of ordered treatment effects come from models that include a full set of matching-cell fixed effects. Estimate of ordered models also include a set of additive controls that should be unrelated to the instruments conditional on matching controls. Angrist and Imbens (1995) show that the IV estimand in models with an ordered treatment, saturated covariate controls, and a saturated first stage (that is a first stage that interacts Z with $\{d_j\}$), can be written as an average causal effect of a one-unit increase in treatment intensity for ordered-treatment compliers. In practice, we omit interactions of Z with $\{d_j\}$ from the first stage, except where required to identify models with multiple endogenous regressors. This omission is of little empirical consequence.

B. *Weakening Assumption 4*

The potential outcomes model described by equations (1) and (2) can be modified to allow legacy enrollment to change legacy-year and later potential outcomes to differing degrees. Identification in this case requires a covariate, so the notation here reflects this. Suppose that

$$E\left[Y_z^l | X,\, T = s\right] = \alpha_{1s}(X) + z\gamma_s(X) \tag{A6}$$

$$E\left[Y_{zd}^g | X,\, T = s\right] = \alpha_{2s}(X) + \lambda z\gamma_s(X) + d\beta_s, \tag{A7}$$

where λ is a parameter assumed to lie in the unit interval. Equations (A6) and (A7) extend equations (1) and (2) with additive effects for a Bernoulli covariate, X. Theorem 2, below, shows that the addition of covariate-instrument interactions identifies the more general model.

THEOREM 2: *Suppose Assumptions* 1 *and* 2 *hold conditional on a Bernoulli covariate*, X, *and that the conditional mean functions for potential outcomes satisfy* (A6) *and* (A7). *Suppose also that the takeover first stage varies with* X, *so that* $P(D_1 > D_0 | X = 0) \neq P(D_1 > D_0 | X = 1)$. *Then the IV estimand for a regression of* Y^g *on the pair* $\left(Y^l,\, D\right)$, *treated as endogenous and instrumented with* $(Z,\, ZX)$, *while controlling for exogenous* X, *identifies the parameters* λ *and* β_c *in equation* (A7).

PROOF.

Independence, monotonicity, and equations (A6) and (A7) imply

$$E\left[Y^l | Z = 1,\, X\right] - E\left[Y^l | Z = 0,\, X\right] = \sum_{s \in \{a,\, n,\, c\}} \gamma_s(X) P(T = s | X), \text{ and}$$

$$
\begin{aligned}
\text{(A8)}\quad & E\left[Y^g \mid Z = 1, X\right] - E\left[Y^g \mid Z = 0, X\right] \\
& = \sum_{s \in \{a, n, c\}} \lambda \gamma_s(X) P(T = s \mid X) + \beta_c P(T = c \mid X) \\
& = \lambda\Big(E\left[Y^l \mid Z = 1, X\right] - E\left[Y^l \mid Z = 0, X\right]\Big) \\
& \quad + \beta_c (E[D \mid Z = 1, X] - E[D \mid Z = 0, X]).
\end{aligned}
$$

This completes the proof since (A8) is the reduced form for the IV procedure described in the theorem. ∎

Appendix Table A3 reports estimates motivated by theorem 2 that use 22 interactions of the grandfathering instrument with baseline year, grade, and SPED/SPS bin cells instead of the single interaction the theorem requires (the model for UP includes 41 interactions with baseline year, grade, and SPED/school cells). Consistent with Figure 2, legacy year effects in RSD are estimated to be about -0.09σ for math and -0.03σ for ELA. Although λ is estimated to be about a half, RSD takeover effects estimated under a weakened Assumption 4 are similar to those estimated under gains exclusion, as can be seen by comparing the results in columns 2 and 3. As suggested by Figure 3, the legacy year treatment effect for those grandfathered into UP is positive for math and negative for ELA. These estimates are reported in column 4 of Table A3. In this case, λ is estimated to be about 0.5 for math and about 0.4 for ELA. Comparing the results in columns 5 and 6 shows the UP takeover effects reported here are also similar to those estimated under gains exclusion.

TABLE A1—RSD FULL CHARTER TAKEOVERS FROM 2008–2009 TO 2012–2013

Closure year	Legacy school	Charter legacy?	Legacy grades	Takeover school	Takeover charter network	"No Excuses" network?	Takeover grades	Study takeover?
2010	A.D. Crossman: Esperanza Charter	Yes	K–8	Esperanza Charter School	Choice		K–8	
	John Dibert Elementary		PK–8	John Dibert Community School	FirstLine	Yes	PK–8	Yes
	Laurel Elementary		PK–8	SciTech Academy at Laurel Elementary	ReNEW	Yes	PK–8	Yes
	Live Oak Elementary		PK–8	Batiste Cultural Arts Academy at Live Oak Elementary	ReNEW	Yes	PK–8	Yes
	Harney Elementary		PK–8	Edgar P. Harney Spirit of Excellence Academy	Spirit of Excellence		K–8	Yes
	Gentilly Terrace Elementary		PK–8	Gentilly Terrace School	New Beginnings		PK–8	Yes
2011	Harriet Tubman Elementary	Yes	PK–8	Harriet Tubman Charter School	Crescent City	Yes	K–8	
	Joseph S. Clark Senior High		9–12	Joseph S. Clark Preparatory High School	FirstLine	Yes	9–12	
	Sarah Towles Reed Elementary		PK–8	Dolores T. Aaron Elementary	ReNEW	Yes	PK–8	Yes
2012	McDonogh #42 Charter	Yes	PK–8	McDonogh 42 Elementary Charter School	Choice		PK–8	
	Joseph A. Craig School		PK–8	Joseph A. Craig Charter School	Friends of King	Yes	PK–8	Yes
	John McDonogh Senior High		9–12	John McDonogh High School	Future is Now		9–12	
2013	Pride College Preparatory Academy	Yes	K-5	Mildred Osborne Charter School	Arise Academy		PK–6	
	Crocker Arts and Technology School	Yes	PK-5	Lawrence D. Crocker College Prep	New Orleans College Prep		PK–5	
	Paul B. Habans Elementary School		PK-6	Paul Habans Charter School	Crescent City	Yes	PK–6	Yes
	Murray Henderson Elementary School		1–5	Paul Habans Charter School	Crescent City	Yes	PK–6	Yes
	H.C. Schaumburg Elementary School		PK-8	Schaumburg Elementary	ReNEW	Yes	PK–8	Yes
	Abramson Science and Technology School		K-8	Schaumburg Elementary	ReNEW	Yes	PK–8	Yes

Notes: This table lists RSD's full charter takeovers from the 2008–2009 to the 2012–2013 academic years. Study takeovers are those involving a public-to-charter middle school takeover. "No Excuses" networks are identified using charter applications and school or network websites. There were no full charter takeovers in the 2008–2009 academic year.

Table A2—RSD and Boston School and Teacher Characteristics

	RSD				Boston					
	Direct-run (1)	Charter (2)	Legacy (3)	Takeover (4)	BPS (5)	Charter (6)	Gavin (7)	UP (8)	Dearborn/ Harbor (9)	Orchard Gardens (10)
Panel A. School characteristics										
Student-teacher ratio	—	—	—	—	12.3	11.8	13.5	11.7	10.1	8.8
Average class size	20.3	19.4	19.9	19.7	—	—	—	—	—	—
Per-pupil expenditures										
Reported	$13,104	$11,056	$11,682	$10,934	$17,948	$14,938	$15,054	$14,586	$19,497	$16,333
Adjusted	$11,104	—	—	—	$15,419	$14,000	$12,119	$13,441	$16,646	$15,289
Panel B. Teacher characteristics										
Average age	—	—	—	—	42	32	41	28	41	30
Proportion young (age $\leq$ 28)	—	—	—	—	0.10	0.40	0.03	0.60	0.13	0.34
Average years of experience in district	12.4	7.0	—	—	12	3	13	1	9	3
Proportion new (experience $\leq$ 1)	0.18	0.28	—	—	0.06	0.28	0.03	1.00	0.11	0.49
Proportion veteran (experience $>$ 5)	0.62	0.36	—	—	0.69	0.15	0.62	0.00	0.59	0.26
Average salary	$48,080	$46,416	—	—	$81,963	$66,696	$77,251	$60,459	$68,861	$61,978

Notes: The statistics in columns 1–4 were calculated using data from the Louisiana Department of Education, http://www.louisianabelieves.com. Statistics in columns 5–10 are based on data from the Massachusetts Student Information Management System and the Educational Personnel Information Management System, as well as information on expenditures and teacher salaries from the Massachusetts Department of Elementary and Secondary Education websites, http://www.doe.mass.edu, http://profiles.doe.mass.edu, and http://www.doe.mass.edu/finance/statistics/ (accessed on October 8, 2014). The online Appendix provides additional documentation of the sources and methods used to construct this table.

Table A3—Relaxing Gains Exclusion

	RSD			UP		
	Legacy score (1)	Outcome score (2)	Outcome gain (3)	Legacy score (4)	Outcome score (5)	Outcome gain (6)
Panel A. Math						
Legacy enrollment	−0.088			0.106		
	(0.040)			(0.046)		
Takeover enrollment		0.186	0.210		0.381	0.336
		(0.033)	(0.037)		(0.037)	(0.037)
Legacy score persistence		0.520			0.533	
		(0.125)			(0.128)	
No. of instruments		23	23		42	42
N	2,553	5,625	5,625	1,028	1,543	1,543
Panel B. ELA						
Legacy enrollment	−0.030			−0.149		
	(0.042)			(0.049)		
Takeover enrollment		0.130	0.141		0.327	0.410
		(0.031)	(0.038)		(0.040)	(0.042)
Legacy score persistence		0.509			0.381	
		(0.138)			(0.106)	
No. of instruments		23	23		42	42
N	2,553	5,621	5,621	1,025	1,539	1,539

Notes: This table compares 2SLS estimates of takeover enrollment effects estimated under alternative assumptions about the persistence of legacy score effects. The outcomes, sample and endogenous variables are as in Table 3 (for RSD) and Table 7 (for UP). The instruments used for columns 2–3 and 5–6 are grandfathering eligibility interacted with baseline year, special education status, and baseline school SPS bin (RSD) or school (UP). The estimates in columns 2 and 5 treat legacy scores and takeover enrollment as endogenous. Column 1 reports the average effect of grandfathering eligibility on legacy scores, estimated by OLS. All models control for matching strata, limited English proficiency, baseline test scores, and year/grade effects. Robust standard errors, are reported in parentheses.

REFERENCES

Abadie, Alberto. 2003. "Semiparametric Instrumental Variable Estimation of Treatment Response Models." *Journal of Econometrics* 113 (2): 231–63.

Abdulkadiroğlu, Atila, Joshua D. Angrist, Susan M. Dynarski, Thomas J. Kane, and Parag A. Pathak. 2011. "Accountability and Flexibility in Public Schools: Evidence from Boston's Charters and Pilots." *Quarterly Journal of Economics* 126 (2): 699–748.

Abdulkadiroğlu, Atila, Joshua D. Angrist, Peter D. Hull, and Parag A. Pathak. 2016. "Charters without Lotteries: Testing Takeovers in New Orleans and Boston: Dataset." *American Economic Review*. http://dx.doi.org/10.1257/aer.20150479.

Abdulkadiroğlu, Atila, Joshua D. Angrist, and Parag Pathak. 2014. "The Elite Illusion: Achievement Effects at Boston and New York Exam Schools." *Econometrica* 82 (1): 137–96.

Abdulkadiroğlu, Atila, Yeon-koo Che, Parag A. Pathak, and Alvin E. Roth. 2015. "Minimal-Envy Efficient Assignment Mechanisms: The Design of New Orleans OneApp." Unpublished.

Angrist, Joshua D. 1998. "Estimating the Labor Market Impact of Voluntary Military Service Using Social Security Data on Military Applicants." *Econometrica* 66 (2): 249–88.

Angrist, Joshua D., Susan M. Dynarski, Thomas J. Kane, Parag A. Pathak, and Christopher R. Walters. 2012. "Who Benefits from KIPP?" *Journal of Policy Analysis and Management* 31 (4): 837–60.

Angrist, Joshua D., and Guido W. Imbens. 1995. "Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity." *Journal of the American Statistical Association* 90 (430): 431–42.

Angrist, Joshua D., Guido Imbens, and Donald B. Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association* 91 (434): 444–55.

Angrist, Joshua D., Parag A. Pathak, and Christopher R. Walters. 2013. "Explaining Charter School Effectiveness." *American Economic Journal: Applied Economics* 5 (4): 1–27.

Angrist, Joshua D., and Jörn-Steffen Pischke. 2008. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton: Princeton University Press.

Ashenfelter, Orley. 1978. "Estimating the Effect of Training Programs on Earnings." *Review of Economics and Statistics* 60 (1): 47–57.

Ash, Katie. 2013. "Charters Adopt Common Application Systems." *Education Week*, September 24, 2013. http://www.edweek.org/ew/articles/2013/09/25/05charter_ep.h33.html (accessed December 17, 2014).

Boston Public Schools (BPS). 2013. "BPS Five-Year Capital Facilities Master Plan, Fiscal Years 2014–2018, Volume II—School Building Overviews." http://www.bostonpublicschools.org/cms/lib07/MA01906464/Centricity/Domain/111/capital_facility_master_plan_fy2014-2018_volume_2_final.pdf (accessed February 4, 2015).

Boston Public Schools (BPS). 2014. "Strengthening the Dearborn." http://www.bostonpublicschools.org/Page/3095 (accessed March 21, 2015).

Center for Research on Education Outcomes (CREDO). 2013a. *National Charter School Study*. https://credo.stanford.edu/documents/NCSS%202013%20Final%20Draft.pdf (accessed October 10, 2014).

Center for Research on Education Outcomes (CREDO). 2013b. *New Schools for New Orleans, Year 2 Report*. http://credo.stanford.edu/pdfs/NSNOYear2Report.pdf (accessed October 10, 2014).

Center for Research on Education Outcomes (CREDO). 2013c. *Charter School Performance in Louisiana*. Center for Research on Education Outcomes. https://credo.stanford.edu/documents/la_report_2013_7_26_2013_final.pdf (accessed October 3, 2015).

Cowen Institute. 2011. *Transforming Public Education in New Orleans: The Recovery School District 2003–2011*. New Orleans: Tulane University. http://www.coweninstitute.com/wp-content/uploads/2011/12/History-of-the-RSD-Report-2011.pdf (accessed October 10, 2014).

Darling-Hammond, Linda. 2012. "Why is Congress Redlining Our Schools?" *The Nation*, January 11, 2012. http://www.thenation.com/article/165575/why-congress-redlining-our-schools (accessed December 17, 2014).

Dee, Thomas. 2012. "School Turnarounds: Evidence from the 2009 Stimulus." National Bureau of Economic Research Working Paper 17990.

Dobbie, Will, and Roland G. Fryer, Jr. 2011. "Are High-Quality Schools Enough to Increase Achievement among the Poor? Evidence from the Harlem Children's Zone." *American Economic Journal: Applied Economics* 3 (3): 158–87.

Dobbie, Will, and Roland G. Fryer, Jr. 2013. "Getting beneath the Veil of Effective Schools: Evidence from New York City." *American Economic Journal: Applied Economics* 5 (4): 28–60.

Dreilinger, Danielle. 2013. "Orleans Parish School Board Pulls Back on OneApp, Lets Schools Choose Students." *Times-Picayune*, June 18, 2013. http://www.nola.com/education/index.ssf/2013/06/orleans_parish_school_board_pu.html (accessed March 27, 2015).

Dreilinger, Danielle. 2016. "Fraud in Testing, Special Education Found at ReNEW Charter School." *Times-Picayune*, January 29, 2016. http://www.nola.com/education/index.ssf/2016/01/renew_charter_fraud_report.html (accessed February 3, 2016).

Duncan, Arne. 2010. "Secretary Duncan Message to School Boards." US Department of Education video, 3:58. May 27, 2010. https://www.youtube.com/watch?v=DlmyBh3dTUYl (accessed November 30, 2014).

Education Resource Strategies. 2013. *School Turnaround in Boston Public Schools*. Boston Public Schools Case Study. http://www.erstrategies.org/cms/files/1488-boston-case-study.pdf (accessed March 11, 2015).

Eyles, Andrew, and Stephen Machin. 2015. "The Introduction of Academy Schools to England's Education." CEP Discussion Paper 1368.

Fryer, Roland G., Jr. 2014. "Injecting Charter School Best Practices into Traditional Public Schools: Evidence from Field Experiments." *Quarterly Journal of Economics* 129 (3): 1355–1407.

Harmon, Lawrence. 2011. "Unlocking Potential at a New Charter School." *Boston Globe*, March 11, 2011. http://www.boston.com/news/education/k_12/articles/2011/03/13/unlocking_potential_at_a_new_charter_school/ (accessed January 19, 2016).

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

Hull, Peter. 2015. "IsoLATEing: Identifying Counterfactual-Specific Treatment Effects by Stratified Comparisons." Unpublished.

Imbens, Guido W., and Joshua D. Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica* 62 (2): 467–75.

Institute for Strategic Leadership and Learning (**INSTLL**)**.** 2013. *Emerging and Sustaining Practices for School Turnaround: An Analysis of School and District Practices, Systems, Policies, and Use of Resources Contributing to the Successful Turnaround Efforts in Level 4 Schools*. http://www.doe.mass.edu/boe/docs/fy2013/2013-06/spec-item1-emergingpractices.docx (accessed March 11, 2015).

Jacob, Brian A. 2004. "Public Housing, Housing Vouchers, and Student Achievement: Evidence from Public Housing Demolitions in Chicago." *American Economic Review* 94 (1): 233–58.

McEachin, Andrew, Richard O. Welsh, and Dominic J. Brewer. 2014. "Student Achievement within a Portfolio Management Model: Early Results from New Orleans." Unpublished.

National Alliance for Public Charter Schools (**NACPS**)**.** 2013. "Back to School Tallies: Estimated Number of Public Charter Schools & Students, 2012–2013." http://www.publiccharters.org/wp-content/uploads/2014/01/NAPCS-2012-13-New-and-Closed-Charter-Schools_20130114T161322.pdf (accessed October 10, 2014).

National Alliance for Public Charter Schools (**NACPS**)**.** 2014a. "Back to School Tallies: Estimated Number of Public Charter Schools and Students, 2013–2014." http://www.publiccharters.org/wp-content/uploads/2014/02/New-and-Closed-Report-February-20141.pdf (accessed August 12, 2015).

National Alliance for Public Charter Schools (**NACPS**)**.** 2014b. *A Growing Movement: America's Largest Charter School Communities*. Washington, DC: NACPS. http://www.publiccharters.org/wp-content/uploads/2014/12/2014_Enrollment_Share_FINAL.pdf (accessed August 12, 2015).

National Alliance for Public Charter Schools (**NACPS**)**.** 2015. "Back to School Tallies: Estimated Number of Public Charter Schools and Students, 2014–2015." http://www.publiccharters.org/wp-content/uploads/2015/02/open_closed_FINAL1.pdf (accessed August 12, 2015).

National Center for Education Statistics (**NCES**)**.** 2011. *The Nation's Report Card: Trial Urban District Assessment: Mathematics, 2011*. Alexandria, VA: US Department of Education. http://nces.ed.gov/nationsreportcard/pubs/dst2011/2012452.aspx (accessed September 6, 2015).

National Center for Education Statistics (**NCES**)**.** 2013. "Digest of Education Statistics, Table 216.20: Number and enrollment of public elementary and secondary schools, by school level, type, and charter and magnet status: Selected years, 1990–91 through 2011–12." http://nces.ed.gov/programs/digest/d13/tables/dt13_216.20.asp (accessed December 17, 2014).

National Center on Time and Learning. 2013. *Transforming Schools through Expanded Learning Time: Orchard Gardens K–8 Pilot School, Update 2013*. http://www.timeandlearning.org/sites/default/files/resources/orchardgardencasestudy.pdf (accessed March 11, 2015).

Oreopoulos, Philip. 2006. "Estimating Average and Local Average Treatment Effects of Education when Compulsory Schooling Laws Really Matter." *American Economic Review* 96 (1): 152–75.

Ravitch, Diane. 2010. *The Death and Life of the Great American School System: How Testing and Choice Are Undermining Education*. New York: Basic Books.

Rothstein, Richard. 2004. *Class and Schools: Using Social, Economic, and Educational Reform to Close the Black-White Achievement Gap.* New York: Teachers College Press.

Rothstein, Richard. 2011. "Grading the Education Reformers." *Slate*, August 29, 2011. http://www.slate.com/articles/arts/books/2011/08/grading_the_education_reformers.html (accessed November 30, 2014).

Roy, Arthur. 1951. "Some Thoughts on the Distribution of Earnings." *Oxford Economic Papers* 3 (2): 135–146.

Toness, Bianca Vázquez. 2010. "Charter School Will Do Things the Hard Way." 90.9WBUR. http://www.wbur.org/2010/12/09/charter-challenges (accessed October 17, 2014).

Tuttle, Christina Clark, Brian Gill, Philip Gleason, Virginia Knechtel, Ira Nichols-Barrer, and Alexandra Resch. 2013. *KIPP Middle Schools: Impacts on Achievement and Other Outcomes*. Princeton: Mathematica Policy Research.

UP Academy. 2010. "Application for a Massachusetts Horace Mann III Public Charter School: UP Academy Charter School of Boston." Submitted November 8, 2010. http://www.doe.mass.edu/charter/finalists/10/UPApp.doc (accessed April 16, 2014).

US Department Of Education. 2009. "Guidance on School Improvement Grants Under Section 1003(g) of the Elementary and Secondary Education Act of 1965." http://www2.ed.gov/programs/sif/guidance-20091218.doc (accessed October 10, 2014).

Walters, Christopher R. 2014. "The Demand for Effective Charter Schools." National Bureau of Economic Research Working Paper 20640.