A Data Appendix

The data used for this study come from several sources. Lists of charter applicants and lottery winners are constructed from records provided by individual charter schools. Information on schools attended and student demographics come from the Student Information Management System (SIMS), a centralized database that covers all public school students in Massachusetts. Test scores are from the Massachusetts Comprehensive Assessment System (MCAS). Advanced Placement (AP) and Scholastic Aptitude Test (SAT) scores are provided by the College Board. College attendance information comes from the National Student Clearinghouse (NSC). This Appendix describes each data source and details the procedures used to clean and match them.

Lottery Data

Data description and sample restrictions

Our sample of applicants is obtained from records of lotteries held at six Massachusetts charter schools between 2002 and 2009. The participating schools and lottery years are listed in Table A1. A total of 26 school-specific entry cohorts are included in the analysis. Lotteries for three participating schools, Match, Codman Academy and City on a Hill, were conducted for entry to 9th grade; two schools, Boston Preparatory and Academy of the Pacific Rim, held lotteries for 6th grade entry. Records for Boston Collegiate are from 5th grade lotteries.

The raw lottery records typically include applicants names, dates of birth, contact information and other information used to define lottery groups, such as sibling status. The first five rows in Table A1 show the sample restrictions we impose on the raw lottery records. We exclude duplicate applicants and applicants listed as applying to the wrong entry grade. We also drop late applicants, out-of-area applicants, and sibling applicants, as these groups are typically not included in the standard lottery process. Imposing these restrictions reduces the number of lottery records from 9,256 to 8,851.

Lottery offers

In addition to the data described above, the lottery records also include information regarding offered seats. We used this information to reconstruct indicator variables for whether lottery participants received randomized offers. We make use of two sources of variation in charter offers, which differ in timing. The *initial offer* instrument captures offers made on the day of the charter school lottery. The *ever offer* instrument captures offers made initially or later, as a consequence of movement down a randomly sequenced waiting list. The pattern of instrument availability across schools and applicant cohorts is documented in Panel B of Appendix Table A1. In some years, all applicants eventually received offers, in which case only the initial offer instrument contributes to the analysis; these cases are listed as No Variation for the ever offer instrument. As documented in Table 2, initial and ever offer rates were 31 and 65 percent in our MCAS analysis sample, and these rates were similar in the samples for other outcomes.

SIMS Data

Data description

Our study uses SIMS data from the 2001-2002 school year through the 2011-2012 school year. Each year of data includes an October file and an end-of-year file. The SIMS records information on

demographics and schools attended for all students in Massachusetts public schools. An observation in the SIMS refers to a student in a school in a year, though there are some student-school-year duplicates for students that switch grades or programs within a school and year. The SIMS includes a unique student identifier known as the SASID, which is used to match students from other data sources as described below.

Coding of demographics and attendance

The SIMS variables used in our analysis include grade, year, name, town of residence, date of birth, sex, race, special education and limited English proficiency status, free or reduced price lunch and school attended. We constructed a wide-format data set that captures demographic and attendance information for every student in each year in which he or she is present in Massachusetts public schools. This file uses information from the longest-attended school in the first calendar year spent in each grade. Attendance ties were broken at random; this affects only 0.007 percent of records. Students classified as special education, limited English proficiency, or eligible for a free or reduced price lunch in any record within a school-year-grade retain that designation for the entire school-year-grade. The SIMS also includes exit codes for the final time a student is observed in the database. These codes are used to determine high school graduates and transfers.

We measure charter school attendance in 9th or 10th grade. A student is coded as attending a charter in his or her 9th or 10th-grade year when there is any SIMS record reporting charter attendance in that year. Students who attend more than one charter school within a year are assigned to the charter they attended longest.

MCAS Data

We use MCAS data from the 2001-2002 school year through the 2011-2012 school year. Each observation in the MCAS database corresponds to a students test results in a particular grade and year. The MCAS outcomes of interest are math and English Language Arts (ELA) tests in grade 10. We also use baseline tests taken prior to charter application, which are from 4th grade or 8th grade depending on a students application grade. The raw test score variables are standardized to have mean zero and standard deviation one within a subject-grade-year in Massachusetts. We also make use of scaled scores, which are used to determine whether students meet MCAS thresholds, which are Needs Improvement, Proficient, and Advanced. Unless otherwise noted, we only use the first test taken in a particular subject and grade.

AP and SAT Data

We use AP and SAT data files provided to the Massachusetts Department of Elementary and Secondary Education by College Board. The AP and SAT files include scores on all AP exams and SAT tests for graduation cohorts 2007 and 2012; for student who took the SAT more than once, the file includes only the score for the most recent exam. The AP and SAT files also include SASID identifiers, which are used to merge these outcomes with the SIMS database.

GED Data

Information on GED test-taking is provided by the Massachusetts Department of Elementary and Secondary Education's GED Office. This data includes testing dates and outcomes (pass

or fail) for students taking tests from 2002 to 2010. The GED information is merged to the SIMS administrative database by first, middle, and last name, and birth date. This procedure matched 70 percent of GED tests to records in the SIMS database.

NSC Data

Data on college outcomes comes from the National Student Clearinghouse (NSC) database, which captures enrollment for 94% of undergraduates in Massachusetts. We combine information from three separate searches of the NSC database:

- A 2010 search for all students in the SIMS database between 2002 and 2009 with projected graduation years earlier than 2014, assuming normal academic progress from the last observed grade and year (not restricted to students who graduated high school);
- A 2011 search of students who graduated from Massachusetts public high schools in the class of 2010;
- A 2012 search of all students who graduated from Massachusetts public high schools in the classes of 2003 through 2010;
- A 2013 search of students who graduated from Massachusetts public high schools in the classes of 2003 through 2012.

All students in our charter applicant sample were included in the 2010 NSC search, and Massachusetts high school graduates were included in multiple searches. College types are coded using the first attended college after the last date a student is observed in the SIMS. NSC searches were conducted using criteria like name and date of birth; the NSC files also include SASIDs, which are used to merge the college outcomes with the SIMS database.

Matching Data Sets

The MCAS, AP, SAT and NSC data files are merged to the master SIMS data file using the unique SASID identifier. The lottery records do not include SASIDs; these records are matched manually to the SIMS by name, application year and application grade. In some cases, this procedure did not produce a unique match. We accepted some matches based on fewer criteria where the information on grade, year and town of residence seemed to make sense.

Our matching procedure successfully located most applicants in the SIMS database. The sixth row of Panel A of Table A1 reports the number of applicant records matched to the SIMS in each applicant cohort. The overall match rate across all cohorts was 94 percent (8,342/8,851).

Once matched to the SIMS, each student is associated with a unique SASID; at this point, we can therefore determine which students applied to multiple schools in our lottery sample. Following the match, we reshape the lottery data set to contain a single record for each student. If students applied in more than one year, we keep only records associated with the earliest year of application. Our lottery analysis also excludes students who did not attend a Boston Public Schools (BPS) school at baseline, as students applying from private schools have lower follow-up rates. This restriction eliminates 22 percent of charter applicants. Of the remaining 4,700 charter applicants, 3,671 (78 percent) contribute a score to our MCAS analysis.

Table A1: Lottery Records

Panel A: Lottery Records

Projected Senior Year	2006	2007	2008	2009	2010	2011	2012	2013	All
Total number of records	600	450	940	883	1117	1533	1753	1980	9256
Excluding disqualified applicants	600	450	940	883	1117	1530	1753	1968	9241
Excluding late applicants	590	446	930	880	1117	1530	1733	1968	9194
Excluding applicants from outside of area	590	446	930	880	1114	1529	1733	1950	9172
Excluding siblings	570	437	905	864	1101	1482	1642	1850	8851
Excluding records not matched to the SIMS	509	419	858	816	1055	1395	1547	1743	8342
Reshaping to one record per student-year	437	419	632	594	799	1025	1100	1273	6279
Excluding repeat applications	437	419	629	589	778	1004	1028	1164	6048
In Boston schools at baseline	289	337	511	481	606	849	761	866	4700
Excluding applicants without 10th-grade ELA	232	267	415	378	482	664	570	663	3671

Panel B: Comparison of Ever Offer and Initial Offer Records by Schools and Cohorts

Application Year/School		Boston Preparatory	Academy of Pacific Rim	Boston Collegiate	City on a Hill	Codman Academy	Match
Entry grade		6	6	5	9	9	9
2002	Ever	Not Open	No Records	Yes	Not Oversubscribed	No Records	Yes
	Initial			Yes	Yes		Yes
2003	Ever	Not Open	No Records	Yes	No Records	No Records	Yes
	Initial			Yes			Yes
2004	Ever	Incomplete Records	No Records	Yes	Not Oversubscribed	Not Oversubscribed	Yes
	Initial			Yes	Yes	Yes	Yes
2005	Ever	Not Oversubscribed	Yes	Yes	Yes	Incomplete Records	Yes
	Initial	Yes	Yes	Yes	Yes		Yes
2006	Ever	Yes	Yes		Yes	Incomplete Records	Yes
	Initial	Yes	Yes		Yes		Yes
2007	Ever	Too Young for Follow-up			Yes	No Records	Yes
	Initial				Yes		Yes
2008	Ever				Not Oversubscribed	Yes	Yes
	Initial				Yes	Yes	Yes
2009	Ever				Yes	Yes	Yes
	Initial				Yes	Yes	Yes
N		208	178	265	1889	148	2330

Notes: Panel A summarizes the sample restrictions imposed for the lottery analysis. Disqualified applications are duplicate records and applications to the wrong grade. In Panel B, "Not Oversubscribed" indicates that every applicant received an offer in the relevant cohort. "Yes" means that lottery records with non-missing information on ever offer and initial offer were available, and that some applicants did not get offers. "Incomplete Records" indicates schools and years for which lottery records are inadequate to allow reliable coding of initial or ever offers. The last row shows the number of applicants to each school in the lottery sample excluding applicants without 10th-grade ELA (N = 3,671). Cohorts are too young for follow-up if they don't generate AP, SAT, high school graduation, or college-going outcomes in time for our study. For City on a Hill 2009 and Match 2008 applicants, we impute initial offer using 2008 City on a Hill and 2007 Match initial offer cutoffs. Starting in 2006, Academy of Pacific Rim has operating grades from 5 to 12.

Table A2: Grade 10 and Grade 12 Attrition

	Panel A: Observed 10th-Grade MCAS Scores and Grade 12 In MA Status				*Panel B: Attrition Differentials by Ever Offer and Initial Offer*					
	Either Math or ELA	ELA	Math	Grade 12 MA	ELA		Math		Grade 12 MA	
Projected Senior Year	Mean	Mean	Mean	Mean	Ever Offer	Initial Offer	Ever Offer	Initial Offer	Ever Offer	Initial Offer
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
2006	0.803	0.803	0.803	0.747	0.108	0.031	0.108	0.031	0.015	0.068
					(0.083)	(0.053)	(0.083)	(0.053)	(0.079)	(0.062)
2007	0.795	0.792	0.789	0.774	-0.038	-0.063	-0.034	-0.036	0.011	0.017
					(0.058)	(0.066)	(0.058)	(0.065)	(0.057)	(0.062)
2008	0.820	0.812	0.800	0.765	0.100	-0.034	0.072	-0.035	0.028	-0.011
					(0.064)	(0.043)	(0.066)	(0.045)	(0.070)	(0.050)
2009	0.794	0.786	0.771	0.763	-0.033	-0.061	-0.020	-0.048	-0.037	-0.050
					(0.042)	(0.042)	(0.044)	(0.043)	(0.044)	(0.043)
2010	0.799	0.795	0.785	0.766	0.036	-0.010	0.028	-0.015	-0.033	-0.040
					(0.044)	(0.040)	(0.045)	(0.041)	(0.046)	(0.042)
2011	0.786	0.782	0.760	0.729	-0.005	0.037	-0.005	0.046	0.012	0.029
					(0.031)	(0.032)	(0.032)	(0.033)	(0.033)	(0.035)
2012	0.756	0.749	0.742	0.607	0.037	-0.018	0.057	-0.009	-0.046	-0.059
					(0.052)	(0.036)	(0.052)	(0.037)	(0.057)	(0.041)
2013	0.768	0.766	0.752	-	-0.025	-0.013	-0.020	0.002	-	
					(0.035)	(0.032)	(0.036)	(0.033)		
All Cohorts	0.786	0.781	0.769	0.725	0.005	-0.009	0.007	-0.001	-0.011	-0.011
					(0.016)	(0.014)	(0.016)	(0.014)	(0.019)	(0.017)
N (All Cohorts)						4700		4700		3834

Notes: This table summarizes attrition for 10th-grade MCAS scores and 12th-grade enrollment status for charter school lottery applicants. Columns (1)-(3) show fractions of applicants with observed MCAS scores in each subject among those expected to take the test, assuming normal academic progress after the lottery. Column (4) shows the fraction of applicants with 12th-grade records in the SIMS administrative database. Columns (5)-(10) report coefficients from regressions of indicators for follow-up data on ever offer and initial offer dummies. Regressions also control for risk set dummies.

*significant at 10%; **significant at 5%; ***significant at 1%

Table A3: Lottery Estimates of Effects Excluding Each School

	Excl. School 1	Excl. School 2	Excl. School 3	Excl. School 4	Excl. School 5	Excl. School 6	All Schools
	Effect	Effect	Effect	Effect	Effect	Effect	Effect
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				Panel A: 10th- Grade MCAS			
Standardized ELA	0.398***	0.357***	0.401***	0.416***	0.392***	0.409***	0.411***
	(0.114)	(0.136)	(0.107)	(0.105)	(0.109)	(0.157)	(0.104)
Standardized Math	0.565***	0.701***	0.532***	0.556***	0.607***	0.479**	0.569***
	(0.131)	(0.159)	(0.122)	(0.121)	(0.121)	(0.189)	(0.120)
				Panel B: SAT Outcomes			
Took SAT	0.018	0.077	0.024	0.022	0.042	0.009	0.033
	(0.083)	(0.095)	(0.077)	(0.079)	(0.080)	(0.127)	(0.078)
SAT Composite (2400)	95.6**	183.2***	96.9**	99.6**	95.7**	66.6	102.8**
	(45.9)	(56.0)	(42.8)	(43.6)	(44.6)	(64.1)	(42.9)
				Panel C: AP Outcomes			
Took any AP	0.316***	0.412***	0.279***	0.291***	0.295***	0.085	0.287***
	(0.074)	(0.103)	(0.074)	(0.073)	(0.075)	(0.137)	(0.073)
Score 3 or Higher, any AP	0.089*	0.184***	0.091*	0.093*	0.085	0.021	0.096*
	(0.052)	(0.059)	(0.052)	(0.052)	(0.054)	(0.066)	(0.052)
				Panel D: High School Graduation Outcomes			
Four-year Graduation	-0.128**	-0.034	-0.129**	-0.133**	-0.117*	-0.190*	-0.125**
	(0.065)	(0.089)	(0.063)	(0.063)	(0.065)	(0.100)	(0.063)
Five-year Graduation	-0.010	-0.045	0.001	0.004	0.000	0.088	0.000
	(0.068)	(0.087)	(0.065)	(0.065)	(0.065)	(0.105)	(0.065)
				Panel E: College Enrollment Within 18 Months			
Any	0.102	0.045	0.115	0.115	0.115	0.305*	0.115
	(0.087)	(0.108)	(0.084)	(0.087)	(0.084)	(0.169)	(0.084)
Four-year	0.167**	0.098	0.173**	0.169**	0.173**	0.235	0.173**
	(0.081)	(0.106)	(0.079)	(0.081)	(0.079)	(0.150)	(0.079)

Notes: This table reports 2SLS estimates of the effects of charter attendance, excluding data from each of the six schools in the sample at a time. See Table 1 notes for outcome sample descriptions and Table 3 notes for detailed regression specifications. Sample size is excluded to avoid identifying individual schools.
*significant at 10%; **significant at 5%; ***significant at 1%

Table B1: Match From School Lottery Data to SIMS

Projected Senior Year	Number of applicants	Sample Mean	Ever Offer	Initial Offer
	(1)	(2)	(3)	(4)
2006	570	0.912	0.023	0.034
			(0.040)	(0.027)
2007	437	0.959	-0.017	-0.007
			(0.026)	(0.028)
2008	905	0.950	-0.007	-0.002
			(0.023)	(0.019)
2009	864	0.944	0.005	0.005
			(0.016)	(0.018)
2010	1101	0.959	0.034**	0.024**
			(0.017)	(0.012)
2011	1482	0.941	0.034***	0.043***
			(0.012)	(0.014)
2012	1642	0.942	0.032**	0.051***
			(0.016)	(0.011)
2013	1850	0.942	0.036***	0.026**
			(0.013)	(0.011)
All Cohorts	8851	0.944	0.025***	0.028***
			(0.006)	(0.006)

Notes: This table summarizes the match from the lottery records to the SIMS administrative data. The sample excludes disqualified applicants, late applicants, out-of-area applicants, and siblings. Columns (3) an (4) report coefficients from regressions of an indicator for a successful SIMS match on ever and initial offer dummies. The initial offer dummy is equal to one when a student is offered a seat in any of the charter schools immediately following the lottery, while the ever offer dummy is equal to one for students offered seats at any time. Year-specific regressions control for charter school dummies. Regressions for all cohorts controls for school-by-year dummies.

*significant at 10%; **significant at 5%; ***significant at 1%

Table B2: Lottery Estimates of Effects by Lottery Cohort

		Subsamples by Projected Senior Year		All
		Pre-2009	Post-2010	
		Effect	Effect	Effect
Outcomes		(1)	(2)	(3)
		Panel A: 10th- Grade MCAS		
Standardized ELA		0.111	0.533***	0.411***
		(0.196)	(0.117)	(0.104)
		1292	2379	3671
Standardized Math		0.450**	0.618***	0.569***
		(0.210)	(0.141)	(0.120)
	N	1278	2337	3615
		Panel B: SAT Outcomes		
Took SAT		0.153	-0.015	0.033
		(0.140)	(0.094)	(0.078)
	N	1134	1823	2957
SAT Composite (2400)		61.9	123.8***	102.8**
		(90.5)	(44.8)	(42.9)
	N	722	1175	1897
		Panel C: AP Outcomes		
Took any AP		0.345***	0.258***	0.287***
		(0.133)	(0.095)	(0.073)
Score 3 or Higher, any AP		0.137	0.079	0.096*
		(0.096)	(0.063)	(0.052)
	N	1134	1823	2957
		Panel D: High School Graduation Outcomes		
Four-year Graduation		0.002	-0.201***	-0.125**
		(0.116)	(0.070)	(0.063)
	N	1382	1823	3205
Five-year Graduation		0.072	-0.050	0.000
		(0.099)	(0.085)	(0.065)
	N	1382	1217	2599
		Panel E: College Enrollment Within 18 Months		
Any		0.090	0.047	0.063
		(0.120)	(0.084)	(0.072)
Four-year		0.207*	0.158*	0.170**
		(0.114)	(0.085)	(0.070)
	N	1382	1217	2599

Notes: This table reports 2SLS estimates of the effects of Boston charter attendance in two subsamples split by projected senior year 2009. See Table 3 notes for detailed regression specifications. Means are for non-charter attendees.

*significant at 10%; **significant at 5%; ***significant at 1%

Table B3: Peer Quality as the Endogenous Variable

	First Stage		Second Stage
	Ever Offer	Initial Offer	Effect
Subject	(1)	(2)	(3)
Standardized ELA	0.051*	-0.010	2.840**
	(0.026)	(0.019)	(1.325)
N	3660		
First-stage F	1.9		
Standardized Math	0.067**	-0.020	2.502**
	(0.027)	(0.020)	(1.009)
N	3615		
First-stage F	3.2		

Notes: This table reports first and second stage estimates of the effects of peer quality in 10th grade on MCAS test scores. The sample includes students projected to graduate in 2006 through 2013. The endogenous variable is peer quality in 10th-grade measured as the average baseline ELA (or math) scores for the peers in grade 10. The instruments are ever offer and initial offer dummies. Initial offer is equal to one when a student is offered a seat in any of the charter schools immediately following the lottery, while ever offer is equal to one for students offered seats at any time.All 2SLS regressions control for risk sets, 10th grade calendar year dummies, race, sex, special education, limited English proficiency, subsidized lunch status, and a female by minority dummy. Standard errors are clustered at the school-year level in 10th grade.
*significant at 10%; **significant at 5%; ***significant at 1%