TEACHING TO THE STUDENT: CHARTER SCHOOL EFFECTIVENESS IN SPITE OF PERVERSE INCENTIVES

Sarah R. Cohodes
Education Policy and Social Analysis
Columbia University
Teachers College
New York, NY 10027
cohodes@tc.columbia.edu

Abstract

Recent work has shown that Boston charter schools raise standardized test scores more than their traditional school counterparts. Critics of charter schools argue that charter schools create those achievement gains by focusing exclusively on test preparation, at the expense of deeper learning. In this paper, I test that critique by estimating the impact of charter school attendance on subscales of the Massachusetts Comprehensive Assessment System and examining them for evidence of score inflation. If charter schools are teaching to the test to a greater extent than their counterparts, one would expect to see higher scores on commonly tested standards, higher-stakes subjects, and frequently tested topics. Despite incentives to reallocate effort away from less frequently tested content to highly tested content, and to coach to item type, I find no evidence of this type of test preparation. Boston charter middle schools perform consistently across all standardized test subscales.

doi:10.1162/EDFP_a_00175

1. INTRODUCTION

Charter middle schools in Boston have obtained impressive test score results and strong reputations, resulting in hundreds of children on waitlists, hoping for a chance to enter one of these schools. According to *The Boston Globe* (2011), two Boston middle school charters are in the top ten middle schools in Massachusetts, as ranked by proficiency on the eighth grade state exam. Causal research based on lotteries confirmed the impressive test score results by showing charter school students who won the lottery and attended school outperformed those who did not win the lottery and did not attend (Abdulkadiroglu et al. 2009, 2011). These results are particularly important because they control for selection bias, countering the frequent criticism that charter schools "cream" certain kinds of students.

The mechanisms behind this large impact are unclear, however. Case studies and non-causal quantitative research suggest that long school days and years, low student–teacher ratios, coherent mission and curriculum, and other school characteristics may contribute to charter school success. On the other hand, another potential cause of the charter school effect is score inflation caused by test preparation activities. Score inflation is defined as "increases in scores that do not signal a commensurate increase in proficiency in the domain of interest" that the test is designed to assess (Koretz 2008, p. 34). Two potential causes of score inflation are strategic coaching of predictable characteristics of tests and reallocation of teaching effort to highly tested topics. If charter schools are engaging in these types of activities, their strong results may be due to score inflation, rather than an actual increase in students' comprehension. Currently, there is no quantitative evidence for or against the existence of score inflation at charter schools, but there is anecdotal evidence that charter schools are very test-aware.

The accountability system that charter schools face, which has additional accountability measures on top of No Child Left Behind (NCLB), incentivizes teachers to reallocate to highly tested content and to coach certain types of items in order to raise overall score, but not necessarily increase students' human capital. Using fine-grained data from Massachusetts, I investigate the Boston charter middle school effect more deeply to see if charter students are more successful than their counterparts in other Boston schools on all aspects of the Massachusetts Comprehensive Assessment System (MCAS) and if any of the gains can be explained by score inflation. If charter school students outpace their peers on all elements of the test (rarely tested standards as well as common standards and topics) on science, math, and English/language arts (ELA), and on all types of questions (multiple choice, short answer, and open response) then I will have no evidence of charter schools using test preparation to a greater extent than other schools in Boston.

This is the first study of charter schools that uses item-level information to disaggregate the test score effects in order to determine if charter schools are using test preparation to fuel their test score results. I present results for rarely tested content, including rarely tested curriculum standards, science, and less-emphasized topics to investigate reallocation, and results by item type to investigate coaching.

Although accountability pressure from the state rating system and public competition around test score results might induce teachers to utilize test preparation, I find no evidence of this. Charter school students have large gains on almost all components of MCAS exams, leading me to suggest that their success is not due to differential test preparation, in spite of perverse incentives that might encourage it. The results are robust to adjustments made for attrition and sample matching. Additionally, charter schools do not focus on children on the "bubble" of proficiency—instead, gains are magnified for the least academically prepared.

The organization of the paper is as follows. In section 2, I provide the background and context by describing the charter school impact research, reviewing the relevant details of prior work in Boston, and discussing score inflation. In section 3, I provide a theoretical framework. Section 4 describes the outcome measures, data, and sample. In section 5, I present my identification strategy and in section 6 my results. Section 7 addresses threats to validity. Section 8 concludes.

2. BACKGROUND AND CONTEXT

Charter School Impacts

Lottery-based studies of charter schools have generally found positive results of charter schools on academic achievement. These studies compare students who are offered a seat at a charter school through a lottery with those who are not offered a seat, meaning that the only difference between the two groups is the random offer of charter school attendance. Most of these lottery-based studies are small and city-specific, however. They are also limited to schools that are oversubscribed, which restricts their generalizability. Additionally, lottery-based results may overestimate the underlying citywide results if higher demand occurs at higher-quality schools. Hoxby, Murarka, and Kang's (2009) investigation of New York City charter schools found gains for charter school students in grades 4 through 8. Dobbie and Fryer (2011) focus on one charter school in the Harlem Children's Zone in New York City and found dramatic results, with the causal effect of charter school attendance on math achievement of around a standard deviation over the course of three years in middle school. Interestingly, a recent national lottery-based evaluation of 36 charter

schools found no significant effects overall, but significant gains for attendance at urban charter schools (Gleason et al. 2010).

In Boston, the causal effect of charter school attendance on middle school math scores is 0.4 standard deviations on the MCAS, and the effect on middle school ELA scores is 0.2 standard deviations on the MCAS, for each year of charter school attendance. The results for high schools are similar, though slightly smaller, with about a 0.2 standard deviation gain in both ELA and math per year (Abdulkadiroglu et al. 2009, 2011). The middle schools that participate in the Boston research, updated with additional years and newly opened schools, form the sample for this study.

When examining charter school impacts across Massachusetts, the Boston effect was muted (Angrist et al. 2011, 2013), but when the impacts were disaggregated by urbanicity, urban charters performed at similar levels to the Boston schools.

Results from broad comparisons between charter schools and traditional public schools are more mixed (CREDO 2009; Zimmer et al. 2009). The advantage of these studies is that they include students from both highly demanded and less demanded schools. They cannot adjust for the omitted variable bias inherent in comparing attendees at charters with those who may have never applied to a charter, however. A recent report finds that matching estimators can sometimes replicate lottery-based charter effects, but also finds that regression and fixed effects approaches are less successful at replication, perhaps another reason for the divergence in the literature (Fortson et al. 2012). Results from both lottery-based studies and other comparisons are limited in scope to the general impact of charter school attendance on test outcomes, not the details on these outcomes or the mechanisms behind the effects.

Beyond Charter School Test Impacts

Although the Boston results show large impacts for highly demanded charters, the authors cannot use the test score impacts to investigate the specific mechanisms that lead to the strong results. Quantitative research on charter schools is just beginning to investigate the mechanisms behind test score impacts. To date, charter schools have almost all been treated as a "black-box" where schools produce educational achievement by undetermined mechanisms. Hoxby, Murarka, and Kang's (2009) investigation of New York City charters attempts to peek into the black-box by associating some characteristics of charter schools with their success. They find that charter schools that have a longer school year/day, more minutes of instruction in core subjects, a "small rewards/small punishment discipline" system, a performance pay structure, and/or mission statements that emphasize academic success tend

to have greater test score success than charter schools without such policies (Hoxby, Murarka, and Kang 2009, p. V-5). These associations should not be interpreted causally, because although they use lottery-based estimates, the connection to characteristics is descriptive. Abdulkadiroglu et al. (2011) observe that Boston charter schools have much lower student–teacher ratios, younger teachers, and fewer in-subject licensed teachers, but again, these are descriptive, not causal, associations. Dobbie and Fryer (2013) find that positive charter school results are associated with "frequent teacher feedback, the use of data to guide instruction, high-dosage tutoring, increased instructional time, and high expectations" (p. 28). Angrist, Pathak, and Walters (2013) suggest that the positive impacts for urban Massachusetts charters are partially due to demographics and partially due to adherence to a "no excuses" philosophy. Recent case studies of five high-performing charter schools in Massachusetts, including three schools in this study, found those successful charter schools were characterized by a strong mission and a school culture dedicated to that mission; structures "that support student learning"; a focus on getting the "right" personnel; involved parents; and "classroom procedures that maximize[d] time on task and tightly link[ed] content to the Massachusetts curriculum framework" (Merseth et al. 2009, p. 228). The factors described here may be the determinants of charters' success on test scores.

Score Inflation

Another factor that could influence charter schools' MCAS success is test preparation. If test preparation is about "working more effectively, teaching more, [and] working harder" (Koretz 2008), then charter school test score gains might be due to an increase in these beneficial activities. But other, less benign, kinds of test preparation might be a factor in charters' MCAS success. If test preparation focuses on trivial knowledge of the test or reallocates resources to tested subjects, it could lead to score inflation. Why would potential score inflation in MCAS scores matter? If we think that MCAS outcomes are a measure of future success, not just an academic signpost during school, then test preparation and score inflation impede the inferences that we can draw from MCAS scores. To illustrate, when there is score inflation, a high math MCAS score would give a false impression of future success in math because the high score reflects test preparation rather than increased understanding of the content matter. Thus, if charter school effects are due to test preparation, the inference that they prepare students well for future math courses would be false.

Score inflation can be caused by four types of test preparation: "reallocation, alignment, coaching, [and] cheating" (Koretz 2008, p. 251). Cheating clearly

undermines the purpose of testing and leads to score inflation by increasing test scores with no parallel increase in learning (Jacob and Levitt 2003). Reallocation, alignment, and coaching are more ambiguous. Reallocation and alignment involve focusing resources and teaching on tested (or highly tested) topics and subjects, and cause score inflation when they draw efforts away from other parts of the curriculum that actually contribute to the underlying domain that the test is attempting to measure. Coaching occurs when teaching focuses on trivial aspects of the test, taking away time from meaningful content or focusing understanding of a topic in a specific format or organization. This causes score inflation by giving the impression that students comprehend the underlying domain of the test when actually they have become proficient in test taking methods or problems presented in a specific format.

Reallocation is likely widespread. With the implementation of NCLB, school districts across the nation are spending more time on tested subjects and less time on other subjects (McMurrer 2007; Nichols and Berliner 2007). Effects on test scores can be seen through gains on highly tested content but smaller or no gains on other content. Jacob (2005) finds that the implementation of high-stakes testing in Chicago led to gains on math items that are easy to teach or more common on the assessment, but no gains on other parts of the test, implying that reallocation to highly tested subjects caused the math gains. In Boston high school charters, Merseth (2010) sees impressive results on the MCAS but less-impressive results on college entrance exams, and she suggests that teaching at the schools may focus on material in line with the state exam but not the higher-order cognitive tasks tested on the SAT.[1]

As mentioned earlier, coaching involves teaching students about test-specific aspects of the assessment, rather than content. Some familiarity with test forms is important, but techniques that teach methods of guessing or standard responses to open response questions can inflate scores. Hamilton (2003) describes case studies and nationwide studies where teachers only distribute problems that parallel the formats on the test and change their instruction to mirror the format of state exams. Koretz (2008) describes methods like the process of elimination on multiple choice exams that, if taught, would increase students' test-taking skills but not the knowledge that tests are trying to assess.

Despite their successes, Boston charters are not immune to the accountability pressures that might induce test preparation and result in score inflation. Although widely perceived as successful schools because of their MCAS

1. Merseth reports 100 percent participation rates for taking the SAT at the three Boston charters for which she reports results (Academy of the Pacific Rim, Boston Collegiate Charter School, and MATCH). And although she reports the SAT results as less impressive than MCAS results, all three schools exceed the average Boston Public Schools (BPS) SAT score, even though only around 65 percent of Boston students take the SAT. The different compositions of who takes the test may account for the lack of a wider test score gap between the charters and BPS.

scores, NCLB's Adequate Yearly Progress (AYP) rankings identify most Boston area charters as needing improvement. In 2011, the only Boston charter middle schools not identified as in "improvement" or "corrective action" status under NCLB's standards for subgroups were Edward Brooke and Excel (Massachusetts Department of Elementary and Secondary Education 2011a). Boston charter schools, like many other schools in the nation, have the threat of NCLB sanctions as an incentive to do well on standardized exams. They are also under pressure to maintain high MCAS rankings that are widely trumpeted. Finally, charter schools must be renewed every five years in Massachusetts. Although renewals are not solely based on test scores, academic achievement is part of the renewal process. These triple pressures might encourage test preparation which would cause score inflation. In section 3, I describe in more detail how accountability systems can distort behavior to induce score inflation.

There is also evidence that Boston charter schools are very test conscious. Merseth et al.'s (2009) in-depth study of five charters, three of which are included in this study, indicates that teachers and administrators are very test-aware. Merseth et al. report that curriculum is carefully prepared to match with the Massachusetts Curriculum Frameworks. Teachers use publicly available MCAS items from prior years and they use assessments similar to the MCAS, and teachers constantly track their students' progress on content that is tested. Nevertheless, these test-aware behaviors need not lead to score inflation if the test preparation activities involve teaching more or better, rather than reallocating time to tested subjects or coaching on trivial details.

Implications

Boston middle school charters produce large gains for their students on the MCAS. The mechanisms behind Boston charter middle schools' success on the MCAS are unclear, however. They may be due to structural reasons, like longer school days and years, or low student–teacher ratios. They may be due to curriculum and planning efforts. Or they may be due to differential test preparation that results in score inflation. The purpose of this paper is to attempt to discover more details on this apparent success. I do so by disaggregating the MCAS scores so as to separate MCAS outcomes that are susceptible to test preparation from those that are not.

By determining if charters do not perform consistently across all measures of the test, I can look for evidence of test preparation. For instance, a particularly large effect on the multiple choice outcome, but little or no effect on the open response or short answer outcomes, might indicate coaching to item type. Similarly, a particularly large effect on standards that are tested most frequently, but little or no effect on standards tested rarely, might indicate

reallocation within mathematics to highly tested topics. For an additional check for this type of reallocation, I also exploit the fact that science is less emphasized in the accountability system and investigate whether science gains are similar in size to math and ELA gains.

3. THEORETICAL FRAMEWORK

Several articles (Jacob and Levitt 2003; Muralidharan and Sundararaman 2011; Barlevy and Neal 2012) have framed score inflation as a principal–agent problem. Accountability systems are put into place by state education agencies and the federal government in order to improve student achievement, but individual actors in the education system have an incentive to change their behavior so as to increase measured student achievement, and not necessarily students' underlying knowledge. Jacob and Levitt (2003) argue that accountability incentivizes cheating, and find overt cheating in 4 to 5 percent of Chicago classrooms. On the other hand, Muralidharan and Sundararaman (2011) argue that although a teacher incentive pay system in India might induce perverse responses, there is no evidence of such responses. Barlevy and Neal's (2012) theoretical incentive scheme also induces socially optimal responses.

Accountability systems may be formal, such as those prescribed by NCLB and state educational agencies. In Massachusetts, charters face an additional accountability system with five-year reviews from their authorizing agency (i.e., the state). In its reviews, the state requires charters to have "academic program success," "organizational viability," and be "faithful to the terms of the charter." Student performance is accounted for by the academic program requirements, which, prior to 2013, included MCAS proficiency or growth toward proficiency and AYP. The factors are also accounted for in the charter faithfulness requirement, as many charter school missions include an explicit focus on academic success. Although there are many other aspects of the reauthorization process, student academic achievement is quite important. Charter schools have similar pressures under the authorization process, the state accountability system, and NCLB, because they all rely on MCAS and proficiency levels or progress toward proficiency. Accountability systems may also be informal, such as pressure exerted by publicity around test scores and school rankings. This might be operationalized by parents with increased pressure on school leaders and teachers, or by parents moving their children out of lower-performing schools. It could also be enforced by principals, who have greater control over teacher hiring and firing than in traditional public schools.[2]

2. Note that some accountability pressures are greater for charter schools than traditional public schools—reauthorization and teacher personnel decisions. This does not mean that I cannot

To describe potential score inflation in Boston charters, I draw heavily on the model used by Muralidharan and Sundararaman (2011), with some modifications. Teachers (who may be encouraged in a particular direction by their school leaders, both of whom are agents in this context), under the various formal and informal accountability systems described here, can spend time on two topics, T_1, frequently tested content, and T_2, infrequently tested content. In the context of this study, math and ELA would be considered frequently tested content, whereas science is infrequently tested. Within subjects, some curriculum standards are tested frequently and others are not (for details, see section 4 on outcomes). Additional time spent on frequently tested topics is represented by t_1 and additional time on infrequently tested topics is represented by t_2.

Both frequently and infrequently tested topics contribute to the production of gains in human capital:

$$H = f_1(t_1) + f_2(t_2) + \varepsilon, \tag{1}$$

where H is unobserved gains in human capital, f_1 and f_2 are the marginal effects on human capital gains of time spent on t_1 and t_2, and ε is random error including all other factors that contribute to a student's gains in human capital. An education accountability system (the principal in the classic principal–agent problem) does not assign rewards and punishments to schools based on H, which is unobserved, but on an observable test score measure, Y. Test scores are also a function of time spent on frequently and infrequently tested content:

$$Y = g_1(t_1) + g_2(t_2) + \eta, \tag{2}$$

where g_1 and g_2 are, respectively, the marginal effects of time spent on t_1 and t_2 on test scores and η is random error including all other factors that contribute to a student's test score. The key feature of this analysis is that the causal charter school effect, measured by exploiting the charter school lottery, can be broken into score subscales representing t_1 and t_2. Unlike a traditional principal–agent problem, an educational accountability system does not offer an explicit wage based on Y, but it offers school-level rewards and punishments (which for charters may include closure), perhaps consequences for individual teachers depending on how a school leader uses test scores (increased professional development, increased evaluation, more freedom, job security, termination), and psychological comfort from meeting accountability

compare the two types of schools, however, only that charter school leaders and teachers might face even more incentives to teach to the test.

goals. These consequences do not directly affect salary or bonuses in most schools, but they do affect the non-pecuniary benefits of working in a school and can be considered part of a wage that is paid in utility.

Thus the accountability system offers a wage in utils, U, that is a function of the test score:

$$U = E[S] + E[Y] - E[C(t_1) + C(t2)], \tag{3}$$

where $E[S]$ is the expected utility of the teacher's salary, $E[Y]$ the expected utility or disutility of the non-pecuniary benefits of test scores (note that $E[Y]$) may be negative) measured in dollars, and $E[C(t_1) + C(t_2)]$ is the expected utility of the costs associated with the effort of teaching. When trying to find an optimal contract, the next step in this model is to determine a bonus associated with Y that induces optimal behavior. Here, these equations are sufficient to discuss how incentives from an accountability system may distort teacher behavior.

An increase (decrease) in test scores will increase (decrease) teacher utility. Additionally, if $g_1(t_1) > g_2(t_2)$ and $C(t_1) < C(t_2)$, reallocating time from infrequently tested items (T_2) to frequently tested items (T_1) will increase utility through two channels. First, when $g_1(t_1) > g_2(t_2)$ test scores will increase. Second, when $C(t_1) < C(t_2)$, costs will decrease. We expect $C(t_1) < C(t_2)$ if more curricular materials are provided for highly tested items and collaboration between teachers is easier for such items so that shifting time to t_1 lowers costs. Additionally, when T_1 is more emphasized on the test than T_2, it is likely that $g_1(t_1) > g_2(t_2)$ since additional t_1 will payoff on many items whereas additional t_2 will contribute to relatively few points on an exam. The most important question is whether f_1 has the same functional form as f_2 and both have non-decreasing returns. If both content areas influence gains in students' underlying human capital equally, it does not matter if teachers reallocate between T_1 and T_2. But if f_1 has decreasing marginal returns or if $f_2 > f_1$, reallocation to T_1 incentivized by the accountability system will lower human capital gains for students.

I argue that it is possible to separate Y into two components, Y_1 and Y_2, which in turn correspond to T_1 and T_2. For example, Y_1 measures performance on frequently tested content and Y_2 measures performance on infrequently tested content. I can then observe whether teachers respond to the incentive system that encourages them to increase Y by focusing on T_1, as measured by Y_1, or on T_2, as measured by Y_2.

Similar interpretations can be made if T_1 represents test preparation activities that increase Y but do not increase H (i.e., coaching) and T_2 represents other classroom activities that increase both Y and H.

4. OUTCOMES, DATA, AND SAMPLE

Outcomes

Each of the outcome measures attempts to highlight a different way that instruction, and thus test scores, can be manipulated or reallocated. The outcome data come from detailed information from individual-level MCAS results. Developed as a result of the 1993 Massachusetts Education Reform Act, which also allowed charters in the state, the MCAS has been the state's standardized test system since 1998. Since 2006, math and English/language arts have been tested in all of the relevant grade levels, and science is tested in eighth grade.

Using the detailed MCAS results, I added further information from the MCAS to create outcome variables that go beyond subject scores. Massachusetts makes public the question type, topic, difficulty, correct answer, and, since 2007, corresponding Massachusetts curriculum standard for each MCAS question (Massachusetts Department of Elementary and Secondary Education 2011b).[3] Indeed, the state even publishes the actual MCAS question. Thus, when I merged these data with item-level responses, I was able to identify each question that an individual student answered correctly and create outcome metrics based on subsets of questions.

The outcomes are grouped in three ways: rare standards, question type, and topic. Information on standards was first available for the spring 2007 MCAS, so outcomes using rarely tested standards have a restricted time range. I refer to this as the *rare standards sample*. Question type and question topic outcomes are available for all MCAS administrations, so I refer to these outcomes as covering the *full sample*. Each of the outcome measures is a standardized raw score of points in the category by subject, grade, and year. For reference, I also report outcomes for overall standardized score in each subject ("all items") in both the rare standards and full samples.

The MCAS exams consistently test each of the outcomes in similar proportions across years, making the frequently tested standards, question types, and topics on the test predictable. See table 1 for details. For instance, in math, multiple choice items always account for about 30 points, short answer items about 5 points, and open response items about 19 points (the test format changed slightly in 2010). Topic areas also follow a consistent pattern across years.

The MCAS outcomes used here make up about 80 percent of the MCAS exam; the other 20 percent includes items for equating and trial purposes, which are not reported or included in score calculation but are similar in type

3. Beginning in 2012, standards were categorized both by state standards and Common Core standards. Thus, I limit my sample to 2011 and prior years.

Table 1. Average Points Possible on MCAS Categories

	Math			ELA			Science
Subscale Outcome	**Grade 6 (1)**	**Grade 7 (2)**	**Grade 8 (3)**	**Grade 6 (4)**	**Grade 7 (5)**	**Grade 8 (6)**	**Grade 8 (7)**
Total Points Possible	54.0	54.0	54.0	52.0	52.0	52.0	54.0
Standard deviation (SD)	(11.8)	(12.1)	(12.6)	(8.3)	(8.6)	(8.6)	(10.1)
Rare Standards Sample							
Rare	8.4	8.2	10.2	1.6	1.8	3.8	5.6
SD	(2.4)	(2.2)	(2.7)	(1.0)	(0.9)	(2.3)	(1.6)
Somewhat Common	19.6	13.0	9.8	9.8	7.4	9.4	13.2
SD	(4.6)	(3.2)	(2.7)	(2.3)	(2.2)	(3.5)	(3.0)
Common	26.0	32.8	34.0	40.6	42.8	38.8	35.2
SD	(6.2)	(7.6)	(8.4)	(6.6)	(7.0)	(7.7)	(6.8)
Full Sample							
Multiple Choice	29.8	30.0	30.0	36.0	36.0	36.0	35.3
SD	(6.4)	(6.6)	(6.9)	(6.1)	(6.3)	(6.3)	(6.6)
Short Answer	5.3	5.3	5.3	-	-	-	-
SD	(1.5)	(1.6)	(1.7)				
Open Response	19.0	18.7	18.7	16.0	16.0	16.0	18.7
SD	(4.8)	(4.9)	(5.2)	(2.9)	(3.1)	(3.1)	(4.4)
Geometry	7.3	7.0	7.0	-	-	-	-
SD	(1.9)	(2.0)	(2.1)				
Measurement	7.1	7.0	7.0	-	-	-	-
SD	(2.2)	(2.1)	(2.3)				
Number Sense & Operations	17.6	13.8	14.0	-	-	-	-
SD	(4.3)	(3.5)	(3.7)				
Patterns, Algebra, & Relations	14.0	15.0	15.0	-	-	-	-
SD	(3.3)	(3.5)	(3.8)				
Data Analysis, Stat., & Prob.	8.0	11.2	11.0	-	-	-	-
SD	(2.2)	(2.8)	(2.7)				
Reading	-	-	-	45.6	47.2	46.0	-
SD				(7.4)	(7.9)	(7.7)	
Language and Literature	-	-	-	6.4	4.8	6.0	-
SD				(1.4)	(1.6)	(1.4)	
Earth and Space Science	-	-	-	-	-	-	13.5
SD							(2.9)
Life Science	-	-	-	-	-	-	14.0
SD							(3.0)
Physical Science	-	-	-	-	-	-	13.2
SD							(3.2)
Technology and Engineering	-	-	-	-	-	-	13.3
SD							(2.9)

Notes: For the test years that contribute to these averages, see table A.1. There is little variation across years in points possible in each category. Statewide standard deviations (SD) are underneath points possible. These are the standard deviations across the whole time period; yearly standard deviations are quite similar.

and topic to the common 80 percent of items (Massachusetts Department of Elementary and Secondary Education 2007). Thus schools and teachers can predict the format and topic of the MCAS each year. This predictability may lead to test preparation, as teachers can anticipate these features of each year's exam.

Rare Standards

For MCAS exams from spring 2007 to 2011, I determined which standards were given the most and least weight and divided the standards into terciles of rare standards, somewhat common standards, and common standards. This outcome allows me to assess whether charter school students do better on frequently assessed standards than on standards only assessed occasionally (to return to the theoretical model, T_1 and T_2). For instance, a question about Massachusetts standard 8.N.11:

> Determine when an estimate rather than an exact answer is appropriate and apply in problem situations.

was asked only once between 2007 and 2011. In contrast, questions about Massachusetts standard 8.M.3:

> Demonstrate an understanding of the concepts and apply formulas and procedures for determining measures, including those of area and perimeter/circumference of parallelograms, trapezoids, and circles. Given the formulas, determine the surface area and volume of rectangular prisms, cylinders, and spheres. Use technology as appropriate.

were asked twenty-one times in 2007–2011, five times in 2007, 2008, and 2009, four times in 2010, and twice in 2011 (Massachusetts Department of Elementary and Secondary Education 2000). Although the second standard likely encompasses more concepts than the first standard, it is difficult to determine whether one or the other is more important for overall understanding of mathematics.

Question Type

Question type outcomes are multiple choice, short answer, or open response. Only the mathematics exams have short answer questions. Multiple choice questions and short answer questions are each worth one point on the exam and open response questions are worth four points, with students scoring zero to four on each open response. The format of question types was only

changed once in the relevant period, with the math and science exams adding four multiple choice questions and subtracting one open response question in 2010. The format of the ELA exam was never changed in the relevant time period.

Topic

Question topic outcomes are specific to subject. For math they include geometry; measurement; number sense and operations; patterns, relations, and algebra; and data analysis, statistics, and probability. For ELA they include reading, and language and literature; and for science they include earth science, biology and life sciences, physical sciences, and technology and engineering. In math, number sense and operations and patterns, algebra, and relations are the most frequently tested topics, followed by data analysis, statistics, and probability. Geometry and measurement are tested the least in the middle school grades. In ELA, reading makes up the majority of the exam and language and literature items only make up a small portion of the test. Science topics are tested evenly. Across subjects, topic divisions are consistent across time. For instance, in ELA, reading accounts for 44 to 48 points on the exam and language and literature 4 to 8 points, depending on the test year.

Data

The data for this analysis come from statewide data sets provided by the Massachusetts Department of Elementary and Secondary Education, as well as lottery records collected from individual charter schools in Boston. The state provided data on students' demographic backgrounds, program participation, and school attendance for school years 2001–02 through 2010–11, and MCAS scores in math, ELA, writing, and science. I assigned students to their most attended school in each year, except that students who attended at least one charter school were assigned to the charter school even if it was not their most attended school. Thus, a student who attended a charter school for one month and a student who attended a charter school for one year were both assigned to the charter school for that year. Because I attribute a full year of attendance and the students' tests scores to the charter schools, no matter how long the student attended, my results based on years of attendance can be considered a lower bound on the effect of attending a year of charter school.[4]

In addition to the state data, lottery records were collected from each charter school for the main entry grade in each school (fifth or sixth grade). Lotteries

4. Results where students are assigned to their most attended school, without an exception for charter schools, are quite similar. As predicted, these results are larger, but only by about $0.01–0.03\sigma$, indicating that my conservative assignment rule makes little difference in the conclusions of this study.

were coded to identify students offered a seat at the charter school, to identify students who were never offered admission to the charter school, and to identify students who did not receive admissions offers randomly, such as students with sibling priority. Not all of the Boston area middle schools that admitted students for middle school entry in fifth or sixth grade were able to contribute records for lottery-based analysis. Two charter schools that contained middle school grades closed, two had insufficient lottery records, and two admitted the majority of their students at the kindergarten level. Appendix table A.10 includes details on school participation. The state data were combined with the lottery data through a matching process, which was then assembled into the analytic data set.

Because my focus is on middle school outcomes, I limit my data set to students with baseline information from the grade of application to a charter (either fourth or fifth grade) who entered charter school lotteries in spring 2002 to spring 2010. The available outcome scores vary with subject and grade level and are detailed in table A.1.

5. METHODS

I estimate the causal effect of attendance at a charter school on student achievement in the same way as Abdulkadiroglu et al. (2009, 2011). Because my intention is to disaggregate the charter school effect and determine if it is due to score inflation, however, the outcome measures are standardized components of the MCAS instead of subject scores, and are estimated separately by grade level, rather than pooled.

If all applicants who received an offer for a seat at a charter school attended that charter school and no applicants who did not receive an offer attended (i.e., if all applicants were all compliers) ordinary least squares regression using a variable representing the receipt of an offer would be sufficient to estimate the effect of charter school attendance on outcomes. In this case, however, because some applicants who received an offer to attend a charter school chose not to attend and a few students who lost the lottery ultimately attended a charter school,[5] I use an instrumental variables approach to estimate the causal effect of charter school attendance on the outcomes of interest.

The causal effect of a year of charter school attendance on a test score outcome component is represented in equation 4, the second stage of the instrumental variables estimation:

$$\gamma_{it} = \alpha_t + \sum_j \delta_j d_{ij} + \beta' X_i + \rho S_{it} + \varepsilon_{it}. \quad (4)$$

5. These students likely were on the waitlist and were offered seats late in the school year or entered a lottery for a grade or obtained sibling preference subsequent to the entry year.

Here, γ_{it} is the grade-level–specific test score based outcome of interest; S_{it} indicates the number of years of attendance, including repeated grades, at any charter school after the lottery at time t; X_i is a vector of student level demographic and test score control variables determined before the lottery; and ε_{it} is an error term. I also include a set of year-of-outcome fixed effects, α_t, and a set of lottery fixed effects $\sum_j \delta_j d_{ij}$, that represent the charter school lottery risk set.[6]

Because attendance at a charter school is not randomly assigned, I use the charter school lottery offer, which is randomly assigned, as an instrument for years of charter school attendance.[7] In equation 5, I represent the first stage:

$$S_{it} = \gamma_t + \sum_j \lambda_j d_{ij} + \kappa' X_i + \pi Z_{it} + \eta_{it}. \tag{5}$$

Here, S_{it} is estimated by X_i, a vector of student baseline demographic and test score control variables; $\sum_j \lambda_j d_{ij}$, a set of lottery fixed effects that represent the charter school lottery risk set; γ_t a set of year-of-outcome fixed effects; η_{it}, an error term; and the instrument, Z_i, which is a dummy variable that indicates if a charter school lottery applicant has received an offer to attend at least one charter school (sometimes referred to as winning the lottery).

In summary, π is the first stage effect, which in this case is the difference between the average number of years a student offered a seat at a charter school attends a charter school and the average number of years a student not offered a seat at a charter school attends a charter school. The causal effect of S_{it}, a year of charter school attendance, on γ_{it}, the test score component, is ρ, which I also refer to as the local average treatment effect. The treatment effect is local because it applies only to compliers, and because it is estimated using a partial compliance estimator, it can also be referred to as the average causal effect (Angrist and Imbens 1995). The associated reduced form or intent-to-treat effect, or effect of Z_i on γ_{it}, is found in an equation similar to equation 4 where Z_i is substituted for S_{it}. The coefficient of interest is ρ, which is the causal effect of a year of charter attendance, and is the ratio of the reduced form coefficient (difference in test based outcome between those offered a seat

6. The charter school lottery risk set for any given applicant is a dummy variable representing the charter school entry grade lottery or lotteries to which the applicant has applied. For instance, applicants applying only to charter school A would be in one risk set, applicants applying only to charter school B would be in another risk set, and applicants applying to both charter schools A and B would be in a third risk set. In Massachusetts, each charter runs its lottery independently, and students can apply to multiple charter schools. Because I only include lotteries for the main entry grades at schools, risk sets do not include late or repeat applications.

7. I exclude siblings, because they are guaranteed admission to charter schools. I also exclude late applicants and applicants from out-of-area, who are sent to the bottom of the waitlist. I also verify the lottery by comparing pretreatment covariates in table A.2, finding in a joint F-test that there is no difference between the groups.

and those not offered a seat) to the first stage coefficient (difference in years of attendance at a charter school between those offered a seat and those not offered a seat).

6. RESULTS

I fit the two-stage least squares (2SLS) model described earlier for each of my MCAS outcomes, such as math multiple choice score and science rare standards score.[8] I then inspect these outcomes to determine the composition of the middle school effects and its consistency or inconsistency across sections of the MCAS. I can also look for the effects of differential test preparation.

By comparing treatment effects across MCAS outcomes, I can see if the treatment effect for one or more of the outcomes has a larger response than the treatment effect on other outcome types. For the question type outcomes (multiple choice, short answer, and open response), differential success across outcomes may indicate that charter schools have coached to that question type to a greater extent than the other Boston schools attended by charter lottery losers. Likewise, by comparing treatment effects across standard frequency and subject topics, I can observe if results by standards frequency and topic are substantially different from each other. If charter school students are much more successful on common standards rather than typical standards, or certain frequently tested math topics rather than others, I would have evidence that charter schools are reallocating effort to teaching certain math standards and topics to a greater extent than other schools in Boston.

There are two important caveats. First, if charter schools are using coaching or reallocation with the same frequency as Boston Public Schools (BPS), I expect to see no difference in treatment effects due to coaching or reallocation. For instance, if both charter schools and other public schools are teaching students guessing strategies for multiple choice items, the subscore for multiple choice items would not stand out, even if test preparation occurred. Additionally, if charter schools are effective at coaching across all types of test questions, or are reallocating from untested subjects to all tested standards and topics, then I could not identify a coaching or reallocation effect, because no outcome would stand out. If charters are coaching a particular item type more than the comparison schools and more than other item types, however, I would expect to see a differential treatment effect on that item type subscale. Similarly, if charter schools are reallocating to common standards or more highly tested topics within a subject, I would expect to see higher scores on the more

8. Throughout this paper, I control for both baseline demographic characteristics and baseline test scores, which reduces the sample slightly. I focus on this model because it is the preferred model in prior work on Boston charters. Results are similar for a model that does not control for demographics or test scores and one that only controls for baseline demographics.

Table 2. First Stage Effect of a Lottery Win on Years of Attendance at a Charter School

	Math			ELA			Science
	Grade 6 (1)	Grade 7 (2)	Grade 8 (3)	Grade 6 (4)	Grade 7 (5)	Grade 8 (6)	Grade 8 (7)
Rare Standards Sample							
Years in Charter School	0.718***	1.018***	1.395***	0.717***	1.021***	1.397***	1.398***
	(0.071)	(0.090)	(0.141)	(0.070)	(0.089)	(0.140)	(0.141)
N	2,683	2,194	1,756	2,677	2,172	1,751	1,755
Full Sample							
Years in Charter School	0.699***	1.039***	1.402***	0.706***	1.049***	1.402***	1.405***
	(0.061)	(0.085)	(0.134)	(0.066)	(0.080)	(0.134)	(0.135)
N	3,317	2,373	1,891	2,987	2,488	1,889	1,890

Notes: This table reports coefficients on regressions predicting years spent in a charter using the offer of enrollment at a charter school. Each outcome cell is estimated by a separate regression. All regressions include baseline demographic controls, baseline test score controls, lottery risk sets (which are a set of dummies for the combination of schools applied to by year), and year of test and year of birth dummies. The sample is restricted to charter school applicants without sibling priority in the lottery, who attended a public or charter school in their year of application, and who have baseline demographic characteristics. Regressions use robust standard errors and are clustered by school by year.
***Significant at the 1% level.

frequently tested items and lower scores on the less frequently tested standards and topics.

First Stage

In table 2, I present the first stage results that show that the offer of a seat at a charter school does predict future attendance at charter schools. Results are similar across samples and subjects. By sixth grade, on average, students who are offered a seat through the charter lottery attend about 0.7 years more of charter school than students who did not receive an offer of a seat. By seventh grade, on average, students who are offered a seat through the charter lottery attend a charter for a full year more than students who did not receive an offer. By eighth grade, the difference is almost a year and a half.

The first stage estimate may be less than the total potential time a student could attend a charter for two reasons. First, only 70 percent of students who win the lottery at one of the oversubscribed charter schools attend a charter school. Second, a third of the students who did not win a seat through an oversubscribed lottery nonetheless attended some charter school for some time. These latter students could attend a charter by entering at a later grade, obtaining sibling preference, getting a spot off the waitlist late in the school year, or attending a charter not included in the lottery sample.

Table 3. Reduced Form Effect of a Lottery Win on MCAS Outcomes

	Math			ELA			Science
	Grade 6	Grade 7	Grade 8	Grade 6	Grade 7	Grade 8	Grade 8
Subscale Outcome	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
Rare Standards Sample							
All Items	0.340***	0.303***	0.348***	0.139***	0.242***	0.171***	0.410***
	(0.038)	(0.049)	(0.056)	(0.032)	(0.043)	(0.051)	(0.063)
Rare	0.404***	0.367***	0.280***	0.182***	0.116**	0.090	0.279***
	(0.046)	(0.059)	(0.060)	(0.060)	(0.051)	(0.059)	(0.060)
Somewhat Common	0.393***	0.286***	0.293***	0.148***	0.191***	0.151***	0.259***
	(0.044)	(0.048)	(0.058)	(0.037)	(0.046)	(0.057)	(0.061)
Common	0.243***	0.258***	0.355***	0.116***	0.241***	0.178***	0.440***
	(0.035)	(0.047)	(0.056)	(0.034)	(0.044)	(0.052)	(0.066)
N	2,683	2,194	1,756	2,677	2,172	1,751	1,755
Full Sample							
All Items	0.369***	0.326***	0.385***	0.122***	0.233***	0.183***	0.418***
	(0.036)	(0.046)	(0.057)	(0.031)	(0.039)	(0.048)	(0.060)
Multiple Choice	0.383***	0.356***	0.380***	0.129***	0.205***	0.157***	0.416***
	(0.038)	(0.050)	(0.054)	(0.030)	(0.039)	(0.045)	(0.063)
Short Answer	0.377***	0.309***	0.354***	-	-	-	-
	(0.043)	(0.053)	(0.063)	-	-	-	-
Open Response	0.278***	0.234***	0.342***	0.068	0.208***	0.190***	0.351***
	(0.035)	(0.042)	(0.060)	(0.044)	(0.050)	(0.067)	(0.057)
N	3,317	2,373	1,891	2,987	2,488	1,889	1,890

Notes: This table reports coefficients on regressions predicting MCAS outcomes using the offer of enrollment at a charter school. Each outcome cell is estimated by a separate regression, using subscales standardized in the statewide sample by subscale and grade. All regressions include baseline demographic controls, baseline test score controls, lottery risk sets (which are a set of dummies for the combination of schools applied to by year), and year of test and year of birth dummies. The sample is restricted to charter school applicants without sibling priority in the lottery, who attended a public or charter school in their year of application, and who have baseline demographic characteristics. Regressions use robust standard errors and are clustered by school by year.
***Significant at the 1% level; **significant at the 5% level.

Reduced Form and 2SLS

In table 3, I present the reduced form results for rare standards, science, and question type. These results show the effect of being offered a seat at an oversubscribed charter school on MCAS subscale outcomes. Recall that the outcomes are standardized subscores, so that a statistically significant reduced form effect can be interpreted as the additional standard deviations (σ) correct on the MCAS subscore that a student offered a seat at a charter school scores compared with students not offered a seat. In table A.3, I present

Table 4. 2SLS: Effect Attending a Charter School, Per Year of Attendance, on MCAS Outcomes

	Math			ELA			Science
Subscale Outcome	**Grade 6 (1)**	**Grade 7 (2)**	**Grade 8 (3)**	**Grade 6 (4)**	**Grade 7 (5)**	**Grade 8 (6)**	**Grade 8 (7)**
Rare Standards Sample							
All Items	0.474***	0.298***	0.250***	0.194***	0.237***	0.122***	0.293***
	(0.048)	(0.039)	(0.033)	(0.043)	(0.038)	(0.032)	(0.037)
Rare	0.563***	0.360***	0.200***	0.236***	0.113**	0.064	0.200***
	(0.061)	(0.051)	(0.035)	(0.083)	(0.048)	(0.040)	(0.039)
Somewhat Common	0.548***	0.281***	0.210***	0.206***	0.187***	0.108***	0.185***
	(0.058)	(0.037)	(0.033)	(0.050)	(0.043)	(0.037)	(0.041)
Common	0.339***	0.254***	0.254***	0.162***	0.236***	0.127***	0.314***
	(0.044)	(0.039)	(0.035)	(0.045)	(0.039)	(0.032)	(0.038)
N	2,683	2,194	1,756	2,677	2,172	1,751	1,755
Full Sample							
All Items	0.528***	0.313***	0.274***	0.173***	0.222***	0.131***	0.297***
	(0.050)	(0.036)	(0.033)	(0.042)	(0.033)	(0.030)	(0.036)
Multiple Choice	0.548***	0.343***	0.271***	0.182***	0.195***	0.112***	0.296***
	(0.053)	(0.040)	(0.032)	(0.039)	(0.033)	(0.028)	(0.038)
Short Answer	0.540***	0.297***	0.252***	-	-	-	-
	(0.059)	(0.044)	(0.037)	-	-	-	-
Open Response	0.398***	0.226***	0.244***	0.096	0.198***	0.135***	0.250***
	(0.050)	(0.034)	(0.038)	(0.061)	(0.046)	(0.044)	(0.036)
N	3,317	2,373	1,891	2,987	2,488	1,889	1,890

Notes: This table reports coefficients on regressions predicting MCAS outcomes using the offer of enrollment at a charter school. Each outcome cell is estimated by a separate regression, using subscales standardized in the statewide sample by subscale and grade. All regressions include baseline demographic controls, baseline test score controls, lottery risk sets (which are a set of dummies for the combination of schools applied to by year), and year of test and year of birth dummies. The sample is restricted to charter school applicants without sibling priority in the lottery, who attended a public or charter school in their year of application, and who have baseline demographic characteristics. Regressions use robust standard errors and are clustered by school by year.
***Significant at the 1% level; **significant at the 5% level.

results pooled across grade levels for outcomes that are similar across grade levels.[9]

The causal effects on MCAS outcomes of attending a year of charter school are simply the ratio of the reduced form coefficients in table 3 to the first stage coefficients in table 2. In table 4, I show the 2SLS results, or average causal effects, on MCAS subscale outcomes per year of attendance at a charter. Because the causal effects are per year of charter school attendance, the

9. When results are not grade-specific, pooled results show similar findings to the disaggregated results.

intention-to-treat effects in table 3 will be scaled up or down to be equivalent to a year of charter school. Thus the sixth grade 2SLS effects are larger than those reported in table 3. The seventh grade 2SLS effects are about the same, and the eighth grade 2SLS results are smaller than the corresponding reduced form results.

In table A.4 I also report in raw MCAS score points the mean outcome score for lottery applicants who did not win a seat in the lottery and those who did. The difference between the two means are roughly equivalent to the reduced form estimates. (The reduced form estimates also include control variables to increase statistical precision.) The mean scores give context to the causal effects that report in the tables in standard deviation units. On the overall scores, students offered a seat in the lottery tend to outscore their counterparts not offered a seat by 3.5–4 MCAS raw points in math, 0.5–2 MCAS raw score points in ELA, and 3–4 raw score points in science, depending on the particular sample. In ELA, the difference is only one multiple choice item on the test, but in math and science the difference is as large as 3–4 multiple choice items or the full score on an open response item. Because these overall gaps are spread across multiple subscales, and some subscales are only a few MCAS points themselves, differences in raw score points between offered and non-offered students will be smaller.

Rare vs. Common Standards

To examine whether charter schools are reallocating more than public schools from less frequently tested topics within each subject, table 3 presents results for the reduced form and table 4 for the 2SLS results of the charter school impact on rarely tested standards, somewhat common standards, and common standards.[10] There is no conclusive pattern. Within each subject, subscores are within 0.05σ to 0.2σ of each other. As a whole, results by standards are positive, significant, and fairly large for all but one subscale: rare items in eighth grade ELA. This single nonsignificant result may be due to chance, given the large number of outcomes I am testing, or it may be due to some reallocation away from rare standards in eighth grade ELA. But as a whole, the pattern across the standards outcomes does not suggest a pattern of reallocation away from the least frequently tested items.

This setup assumes that each MCAS is a weighted random draw of items, with items weighted toward common standards, and that the 2007–11 exams are similar in standards distribution to past exams. Teachers observe this

10. This sample is limited to MCAS years 2007–11 because the state only began making item level information available in 2007. In 2012, the state began transitioning to Common Core standards, so I limit my period of examination to the time where data are available and there is one consistent set of standards.

over time and would have the opportunity to focus on the most common standards. However, perhaps teachers only focus on last year's exam and then reallocate their time away from untested standards. To test for this, I create variables indicating items with standards not on last year's test and items with standards on last year's test. This is only possible in sixth and eighth grade math and eighth grade science, as seventh grade math and all years of ELA standards are tested on every MCAS. The sample for this analysis is also limited to MCAS 2008–11 administrations, because I need both item-level standards data (2007–11) and information about last year's exam (so 2007 cannot be included). I present results from this analysis in table A.5. Again, there is no consistent pattern across subscales, with charter school students outperforming comparison students on both standards that were not tested in the previous year and on standards that were tested in the previous year.

To return to the theoretical framework outlined in section 3, the content of commonly tested standards (or those on last year's test) would correspond to T_1 and the content related to rarely tested standards (or those not on last year's test) would correspond to T_2. I directly observe the MCAS scores related to this content, Y_1 and Y_2, respectively. Because the test score outcomes are of the same magnitude and significance level, I conclude that, in spite of incentives that may encourage differential test preparation, I do not have evidence of reallocation across standards.

Low vs. High Stakes Subjects

As shown in the previous section, I find no evidence of reallocation within subject content on the MCAS from frequently tested standards to less frequently tested standards. Schools and teachers may not be reallocating their efforts within a subject, however, but rather away from less-tested subjects toward highly tested subjects. Nationally, the Center on Education Policy reports school districts increasing instructional time on tested subjects and decreasing time on subjects like science, social studies, foreign languages, arts, and physical education since the implementation of NCLB (McMurrer 2007). Although I cannot directly compare instructional time, I can investigate whether charter schools in Boston have a similar impact on science as on math and ELA and, for the first time, present results on science for Boston charters.

Although science is tested in Massachusetts, it is tested only once in grades 6 through 8 and results from the test do not enter the calculation of AYP during the study time period.[11] Similarly, they are not emphasized in the public presentation of results—each year *The Boston Globe* publishes proficiency MCAS rankings by district and schools. The science results are in a panel far below

11. Massachusetts began including MCAS science scores in AYP calculations in 2012.

the math and ELA rankings (*The Boston Globe* 2011). Because charters do not face the same accountability pressure for science results, they might reallocate their efforts away from science toward math and ELA. If so, I would expect the effect on science of winning the lottery (table 3, column 7) and the average causal response of attending a charter school (table 4, column 7) "all items" scores to be much smaller in magnitude and potentially not significant. I find, however, that results for the 8th grade science MCAS are quite similar to the results for the 8th grade math MCAS. The 2SLS effect in the full sample is about a 0.25σ gain in math test scores and 0.29σ gain in science test scores, per year of attendance at a charter school. These gains are of similar size and are both significant at the 0.001 level. Thus, I find no conclusive evidence of reallocation away from science. Similar to the interpretation of the standards findings, my comparison of high versus low stakes subjects is represented in the theoretical framework where T_1 corresponds to math and ELA, and T_2 corresponds to science. I find similar test scores for each test type.

This finding is somewhat analogous to the findings from a recent evaluation of teacher incentives in India (Muralidharan and Sundararaman 2011). Like the pressures in Massachusetts from NCLB, which incentivize math and ELA but not other subjects, in Muralidharan and Sundararaman's experiment teachers were explicitly rewarded for student achievement in math and reading, but not in science or social studies. They found significant gains in all subjects—suggesting that teachers increase their efforts across all topics when they are facing incentives, that academic press on students transfers across subjects, or that there is spillover from highly incentivized subjects to non-incentivized subjects.

Multiple Choice vs. Open Response

The bottom panels of tables 3 and 4 present reduced form and 2SLS results for all question types. Investigating question type should allow me to see evidence of coaching by question type. For instance, if charter schools were coaching a particular strategy on open response questions more than traditional schools did, I would expect to see a higher relative score for open response questions than for other question types. It is not entirely clear which question type would benefit the most from coaching. Multiple choice items can be coached with test-taking techniques like the process of elimination or encouraging students to guess (there is no penalty for guessing on the MCAS). Open response items can be coached by encouraging students to write down any answer, instead of leaving the response blank, or to use key words to signal structure. If there is differential coaching across question types, however, perhaps because it is easier to coach to one item type, it could appear with different effect sizes

across question type. In this case, difficult to coach items would be represented by T_1 in the model and easy to coach items are represented by T_2.

In general, charter school students do just as well on each type of question as they do on the subject as a whole. For example, in sixth grade, the overall 2SLS effect on the math MCAS is 0.53σ and scores by question type are quite similar: multiple choice, 0.55σ; short answer, 0.54σ; and open response, 0.40σ. In one case (sixth grade ELA), the 2SLS effect for one question type is not significant, although there are significant results for the other question type: overall ELA gains of 0.17σ, multiple choice gains of 0.18σ, and a nonsignificant positive result of 0.10σ for open response. This exception may be due to chance (given the large number of outcomes I am examining, it's not surprising that one would not be significant), or it may be due to a lack of emphasis on writing in sixth grade. Either way, I still conclude that, for the most part, charter schools outperform their peers in traditional public schools on all question types and see no direct evidence of coaching to question type.[12]

Infrequently vs. Frequently Tested Topics

Reduced form results by MCAS topic are presented in table 5 and 2SLS results in table 6. Examining content topics is a similar exercise to examining rarely tested standards. Some topics are consistently tested less frequently—geometry and measurement in math and language and literature in ELA. If charter students perform less well on less frequently tested content areas, I would have evidence of reallocation within subject to more highly tested content areas.

I find, however, that unlike students in Chicago, where the introduction of high-stakes testing resulted in differential effects by question topic (Jacob 2005), charter school students do better than comparison students on all topics on the subject exams. Although there is some fluctuation in the magnitude of effects across topics and grades, all show strongly significant positive results. Therefore, although I cannot rule out reallocation within math topics to those more frequently tested on the MCAS, I have no evidence of it. If both charter schools and the schools that charter lottery losers attend are reallocating their teaching efforts within the math exam to comparable extents, I also would not be able to detect evidence of reallocation.

12. Another possibility is that charter school students have more interim assessments than their counterparts in traditional public schools and that this familiarity generates the success across all item types. I cannot directly test the number of interim assessments in the two sectors, as this is not reported in the data. However, BPS uses both required and teacher-generated formative assessments through Assessment Technology Incorporated, which exposes students to standardized testing in the traditional public school setting as well.

Table 5. Reduced Form, Additional Outcomes: Effect of a Lottery Win on MCAS Topics

	Math			ELA			Science
	Grade 6	Grade 7	Grade 8	Grade 6	Grade 7	Grade 8	Grade 8
Subscale Outcome	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Full Sample							
Geometry	0.375***	0.325***	0.366***	-	-	-	-
	(0.042)	(0.046)	(0.068)	-	-	-	-
Measurement	0.393***	0.351***	0.327***	-	-	-	-
	(0.038)	(0.050)	(0.056)	-	-	-	-
Number Sense & Operations	0.257***	0.197***	0.334***	-	-	-	-
	(0.042)	(0.041)	(0.056)	-	-	-	-
Patterns, Algebra, & Relations	0.280***	0.263***	0.320***	-	-	-	-
	(0.034)	(0.050)	(0.053)	-	-	-	-
Data Analysis, Statistics, & Probability	0.282***	0.278***	0.375***	-	-	-	-
	(0.037)	(0.048)	(0.058)	-	-	-	-
Reading	-	-	-	0.197***	0.183***	0.128**	-
	-	-	-	(0.040)	(0.041)	(0.050)	-
Language and Literature	-	-	-	0.098***	0.226***	0.183***	-
	-	-	-	(0.032)	(0.040)	(0.050)	-
Earth and Space Science	-	-	-	-	-	-	0.309***
	-	-	-	-	-	-	(0.062)
Life Science	-	-	-	-	-	-	0.432***
	-	-	-	-	-	-	(0.062)
Physical Science	-	-	-	-	-	-	0.453***
	-	-	-	-	-	-	(0.063)
Technology and Engineering	-	-	-	-	-	-	0.240***
	-	-	-	-	-	-	(0.054)
N	3,317	2,373	1,891	2,987	2,488	1,889	1,890

Notes: The notes for this table are the same as the notes for table 3, only the outcomes differ.
***Significant at the 1% level.

7. THREATS TO VALIDITY

Matching

Students offered a seat in a charter school lottery are more likely to be matched to the state database than students not offered a seat. This is likely due to lottery losers being more likely to enter private school. If these unmatched students are substantially higher performing than the matched lottery losers, however, their omission from the results would bias my findings upward. To address this possibility, I present results in table A.7 that include only applicants from the 2002 and 2009 spring lotteries, which do not have a significant difference in match rates between the offered and non-offered groups (table A.6). I only show grade 6 results because of small sample sizes for the higher grades.

Table 6. 2SLS, Additional Outcomes: Effect Attending a Charter School, Per Year of Attendance, on MCAS Topics

	Math			ELA			Science
	Grade 6	Grade 7	Grade 8	Grade 6	Grade 7	Grade 8	Grade 8
Subscale Outcome	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
Full Sample							
Geometry	0.537***	0.313***	0.261***	-	-	-	-
	(0.062)	(0.038)	(0.041)	-	-	-	-
Measurement	0.562***	0.338***	0.233***	-	-	-	-
	(0.055)	(0.042)	(0.033)	-	-	-	-
Number Sense & Operations	0.367***	0.190***	0.238***	-	-	-	-
	(0.061)	(0.036)	(0.035)	-	-	-	-
Patterns, Algebra, & Relations	0.401***	0.253***	0.228***	-	-	-	-
	(0.045)	(0.040)	(0.033)	-	-	-	-
Data Analysis, Statistics, & Probability	0.403***	0.268***	0.267***	-	-	-	
	(0.049)	(0.038)	(0.037)	-	-	-	
Reading	-	-	-	0.279***	0.175***	0.091***	-
	-	-	-	(0.058)	(0.039)	(0.034)	-
Language and Literature	-	-	-	0.138***	0.215***	0.131***	-
	-	-	-	(0.042)	(0.035)	(0.031)	-
Earth and Space Science	-	-	-	-	-	-	0.220***
	-	-	-	-	-	-	(0.038)
Life Science	-	-	-	-	-	-	0.307***
	-	-	-	-	-	-	(0.038)
Physical Science	-	-	-	-	-	-	0.322***
	-	-	-	-	-	-	(0.039)
Technology and Engineering	-	-	-	-	-	-	0.171***
	-	-	-	-	-	-	(0.036)
N	3,317	2,373	1,891	2,987	2,488	1,889	1,890

Notes: The notes for this table are the same as the notes for table 4, only the outcomes differ.
***Significant at the 1% level.

Although there is some volatility in the results, as a whole they are just as large or even larger than the findings for the full sample, leading me to conclude that differential match rates are not biasing the results.

Attrition

If students leave the sample at different rates based on their offer or lack of an offer of a seat at a charter school, the results may be biased if students who leave differ in unobserved ways from students who stay. Table 7 shows that there is no significant differential attrition between students offered and not offered a

Table 7. Attrition

	Math		ELA		Science	
	Proportion of Non-offered with MCAS	Difference	Proportion of Non-offered with MCAS	Difference	Proportion of Non-offered with MCAS	Difference
	(1)	(2)	(3)	(4)	(5)	(6)
Has 6th Grade Outcomes	0.874	−0.004	0.875	−0.006	-	-
		(0.007)		(0.007)		-
N	1,494	3,410	1,332	3,052	-	-
Has 7th Grade Outcomes	0.880	0.012	0.875	0.013	-	-
		(0.008)		(0.009)		-
N	915	2,396	1,014	2,544	-	-
Has 8th Grade Outcomes	0.859	0.002	0.861	−0.001	0.859	0.001
		(0.009)		(0.009)		(0.009)
N	752	1,920	705	1,913	752	1,920

Notes: This table reports coefficients on regressions of an indicator variable equal to one if the outcome test score is non-missing on an indicator variable equal to one if the student was offered a seat in the lottery. The regressions are separate for grade level of outcome. All regressions include baseline demographic controls, baseline test controls, lottery risk sets (which are a set of dummies for the combination of schools applied to by year), and year of birth dummies. The sample is restricted to charter school applicants without sibling priority in the lottery, who attended a public or charter school in their year of application, and who have baseline demographic characteristics. Standard errors are robust.

seat. In case there are unobserved patterns among attriters that could influence outcomes, I refit my results including attriters, by using baseline test scores as substitutes for missing middle grade outcomes (baseline math score is used for all math and science outcomes, and baseline ELA score is used for ELA outcomes). This model assumes that students with missing outcomes continue to perform at the same level as at baseline. In actuality, performance at the exact same level between baseline grade and middle school is unlikely, but it is a good proxy because test scores are strongly correlated across grades ($r \approx 0.75$). With baseline scores assigned for missing outcomes, the findings are essentially the same as those presented in section 6 (for brevity, in table A.8 I present only the 2SLS results). Because there is little to no difference between the original findings and the results with baseline test scores assigned to missing outcomes, I conclude that the findings are not biased by selective attrition.

Reallocation between Students

Instead of reallocating resources to highly tested areas in order to boost scores, charter schools may be reallocating resources to particular students to increase

test scores. Focusing on students for whom intervention is mostly likely to influence proficiency categorization could increase test scores due to differential treatment effects by student type. Several studies have found that schools and teachers focus on students who are on the verge of proficiency (which is the test score outcome used in AYP calculations), perhaps to the detriment of other students. Neal and Schanzenbach (2010) show differential test score increases for students in Chicago in the middle of the test score distribution, the so-called "bubble kids," and a case study from Texas demonstrates this is an explicit pattern in some schools (Booher-Jennings 2005).

In order to determine if charter schools are focusing on students on the verge of or just above proficiency to a greater degree than their traditional school counterparts, I include interaction terms in the model that estimates the effect of charter school attendance for students within four scaled score points of the baseline proficiency threshold in the baseline grade. For example, the proficiency threshold is 240, so students scoring 236 and 238 are considered near and underneath, respectively, the threshold in their baseline year, and students scoring 240 and 242 are considered near and above, respectively, the threshold in their baseline year.[13] This baseline definition attempts to both measure prior proficiency level in the way a school or teacher would when examining the records of individual students, and also to avoid concerns about endogeneity. I present interaction results only for grade 6 outcomes, because these are the closest to when prior proficiency is determined.[14]

Because Massachusetts AYP determinations are based on a state calculated Composite Performance Index (CPI) that also gives credit to some scores below proficiency, I also create "near" variables for each kink in the CPI calculation. CPI points are awarded as such: proficient or above (above 240 MCAS points)—100 CPI points; needs improvement high (230–238 MCAS points)—75 CPI points; needs improvement low (220–228 MCAS points)—50 CPI points; warn/fail high (210–218 MCAS points)—25 CPI points; and warn/fail low (200–208 MCAS points)—0 CPI points. Massachusetts also allows schools to achieve AYP through improvement, which involves a specific goal set for each school and subgroup. Improvement is also calculated using the CPI, with its kinked nature, however, which would again put the focus on students near thresholds rather than throughout the achievement distribution.

I investigate the interaction between years of attendance at a charter school and prior scores (table 8 for math and table 9 for ELA). To test whether overall score or specific place in the score distribution is relevant, I do this both for

13. The MCAS is scored in multiples of two, ranging from 200 to 280.
14. Results (not shown) are similar in seventh and eighth grades.

Table 8. 2SLS with Interactions: Math Effect of Attending a Charter School, Per Year of Attendance, on 6th Grade MCAS Outcomes with Interaction Terms

	Prior Score		Prior Near *Proficiency* Threshold		Prior Near *Needs Improvement* High Threshold		Prior Near *Needs Improvement* Low Threshold		Prior Near *Warning* Threshold	
	Main Effect	Interaction	Main Effect	Interaction	Main Effect	Interaction	Main Effect	Interaction	Main Effect	Interaction
Subscale Outcome	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**
Rare Standards Sample										
All Items	0.454***	−0.156***	0.488***	−0.093**	0.480***	−0.060	0.475***	−0.008	0.471***	0.244
	(0.046)	(0.042)	(0.050)	(0.043)	(0.048)	(0.041)	(0.049)	(0.054)	(0.048)	(0.163)
Rare	0.406***	−0.097**	0.574***	−0.079	0.563***	−0.009	0.562***	0.007	0.562***	0.092
	(0.056)	(0.046)	(0.062)	(0.052)	(0.062)	(0.040)	(0.062)	(0.056)	(0.061)	(0.154)
Somewhat Common	0.528***	−0.104*	0.564***	−0.111**	0.550***	−0.020	0.549***	−0.014	0.545***	0.247
	(0.068)	(0.057)	(0.060)	(0.043)	(0.059)	(0.046)	(0.059)	(0.065)	(0.058)	(0.174)
Common	0.417***	−0.158***	0.349***	−0.075	0.347***	−0.087*	0.339***	−0.008	0.336***	0.254*
	(0.044)	(0.041)	(0.046)	(0.047)	(0.044)	(0.045)	(0.044)	(0.052)	(0.044)	(0.153)
Full Sample										
All Items	0.497***	−0.174***	0.545***	−0.130***	0.536***	−0.074*	0.530***	−0.017	0.527***	0.174
	(0.046)	(0.043)	(0.052)	(0.043)	(0.050)	(0.039)	(0.050)	(0.051)	(0.050)	(0.126)
Multiple Choice	0.512***	−0.197***	0.562***	−0.103**	0.555***	−0.068	0.548***	0.003	0.547***	0.139
	(0.048)	(0.045)	(0.054)	(0.044)	(0.053)	(0.044)	(0.053)	(0.057)	(0.053)	(0.125)
Short Answer	0.514***	−0.142***	0.553***	−0.103*	0.551***	−0.110**	0.539***	0.003	0.538***	0.190
	(0.056)	(0.053)	(0.062)	(0.057)	(0.061)	(0.046)	(0.060)	(0.066)	(0.059)	(0.165)
Open Response	0.376***	−0.121***	0.418***	−0.155***	0.404***	−0.057	0.402***	−0.041	0.396***	0.193
	(0.048)	(0.043)	(0.051)	(0.048)	(0.049)	(0.043)	(0.050)	(0.055)	(0.050)	(0.125)

Notes: This table reports coefficients on regressions predicting test-score based outcomes using years spent in charter school and an interaction between years spent in a charter school and a prior test outcome as predicted by the offer of enrollment at a charter school and the offer interacted with the prior test outcome. The remaining notes are the same as those for table 4. Sample sizes are the same as those for sixth grade outcomes in table 4.
***Significant at the 1% level; **significant at the 5% level; *significant at the 10% level.

Table 9. 2SLS with Interactions: ELA Effect of Attending a Charter School, Per Year of Attendance, on 6th Grade MCAS Outcomes with Interaction Terms

	Prior Score		Prior Near *Proficiency* Threshold		Prior Near *Needs Improvement* High Threshold		Prior Near *Needs Improvement* Low Threshold		Prior Near *Warning* Threshold	
	Main Effect	Interaction	Main Effect	Interaction	Main Effect	Interaction	Main Effect	Interaction	Main Effect	Interaction
Subscale Outcome	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**
Rare Standards Sample										
All Items	0.181***	−0.059	0.195***	−0.007	0.202***	−0.074	0.190***	0.072	0.191***	0.280
	(0.040)	(0.036)	(0.045)	(0.036)	(0.042)	(0.052)	(0.042)	(0.063)	(0.043)	(0.221)
Rare	0.231***	−0.022	0.261***	−0.143***	0.234***	0.017	0.235***	0.018	0.235***	0.060
	(0.047)	(0.044)	(0.053)	(0.052)	(0.050)	(0.063)	(0.049)	(0.091)	(0.050)	(0.362)
Somewhat Common	0.193***	−0.059	0.208***	−0.016	0.208***	−0.021	0.195***	0.196**	0.205***	0.103
	(0.047)	(0.044)	(0.053)	(0.052)	(0.050)	(0.063)	(0.049)	(0.091)	(0.050)	(0.362)
Common	0.150***	−0.055	0.160***	0.011	0.171***	−0.090*	0.160***	0.032	0.158***	0.322
	(0.042)	(0.038)	(0.047)	(0.034)	(0.044)	(0.053)	(0.044)	(0.060)	(0.045)	(0.196)
Full Sample										
All Items	0.160***	−0.055	0.172***	0.005	0.182***	−0.083	0.171***	0.033	0.170***	0.211
	(0.039)	(0.035)	(0.044)	(0.036)	(0.041)	(0.051)	(0.041)	(0.064)	(0.042)	(0.179)
Multiple Choice	0.169***	−0.058*	0.185***	−0.018	0.188***	−0.050	0.181***	0.030	0.179***	0.277
	(0.038)	(0.034)	(0.041)	(0.039)	(0.039)	(0.056)	(0.039)	(0.072)	(0.040)	(0.191)
Open Response	0.514***	−0.142***	0.553***	−0.103*	0.551***	−0.110**	0.539***	0.003	0.538***	0.190
	(0.056)	(0.053)	(0.062)	(0.057)	(0.061)	(0.046)	(0.060)	(0.066)	(0.059)	(0.165)

Notes: This table reports coefficients on regressions predicting test-score based outcomes using years spent in charter school and an interaction between years spent in a charter school and a prior test outcome as predicted by the offer of an enrollment at a charter school and the offer interacted with the prior test outcome. The remaining notes are the same as those for table 4. Sample sizes are the same as those for sixth grade outcomes in table 4.
***Significant at the 1% level; **significant at the 5% level; *significant at the 10% level.

overall standarized score, and in a separate model, near each prior CPI relevant threshold: proficient, needs improvement high, needs improvement low, and warn/fail high. If charter schools are focusing on students on the bubble of proficiency (or another score threshold) to a larger extent than their traditional public school counterparts, I would expect the interaction terms for students in the prior year near the threshold category to have a significant positive contribution to the test score impacts (columns 4, 6, 8, and 10). This is the case for one of the math outcomes and one of the ELA outcomes (perhaps, given the large number of coefficients tested, due to chance). Instead, it appears that the charter school effect is largest across all math outcomes and two of the ELA outcomes for students with the lowest prior test scores (column 2). Thus, I find little evidence in test score outcomes that charters are focusing on students on the verge of proficiency or another score threshold at a rate greater than the schools that their counterparts attend. The charter schools are in fact most effective, at least in math, for the many students at the very bottom of the proficiency distribution.

8. CONCLUSION

This paper investigates the details of the large causal impacts of attendance on MCAS outcomes at highly demanded middle school charters in Boston. Despite an incentive structure that would seem to reward teachers and charter schools for focusing on certain aspects of MCAS tests, I find no evidence of test preparation in comparison with traditional public schools. The consistent results across all elements of the test provide no discernible evidence of more reallocation between rare and common standards, low and high stakes subjects, multiple choice and open response questions, and infrequently and frequently tested topics in charter schools compared with traditional public schools. These results remain substantively the same when baseline test scores are assigned to those with missing outcomes or when limited to the sample with the same match rate by offer status. Nor is there evidence that charter schools are focusing on bubble students at a greater rate than other schools in Boston. My analysis strategy cannot conclusively rule out inappropriate test preparation, especially if it is consistent across all aspects of the test or if it is comparable to the test preparation that comparison schools conduct. Nevertheless, the evidence I show here also aligns with recent work showing that Boston charter high school[15] students outperform their counterparts on SAT and advanced placement tests and are more likely to enroll in four-year colleges (Angrist et al., forthcoming). Follow-up work on the Harlem Children's Zone also finds positive outcomes on nonstate standardized test academic and

15. The sample overlap is quite small with the middle schools examined in this study, because few cohorts are currently old enough to observe these outcomes.

social outcomes (Dobbie and Fryer, forthcoming). Combined with this recent evidence from the literature, the lack of any evidence of test preparation in these findings is suggestive that charter school gains are due to building the human capital of their students, rather than just increasing test scores, in spite of incentives that encourage teaching to the test.

I am grateful to Carrie Conaway and the staff at the Massachusetts Department of Elementary and Secondary Education, and the Boston area charter schools for their generous access to data and their time and assistance. Sandy Jencks, Daniel Koretz, Richard Murnane, Lindsay Page, John Willett, and, especially, Joshua Goodman provided helpful comments. I also thank my charter team colleagues, Joshua Angrist, Susan Dynarski, Jon Fullerton, Thomas Kane, Parag Pathak, and Christopher Walters, the Center for Education Policy Research at Harvard University, and the School Effectiveness and Inequality Institute at MIT.

REFERENCES

Abdulkadiroglu, Atila, Joshua D. Angrist, Sarah R. Cohodes, Susan M. Dynarski, Jon Fullerton, Thomas J. Kane, and Parag Pathak. 2009. *Informing the debate: Comparing Boston's charter, pilot, and traditional schools*. Boston, MA: The Boston Foundation.

Abdulkadiroglu, Atila, Joshua Angrist, Susan Dynarski, Thomas J. Kane, and Parag Pathak. 2011. Accountability and flexibility in public schools: Evidence from Boston's charters and pilots. *Quarterly Journal of Economics* 126(2):699–748. doi:10.1093/qje/qjr017

Angrist, Joshua D., Sarah R. Cohodes, Susan M. Dynarski, Jon Fullerton, Thomas J. Kane, Parag A. Pathak, and Christopher R. Walters. 2011. *Student achievement in Massachusetts' charter schools*. Cambridge, MA: Center for Education Policy Research at Harvard University.

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):431–442. doi:10.1080/01621459.1995.10476535

Angrist, Joshua D., Parag A. Pathak, and Christopher R. Walters. 2013. Explaining charter school effectiveness. *American Economic Journal. Applied Economics* 5(4):1–27. doi:10.1257/app.5.4.1

Angrist, Joshua D., Sarah R. Cohodes, Susan M. Dynarski, Parag A. Pathak, and Christopher R. Walters. Forthcoming. Stand and deliver: Effects of Boston's charter high schools on college preparation, entry, and choice. *Journal of Labor Economics*.

Barlevy, Gadi, and Derek Neal. 2012. Pay for percentile. *American Economic Review* 102(5):1805–1831. doi:10.1257/aer.102.5.1805

Booher-Jennings, Jennifer. 2005. Below the bubble: "Educational triage" and the Texas accountability system. *American Educational Research Journal* 42(2):231–268. doi:10.3102/00028312042002231

The Boston Globe. 2011. *2011 MCAS results*. Available www.boston.com/news/special/education/mcas/scores11/. Accessed 10 September 2014.

Center for Research on Education Outcomes (CREDO). 2009. *Multiple choice: Charter school performance in 16 states*. Stanford, CA: CREDO.

Dobbie, Will, and Roland G. Fryer. 2011. Are high quality schools enough to reduce social disparities? Evidence from the Harlem Children's Zone. *American Economic Journal. Applied Economics* 3(3):158–187. doi:10.1257/app.3.3.158

Dobbie, Will, and Roland Fryer. 2013. Getting beneath the veil of effective schools: Evidence from New York City. *American Economic Journal. Applied Economics* 5(4):28–60. doi:10.1257/app.5.4.28

Dobbie, Will, and Roland G. Fryer. Forthcoming. The medium-term impacts of high-achieving charter schools on non-test score outcomes. *Journal of Political Economy*.

Fortson, Kenneth, Natalya Verbitsky-Savitz, Emma Kopa, and Philip Gleason. 2012. *Using an experimental evaluation of charter schools to test whether nonexperimental comparison group methods can replicate experimental impact estimates*. NCEE Technical Methods Report 2012–4019. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Gleason, Phillip, Melissa Clark, Christina Clark Tuttle, and Emily Dwoyer. 2010. *The evaluation of charter school impacts: Final report, NCEE 2010–4029*. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Hamilton, Laura. 2003. Assessment as a policy tool. *Review of Research in Education* 27:25–68.

Hoxby, Carolyn 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.

Jacob, Brian A. 2005. Accountability, incentives and behavior: The impact of high-stakes testing in the Chicago public schools. *Journal of Public Economics* 89(5–6):761–796. doi:10.1016/j.jpubeco.2004.08.004

Jacob, Brian A., and Steven D. Levitt. 2003. Rotten apples: An investigation of the prevalence and predictors of teacher cheating. *Quarterly Journal of Economics* 118(3):843–877. doi:10.1162/00335530360698441

Koretz, Daniel M. 2008. *Measuring up: What educational testing really tells us*. Cambridge, MA: Harvard University Press.

Massachusetts Department of Elementary and Secondary Education. 2000. *Mathematics curriculum framework November 2000*. Available www.doe.mass.edu/frameworks/math/2000/toc.html. Accessed 7 October 2014.

Massachusetts Department of Elementary and Secondary Education. 2007. *MCAS 2007 technical report*. Available www.mcasservicecenter.com/documents/MA/Technical%20Report/TechReport_2007.htm. Accessed 7 October 2014.

Massachusetts Department of Elementary and Secondary Education. 2011a. *2011 lists of Massachusetts schools and districts by NCLB accountability status and accountability and assistance level*. Available www.doe.mass.edu/apa/accountability/2011/improvement.pdf. Accessed 7 October 2014.

Massachusetts Department of Elementary and Secondary Education. 2011b. *2011 Item by item results for grade 06 mathematics*. Available http://profiles.doe.mass.edu/mcas/mcasitems2.aspx?grade=06&subjectcode=MTH&linkid=9&orgcode=00000000&fycode=2011&orgtypecode=0&. Accessed 7 October 2014.

McMurrer, Jennifer. 2007. *NCLB year 5: Choices, changes, and challenges: Curriculum and instruction in the NCLB era*. Washington, DC: Center on Education Policy.

Merseth, Katherine K., Kristy Cooper, John Roberts, Mara Casey Tieken, Jon Valant, and Chris Wynne. 2009. *Inside urban charter schools: Promising practices and strategies in five high-performing schools*. Cambridge, MA: Harvard Education Press.

Merseth, Katherine K. 2010. High-performing charter schools: Serving two masters? In *Hopes, fears, & reality: A balanced look at American charter schools in 2009*, edited by Robin J. Lake, pp. 27–37. Seattle, WA: CRPE, University of Washington-Bothell.

Muralidharan, Karthik, and Venkatesh Sundararaman. 2011. Teacher performance pay: Experimental evidence from India. *Journal of Political Economy* 119(1):39–77. doi:10.1086/659655

Neal, Derek, and Diane Whitmore Schanzenbach. 2010. Left behind by design: Proficiency counts and test-based accountability. *Review of Economics and Statistics* 92 (2):263–283.

Nichols, Sharon Lynn, and David C. Berliner. 2007. *Collateral damage: How high-stakes testing corrupts America's schools*. Cambridge, MA: Harvard Education Press.

Zimmer, Ron, Brian Gill, Kevin Booker, Stephane Lavertu, Tim R. Sass, and John Witte. 2009. *Charter schools in eight states: Effects on achievements, attainment, integration, and competition*. Santa Monica, CA: Rand Corporation.

APPENDIX: ADDITIONAL DATA

Table A.1. Outcome Years

	Math			ELA			Science
	Grade 6 (1)	**Grade 7 (2)**	**Grade 8 (3)**	**Grade 6 (3)**	**Grade 7 (4)**	**Grade 8 (5)**	**Grade 8 (6)**
Rare Items Sample	2007–2011	2007–2011	2007–2011	2007–2011	2007–2011	2007–2011	2007–2011
Full Sample	2004–2011	2006–2011	2006–2011	2006–2011	2006–2011	2006–2011	2006–2011

Notes: Years indicate the spring of the school year, when the MCAS is administered. Information on the standards associated with each item was first published in 2007, thus the limited years for the rare items sample. The seventh grade math, sixth grade ELA, and eighth grade ELA MCAS exams were administered for the first time in spring 2006. The sixth and eighth grade math, seventh grade ELA, and eighth grade science MCAS exams were administered in years prior to those listed, however the first students that participated in the lotteries in the sample take the exam in the years noted.

Table A.2. Covariate Balance between Charter Applicants Offered a Seat and Not Offered a Seat in Charter School Lotteries

	Rare Items Sample		Full Sample	
	Difference (Offered–Not Offered)		Difference (Offered–Not Offered)	
	Coef. (1)	S.E. (2)	Coef. (3)	S.E. (4)
Latino/a	0.048***	(0.017)	0.044***	(0.015)
African-American	−0.038**	(0.019)	0.037**	(0.017)
White	−0.006	(0.015)	−0.006	(0.013)
Asian	−0.001	(0.006)	−0.001	(0.005)
Female	−0.021	(0.020)	−0.005	(0.019)
Free or Reduced Price Lunch	0.028	(0.018)	0.017	(0.017)
Special Education	0.004	(0.015)	0.000	(0.014)
English Language Learner	0.016	(0.012)	0.014	(0.010)
Baseline Standardized Math Score	−0.018	(0.040)	−0.018	(0.037)
Baseline Standardized ELA Score	−0.045	(0.038)	−0.033	(0.035)
Sample Size	3,392		4,036	
p-value from F-test	0.206		0.373	

Notes: This table reports coefficients on regressions of the variable indicated in each row on an indicator variable equal to one if the student was offered a seat at a charter through the lottery. The sample is restricted to charter school applicants without sibling priority in the lottery, who attended a public or charter school in their year of application, and who have baseline demographic characteristics and test scores. All regressions include lottery risk sets (which are a set of dummies for the combination of schools applied to by year), and year of baseline and year of birth dummies. Regressions use robust standard errors. F-tests are for the null hypothesis that the coefficients on winning the lottery in all regressions are all equal to zero. These tests statistics are calculated for the subsample that has non-missing values for all variables tested. Students must have at least one MCAS outcome to be included in the table.
***Significant at the 1% level; **significant at the 5% level.

Table A.3. Sixth, Seventh, and Eighth Grades Combined

	Math			ELA		
	FS	**RF**	**2SLS**	**FS**	**RF**	**2SLS**
Subscale Outcome	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Rare Standards Sample						
All Items	0.986***	0.327***	0.332***	0.988***	0.181***	0.183***
	(0.064)	(0.032)	(0.030)	(0.064)	(0.029)	(0.028)
Rare	0.986***	0.357***	0.363***	1.043***	0.129***	0.124***
	(0.064)	(0.037)	(0.035)	(0.068)	(0.034)	(0.032)
Somewhat Common	0.986***	0.331***	0.336***	0.988***	0.162***	0.164***
	(0.064)	(0.033)	(0.031)	(0.064)	(0.030)	(0.030)
Common	0.986***	0.274***	0.277***	0.988***	0.173***	0.175***
	(0.064)	(0.031)	(0.028)	(0.064)	(0.030)	(0.028)
N		6,633			6,600	
Full Sample						
All Items	0.975***	0.358***	0.367***	0.994***	0.176***	0.177***
	(0.060)	(0.031)	(0.030)	(0.060)	(0.028)	(0.026)
Multiple Choice	0.975***	0.373***	0.383***	0.994***	0.163***	0.164***
	(0.060)	(0.032)	(0.031)	(0.060)	(0.027)	(0.026)
Short Answer	0.975***	0.349***	0.359***	-	-	-
	(0.060)	(0.035)	(0.033)	-	-	-
Open Response	0.975***	0.277***	0.284***	0.994***	0.148***	0.149***
	(0.060)	(0.030)	(0.029)	(0.060)	(0.035)	(0.034)
N		7,581			7,364	

***Significant at the 1% level.

Table A.4. Outcome Means in Raw Score Points

	Not Offered a Seat in the Charter Lottery							Offered a Seat in the Charter Lottery						
	Math			ELA			Science	Math			ELA			Science
Subscale Outcome	Grade 6 (1)	Grade 7 (2)	Grade 8 (3)	Grade 6 (4)	Grade 7 (5)	Grade 8 (6)	Grade 8 (7)	Grade 6 (8)	Grade 7 (9)	Grade 8 (10)	Grade 6 (11)	Grade 7 (12)	Grade 8 (13)	Grade 8 (14)
Rare Standards Sample														
All Items	34.74	34.09	33.20	34.26	34.43	35.20	27.77	38.25	37.58	37.15	35.24	36.29	36.29	31.05
Rare	5.68	5.32	6.70	1.21	1.40	2.38	3.44	6.07	5.96	7.48	1.13	1.40	2.36	3.68
Somewhat Common	12.41	8.40	6.96	6.63	4.83	5.42	6.63	14.21	9.12	7.60	7.20	5.19	5.49	7.13
Common	16.64	20.37	19.54	26.43	28.20	27.41	17.70	17.97	22.50	22.07	26.91	29.70	28.44	20.24
Full Sample														
All Items	34.33	33.90	32.79	34.20	34.47	35.12	32.79	37.94	37.47	37.09	34.99	36.21	36.30	37.09
Multiple Choice	20.85	19.81	20.39	25.58	26.11	26.18	20.94	22.68	21.85	22.71	26.14	27.07	26.90	23.27
Short Answer	3.20	3.66	3.25	1.00	1.00	1.00	1.00	3.60	4.10	3.74	1.00	1.00	1.00	1.00
Open Response	10.28	10.42	9.15	8.62	8.36	8.94	6.88	11.66	11.52	10.63	8.84	9.14	9.40	7.97
Geometry	4.24	4.28	3.79	-	-	-	-	4.78	4.77	4.44	-	-	-	-
Measurement	4.18	3.81	3.88	-	-	-	-	4.65	4.16	4.48	-	-	-	-
Num. Sense & Operations	11.66	8.33	8.43	-	-	-	-	12.93	9.42	9.44	-	-	-	-
Patterns, Alg., & Relations	9.45	9.94	9.50	-	-	-	-	10.25	10.74	10.65	-	-	-	-
Data Analysis, Stat., & Prob.	4.80	7.53	7.20	-	-	-	-	5.34	8.38	8.08	-	-	-	-

Table A.4. Continued.

	Not Offered a Seat in the Charter Lottery							Offered a Seat in the Charter Lottery						
	Math			ELA			Science	Math			ELA			Science
Subscale Outcome	Grade 6 (1)	Grade 7 (2)	Grade 8 (3)	Grade 6 (4)	Grade 7 (5)	Grade 8 (6)	Grade 8 (7)	Grade 6 (8)	Grade 7 (9)	Grade 8 (10)	Grade 6 (11)	Grade 7 (12)	Grade 8 (13)	Grade 8 (14)
Reading	-	-	-	29.95	30.97	31.07	-	-	-	-	30.67	32.69	32.11	-
Language and Literature	-	-	-	4.25	3.49	4.05	-	-	-	-	4.32	3.52	4.19	-
Earth and Space Science							6.96							7.74
Life Science	-	-	-	-	-	-	7.79	-	-	-	-	-	-	8.83
Physical Science	-	-	-	-	-	-	6.47	-	-	-	-	-	-	7.59
Tech. and Engineering	-	-	-	-	-	-	6.59	-	-	-	-	-	-	7.08

Notes: This table reports the mean outcome for students who did not win a seat in the charter school lottery (columns 1–8) and students that did win a seat in the lottery (columns 8–14), in raw score MCAS points. The difference in means roughly corresponds to the reduced form estimates in tables 3 and 5.

Table A.5. 2SLS on Standards Categorized by Last Year's Test

	Math		Science
	6th	8th	8th
Subscale Outcome	**(1)**	**(2)**	**(3)**
Last Year's Standards Sample (2008–2011)			
All Items	0.489***	0.219***	0.269***
	(0.051)	(0.032)	(0.039)
Standards Not on Last Year's Test	0.545***	0.126***	0.220***
	(0.068)	(0.039)	(0.044)
Standards on Last Year's Test	0.462***	0.224***	0.276***
	(0.049)	(0.033)	(0.039)
N	2,276	1,596	1,595

Notes: The notes for this table are the same as those for table 4, with different outcomes, defined by whether or not a standard appeared on last year's test. Seventh grade math and all grades of ELA tested for each standard in almost every test administration, so it is impossible to create these outcomes for those grades and subjects.
***Significant at the 1% level.

Table A.6. 2SLS with Imputed Outcomes for Attriters: Effect Attending a Charter School, Per Year of Attendance, on MCAS Outcomes

	Math			ELA			Science
	Grade 6	Grade 7	Grade 8	Grade 6	Grade 7	Grade 8	Grade 8
Subscale Outcome	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
Rare Standards Sample							
All Items	0.465***	0.288***	0.243***	0.184***	0.226***	0.121***	0.282***
	(0.046)	(0.038)	(0.034)	(0.042)	(0.038)	(0.033)	(0.037)
Rare	0.407***	0.349***	0.195***	0.164***	0.111**	0.064	0.193***
	(0.056)	(0.050)	(0.037)	(0.062)	(0.047)	(0.040)	(0.039)
Somewhat Common	0.528***	0.273***	0.206***	0.194***	0.177***	0.106***	0.177***
	(0.069)	(0.036)	(0.034)	(0.048)	(0.042)	(0.037)	(0.041)
Common	0.431***	0.246***	0.248***	0.153***	0.225***	0.126***	0.303***
	(0.045)	(0.037)	(0.036)	(0.044)	(0.040)	(0.033)	(0.038)
N	2,963	2,366	1,951	2,928	2,372	1,949	1,951
Full Sample							
All Items	0.526***	0.306***	0.268***	0.167***	0.218***	0.131***	0.287***
	(0.049)	(0.036)	(0.034)	(0.041)	(0.034)	(0.030)	(0.036)
Multiple Choice	0.541***	0.335***	0.264***	0.175***	0.191***	0.111***	0.285***
	(0.052)	(0.038)	(0.033)	(0.038)	(0.033)	(0.028)	(0.038)
Short Answer	0.535***	0.295***	0.248***	-	-	-	-
	(0.059)	(0.043)	(0.037)	-	-	-	-

Table A.6. Continued.

	Math			ELA			Science
	Grade 6	Grade 7	Grade 8	Grade 6	Grade 7	Grade 8	Grade 8
Subscale Outcome	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Open Response	0.396***	0.223***	0.238***	0.096	0.199***	0.139***	0.243***
	(0.049)	(0.034)	(0.039)	(0.061)	(0.048)	(0.047)	(0.036)
N	3,561	2,536	2,086	3,237	2,688	2,087	2,086

Notes: The notes for this table are the same as those for table 4, except here baseline scores are used as the outcome for students missing outcome data.
***Significant at the 1% level; **significant at the 5% level.

Table A.7. Match from Lottery Records to SIMS

		Fraction with SIMS Match			
Lottery Cohort	Number of Records (1)	Total (2)	Offered (3)	Not Offered (4)	Offered > Not Offered? (5)
2002	295	0.908	0.934	0.859	Yes
2003	302	0.861	0.873	0.804	No
2004	300	0.887	0.930	0.848	Yes
2005	678	0.934	0.968	0.883	Yes
2006	837	0.952	0.968	0.919	Yes
2007	1,026	0.958	0.983	0.914	Yes
2008	1,225	0.930	0.959	0.881	Yes
2009	1,414	0.897	0.896	0.898	No
2010	1,254	0.923	0.956	0.904	Yes
All	7,331	0.924	0.947	0.894	Yes

Notes: This table summarizes the match from the lottery records to the SIMS data. The sample excludes disqualified applicants, late applicants, out-of-area applicants, and siblings. SIMS = Student Information Management System. Offered > not offered determined from a two-group mean comparison *t*-test with a *p*-value of 0.95.

Table A.8. 2SLS for Cohorts with Same Match Rates: Effect Attending a Charter School, Per Year of Attendance, on MCAS Outcomes

	Math	ELA
	Grade 6	Grade 6
Subscale Outcome	(1)	(2)
Rare Standards Sample		
All Items	0.524***	0.198***
	(0.070)	(0.063)

Table A.8. Continued.

	Math	ELA
Subscale Outcome	**Grade 6 (1)**	**Grade 6 (2)**
Rare	0.516***	0.332**
	(0.061)	(0.131)
Somewhat Common	0.546***	0.181***
	(0.093)	(0.070)
Common	0.481***	0.158**
	(0.070)	(0.071)
N	695	694
Full Sample		
All Items	0.534***	0.210***
	(0.070)	(0.062)
Multiple Choice	0.558***	0.221***
	(0.073)	(0.050)
Short Answer	0.594***	-
	(0.081)	-
Open Response	0.369***	0.092
	(0.070)	(0.117)
N	767	697

Notes: The notes for this table are the same as those for table 4, except here results are only for lottery applicants in 2002 and 2009, when the SIMS match rate across the offered and not offered group was not significantly different. Seventh and eighth grade results are not reported due to small sample size.
***Significant at the 1% level; **significant at the 5% level.

Table A.9. Sample Selection

Applications to Charter Schools with Sufficient Records That Do Not Offer Enrollment to All Applicants	**8,183**
Excluding disqualified applications (wrong grade, repeat application, etc.)	8,159
Excluding late applications	8,092
Excluding out-of-area applications	8,018
Excluding applications with sibling priority	7,331
Excluding applications not matched to state database	6,771
Transforming to one observation to per applicant	5,213
Excluding students without a baseline demographic	4,339
Excluding students without a baseline test score in any subject	4,065
Excluding students without an outcome test score in any subject or grade	3,395

Table A.10. Charter School Participation in Lottery Based Analysis

	Available Spring Lottery Data (1)	Grade Range (2)	Notes (3)
Academy of the Pacific Rim Charter Public School	2005–2010	5–12	
Boston Collegiate Charter School	2002–2010	5–12	
Boston Preparatory Charter Public School	2005–2010	6–11	Initial offer only in 2005.
Dorchester Collegiate Academy Charter School	x	4–5	Opened September 2009. Became K–8 in 2006. Initial offer only in 2006.
Edward Brooke Charter School	2006–2009	K–8	Only middle grade entry lotteries used.
Excel Academy Charter School	2008–2010	5–8	
MATCH Charter Public School	2008–2010	6–12	Opened middle school 2008.
Roxbury Preparatory Charter School	2002–2010	6–8	
Smith Leadership Academy Charter Public School	x	6–8	

Notes: Schools that have entry grade lotteries only in kindergarten are excluded, which excludes Boston Community Charter School and Neighborhood House Charter School. Schools that closed in the relevant time period are excluded, which excludes Fredrick Douglass Charter School (closed 2005) and Uphams Corner Charter School (closed 2010). The remaining schools that do not contribute lotteries to the analysis are not oversubscribed or do not have sufficient lottery records.