GREAT LAKES CENTER
FOR EDUCATION RESEARCH & PRACTICE

Summary of Review

How New York City Charter Schools Affect Achievement estimates the effects on student achievement of attending a New York City charter school rather than a traditional public school and investigates the characteristics of charter schools associated with the most positive effects on achievement. Because the report relies on an inappropriate set of statistical models to analyze the data, however, the results presented appear to overstate the cumulative effect of attending a charter school. In addition, the report does not provide enough technical discussion and detailed description to enable a reader to assess the validity of some aspects of the report's methodology and results. Policymakers, educators, and parents, therefore, should not rely on these estimates until the authors provide more technical detail and the analysis has undergone rigorous peer review.

Executive Summary

The September 2009 report by Caroline Hoxby, Sonali Murarka, and Jenny Kang, *How New York City Charter Schools Affect Achievement*, has the potential to add usefully to the growing body of evidence regarding the effectiveness of charter schools. The report analyzes data from New York City in order to estimate the effects on student achievement of attending a New York City charter school rather than a traditional public school. In addition, the report investigates the characteristics of charter schools in New York that are associated with the most positive effects on achievement. Because of flaws in the report's statistical analysis, however, it appears to overestimate the effect of charter schools in New York City.

The key strength of the report is that it relies on the fact that the majority of students in New York City charter schools were admitted via lottery. The availability of randomized lotteries that determine admission to most of New York City's charter schools provides the researchers with the *opportunity* to obtain highly-credible estimates of the effect of attending a charter school rather than a traditional public school in New York City. Each lottery can be thought of as a small, randomized controlled experiment, with each school, year, and grade in which a lottery is conducted for admission serving as one such experiment. In principle, this allows researchers to estimate the effect of being admitted to a specific charter school, in a specific grade and year, on student achievement in subsequent years. Nonetheless, the mere availability of lotteries does not *guarantee* that the estimates provide credible approximations of how charter schools affect student achievement. In order to take advantage of the opportunity presented by the many lotteries, the researchers must use appropriate statistical models to analyze the data. As de-

scribed below, the statistical models used in most of the analyses are not appropriate.

Several issues in the report's analysis indicate the need for caution in accepting some of the report's conclusions. In particular:

- The report relies on an inappropriate set of statistical models to analyze the data. One feature in particular of the models used—the inclusion of students' test scores from the prior year—will likely lead to mis-estimation of the charter school effects. Because these test scores are measured *after* the lotteries take place, and so could be affected by whether students attend a charter school or not, they destroy the benefits of randomization. This flaw in the analysis affects the estimated effects of charter schooling on student achievement in grades 4-12. The estimated effects of charter schools on achievement by third grade are based on a different statistical model that does not share this flaw, however, and so are more credible.
- The report includes claims regarding the cumulative effects of attending a New York City charter school from kindergarten through eighth grade that are based on an inappropriate extrapolation.
- The report does not include adequately detailed information in some areas to allow a reader to fully assess its methods, results, or generalizability.
- The report uses a criterion for statistical significance that is weaker than that conventionally used in social science research;
- The report describes the variation in charter school effects across schools in a way that may distort the true distribution of effects by omitting many ineffective charter schools from the distribution.

As a result of the flaws in the report's statistical analysis, it likely overstates the effects of New York City charter schools on students' cumulative achievement, though it is not possible—given the information missing from the report—to precisely quantify the extent of overestimation. It may be that New York City's charter schools do indeed have positive effects on student achievement, but those effects are likely smaller than the report claims. Policymakers, educators, and parents should not rely on the report's conclusions regarding charter school effects in grades 4-12 until these issues have been fully investigated and the analysis has undergone rigorous peer review.

I.   INTRODUCTION

The question of whether charter schools are more effective than traditional public schools at improving students' academic achievement is of considerable interest to policymakers, educators, and parents. Some 1.5 million students in the U.S. attend roughly 4,500 charter schools, a number the Obama administration has been pushing states to increase.[1] Although charter schooling is a high-profile topic in educational policy, high-quality, systematic evidence regarding the effects of charter schooling is relatively rare. In the past few years, however, researchers have produced a growing number of studies and reports that attempt to provide evidence regarding the effectiveness of charter schools relative to the local traditional public schools that charter school students would have attended had charter schools been unavailable.[2]

Prominent among these recent reports is a September 2009 report by Caroline Hoxby, Sonali Murarka, and Jenny Kang, "*How New York City Charter Schools Affect Achievement*,"[3] which describes enrollment patterns and average achievement impacts of charter schools in New York City.[4] This report is the second in a series of reports by Hoxby and her colleagues analyzing data from New York City to answer three questions: (1) What kinds of students enroll in New York City charter schools? (2) What are the effects on student achievement of attending a New York City charter school rather than a traditional public school? (3) What features of charter schools are associated with more positive effects of charter schools (that is, what kind of charter schools have the most positive effects on achievement)?

The Hoxby, Murarka, and Kang report is written for a general audience and so contains little of the technical detail regarding data, sample sizes and attrition, and analytic methods that would typically be contained in a scholarly, peer-reviewed publication. Some, but not all, of the relevant supporting technical documentation, however, is contained in an earlier technical report by Hoxby and Murarka, *Charter Schools in New York City: Who Enrolls and How They Affect Their Students' Achievement*.[5] Both the report and the technical report will be referred to throughout this review. In addition, I have had several conversations with Caroline Hoxby over the last few weeks, in which she has graciously answered a number of questions to clarify aspects of the report's data and methodology. Nonetheless, this review will focus primarily on material included in the written reports; additional information provided by Hoxby is described in the endnotes.

Outline of This Review

Section II of this review summarizes the main findings and conclusions of the report. Section III assesses at some length the analyses regarding the effects of charter schools on student achievement (Chapter IV of the report), as these constitute the heart of the report. In particular, Section III calls attention to three specific weaknesses in the report. Section IV discusses a set of additional concerns regarding the report, concluding that some of these concerns are minor, while others require more detailed information to assess. Section V briefly discusses the analyses regarding the variation in charter school effects (Chapter V in the report). The final section of this review (Section VI) explores the usefulness of the report for guiding policy and practice. A technical appendix contains extensive details regarding the statistical issues discussed in the review.

II. FINDINGS AND CONCLUSIONS
   OF THE REPORT

The report contains three main sets of findings, corresponding to the three questions it poses.

What kinds of students enroll in New York City charter schools? The report finds that students who enroll in charter schools in New York are disproportionately non-Hispanic Black and poor, relative to all students in New York City's public schools (see Tables IIa and IIc in the report). Charter school students have average prior test scores similar to those of the average student in New York City schools (see Table IIb in the report). The report notes that these comparisons of prior test scores apply only to the roughly 20% of charter school students for whom test scores are available prior to their enrollment in charter schools—that is, students who enroll in charter schools in grades 4 and higher after attending another New York City school, since the test is initially administered to students in grade 3. These findings are discussed only briefly in this review. Instead greater attention is given to the next two sets of findings described below.

What are the effects on student achievement of attending a New York City charter school rather than a traditional public school? In terms of policy implications, the most important findings of the report are the estimates of the effects of attending a charter school rather than a traditional public school in New York City. These are summarized in Table IVc of the main report. That table indicates that by third grade the average student enrolled in a charter school in early elementary school gains 0.14 and 0.13 standard deviations on the math and English achievement tests, respectively, relative to how well she or he would have scored if enrolled in a traditional public school. The

typical charter school student will have been in a charter school three to four years by the end of third grade, so these estimates imply that charter schools increase student achievement by roughly 0.04 standard deviations per year in grades K-3. Likewise, the report finds that the average charter school student gains 0.12 and 0.09 standard deviations in math and English *each year* in grades 4 through 8, relative to what she or he would have gained each year in a traditional public school. These results are statistically significant at the $p<0.05$ level except for the early elementary school effect on English achievement, which has a p-value of 0.07 (marginally statistically significant by conventional standards).

The report concludes that these are cumulatively large average effects. To put these effects into concrete terms, the report compares the cumulative effect of attending a New York City charter school for nine years (from kindergarten through eighth grade) to the magnitude of average test score differences between students in Harlem and the wealthy New York community of Scarsdale. The estimated cumulative effect would be equal to roughly 66% of the "Scarsdale-Harlem gap" in English and roughly 86% of the gap in math (pages IV-8, IV-9).

The report includes a number of additional estimates of the effects of charter schools on other outcomes, including effects on science and social studies tests, effects on Regents Examination scores, and effects on the probability of graduating with a Regents Diploma. The report finds that charter schools increase students' science and social studies test scores in elementary school. Table IVe reports that charter schools are estimated to improve science test scores by 0.17 standard deviations by fourth grade and an additional 0.23 standard deviations per year from fifth through eighth grade. Although there is no

effect of charter schools on social studies test scores through fifth grade, Table IVe reports an estimated charter school effect on social studies test scores of 0.17 standard deviations per year from sixth through eighth grade. While these estimated effects are large, none are statistically significant by conventional social science standards; each has a p-value of roughly 0.15; the report describes the estimated effects as "marginally statistically significant" (see Table IVe).

The report finds that charter schools significantly ($p<0.05$) increase students' scores on the Regents Examinations by 0.13 to 0.25 standard deviations, depending on the test subject (see Tables IVf and IVg). Finally, the report states that charter schools increase the probability that a student will graduate by age 20 with a Regents Diploma by 7 percentage points for each year a student attends a charter high school. These estimates have p-values of roughly 0.15, however, well above conventional levels of statistical significance; nonetheless, the report describes them as "marginally statistically significant" (see Table IVh).[6]

What features of charter schools are associated with more positive effects of charter schools? The final sections of the report investigate if and how the effects of charter schooling vary across New York City's charter schools. The report finds considerable variation in the effects of individual charter schools; some have annual effects estimated to be greater than 0.20 standard deviations per year; most have annual effects estimated to be between 0 and 0.20 standard deviations per year; and some (enrolling roughly 10% of charter school students) have annual effects estimated to be negative.

The report investigates the associations between individual charter schools' effects and the characteristics of those schools. It finds

that, on average, charter schools with a longer school year, more time devoted to English instruction, a mission statement emphasizing academic performance, performance-based teacher pay, and a disciplinary policy that uses rewards and penalties have larger effects on student achievement than those schools without such policies. The report is careful to note that these associations cannot be interpreted causally.

The report's use of research literature

The report does not present or discuss prior research literature. While this is appropriate because the report intends primarily to describe the results of a single study in New York City, it would have been useful to situate the study within the larger body of scholarship on charter schools and their effects. This would help the reader understand to what extent the findings in New York may be generalizable to other contexts.

III.   REVIEW OF CHAPTER IV:
      THE EFFECTS OF NEW YORK
      CITY'S CHARTER SCHOOLS
      ON ACHIEVEMENT

The Use of Charter School Lotteries

The key design feature of the study is the use of charter school lotteries to identify comparable groups of charter school and traditional public school students. If charter and traditional public school students are different to begin with, then we cannot attribute differences in their later achievement to the effectiveness of charter schools relative to traditional public schools. Lotteries, in principle, can solve this problem, as the report notes, by ensuring that the only way that charter and traditional public school students differ initially is whether they won or lost a lottery coin-toss. Any subsequent difference in their achievement

patterns can then be attributed to the type of school they attend.

The availability of randomized lotteries that determine admission to most of New York City's charter schools therefore provides researchers with the *opportunity* to obtain highly credible estimates of the effect of attending a charter school in New York. Each school, year, and grade in which a lottery is conducted for admission serves as a small randomized controlled experiment that, in principle, allows researchers to estimate the effect of being admitted to a specific charter school, in a specific grade and year, on student achievement in subsequent years. By using a statistical model to average the effects of being admitted to charter schools in each of the hundreds of lotteries, one can obtain an estimate of the average effect of attending a charter school among the population of students who attend charter schools. Only a few other studies of charter schools have relied on lotteries;[7] among those, this study has by far the largest sample of schools and students (more detail below), and so has the potential to provide strong evidence regarding the effects of oversubscribed charter schools, at least within New York City.

Statistical Analysis of Lottery Data

Although the availability of lotteries used for admission to most New York City charter schools provides the *opportunity* to obtain unbiased estimates of charter school effects, the lotteries do not *guarantee* that the estimated effects are credible and unbiased. Several flaws in the statistical models used in the report call into question the report's estimates of the effects of charter schools.

- First, the statistical models used to estimate the effects of charter schools must not introduce bias into the estimates. If the study consisted of a single charter

school lottery in which all lotteried-in students enrolled in a charter school and all lotteried-out students enrolled in a traditional public school, a statistical model would not be needed; we could simply compare the later test scores of the charter and public school students. Because this study includes students who participated in hundreds of lotteries in different years and grades, and who were observed for different numbers of years and in varying grades, however, a statistical model is required to estimate the average effects of charter schooling. As a result, it is important to assess the appropriateness of the statistical models used for these purposes. As described below, the statistical models used to estimate the effects of charter schooling in grades 4-12 are inappropriate.

- Second, because there are relatively few students who have been in a lottery for many years, estimates of cumulative effects of charter schooling over many years must be based on an appropriate extrapolation. The report claims that the annual charter school effects are sufficiently large that a student who attended charter schools from kindergarten through eighth grade would close almost all of the "Scarsdale-Harlem gap" As described below, the report relies on an inappropriate extrapolation to estimate the cumulative effect of attending a charter school for many years.

- Third, the reader should have access to sufficiently detailed information to understand what set of students and schools are used to estimate the effects. For example, the report should include adequately detailed information to allow a reader to determine the extent to which the estimated effects are based primarily on data from a few charter schools or

from all charter schools. Although some such information is included in the technical report, much is absent. The report does not include adequately detailed information in some areas to allow a reader to fully assess its methods or generalizability.

Appropriateness of the Statistical Models

Bias due to inclusion of prior test scores in the statistical model. Page III-6 of the report describes the statistical model used in the report to estimate the effects of charter schooling. One feature of this model is that it relies on multiple observations of each student, one for each post-lottery year in which a student has a test score in the database. So a student who participated in a lottery to enroll in fourth grade in a charter school in 2004-05 will have (typically) four observations in the data: a fourth grade observation from 2004-05; a fifth grade observation from 2005-06; a sixth grade observation from 2006-07; and a seventh grade observation from 2007-08. The regression model then predicts a student's score in a given year as a function of whether he or she was enrolled in a charter school, controlling for his or her prior year's test score (which statisticians call a "lagged" score). Because this student participated in a lottery to enroll in a charter in school in fourth grade, the student's prior year's test scores in all but the first observation will have been measured *after* the lottery took place.

There are two primary problems with this statistical model. First, because the prior year's test score is measured *after* randomization (except for the first year a student is in a charter school), the model destroys the randomization that is the strength of the study's design. As discussed below, this will likely lead the models to overstate the effects of charter schools. A second problem

resulting from the inclusion of test scores measured after randomization in the statistical model is that these test scores are measured with error (i.e., test scores do not perfectly measure students' academic achievement). This also will lead to the models to overstate the effects of charter schools. Both issues are discussed in more detail below.

Controlling for a test score measured after randomization destroys the study's claim to validity based on the lotteries' random assignment. To see this, note that the statistical model that controls for the prior year's test scores implicitly compares charter school students to traditional public school students who had the same test score the prior year. While that may sound reasonable, it assumes that charter and traditional public school students with the same prior test scores can stand as valid counterfactuals for one another. That is, it assumes that students in charter schools, had they instead attended a traditional public school in a given year, would have learned the same amount in that year as those students in traditional public schools who started the year at the same achievement level. This would be a valid assumption if students were randomly assigned, *each year*, to attend a charter school or a traditional public school. If this were the case, students assigned to charter and traditional public schools who had the same prior year's test score would be, on average, the same as one another, so a comparison of their subsequent test scores would provide a valid estimate of the effect of attending a charter school in that year. But because enrollment to a charter school is not randomly assigned every year, students who are in a charter school cannot be assumed to be the same, in every way (including how much they would learn in a given year in a given school), as students in traditional public schools who start the year with the same level of achievement.

The statistical model used in the report thus relies on an unverifiable assumption. Although the existence of the lotteries provides an opportunity to obtain unbiased estimates of charter school effects without relying on such strong assumption, the report does not rely on the randomization, but instead relies on this unverifiable assumption about the similarity of students with the same prior year's test scores. If there were no lotteries, this might be defensible (or might at least be the best one could do), but given the existence of the lotteries, a much more straightforward and defensible analysis is possible. The estimated effects from the models containing lagged scores should not be treated as unbiased estimates.

Moreover, a relatively straightforward statistical analysis indicates that the bias due to including the lagged test scores in the model will likely tend to exaggerate the true effects (if the true effects are positive, the lagged score model will yield estimates that are too large; if the true effects are negative, the lagged score model will yield estimates that are too negative). See the technical appendix for more detailed discussion of this.

Finally, it is important to note that the estimated cumulative effects of charter schools in grades K-3 shown in the report are not subject to this type of bias because the models estimating them do not contain any prior test scores. The estimated annual effects of charter schools in grades 4-8, on the Regents Examinations, and on graduating with a Regents Diploma, however, are all subject to this type of bias, because they each rely on models that include prior test scores measured after the lotteries.

Measurement error in lagged test score will lead the estimates to overstate the charter school effect. A secondary problem resulting from the inclusion of test scores measured after randomization in the statistical model is that these test scores are measured with error.

It is a well-established fact that statistical models that condition on a variable measured with error will yield biased estimates.[8] In the models used here, the bias resulting from measurement-error will tend to exaggerate the effects of charter schools (if the true effects or charter schools are positive, the estimated effects will appear be larger than the true effects; if the true effects are negative, the estimated effects will appear as more negative than the true effects; for technical detail on this point, see the Appendix to this review). As noted above, the models used to estimate the effects of charter schooling by third grade are not subject to this type of bias.

The inclusion of lagged scores measured with error in the statistical models may account, in part, for the results showing that charter school effects are larger in grades 4-8 than in grades K-3. Returning to Table IVc, for example, the reported annual effects of charter schools in the early elementary grades are roughly 0.04 standard deviations per year (0.14 s.d. over 3-4 years), while the estimated annual effects in grades 4-8 are two to three times times larger (0.09-0.12 standard deviations per year). These larger figures are likely inflated by measurement-error-induced bias of the type described above (in addition to bias due to the fact that charter school enrollment is not random, conditional on students' prior scores). Without knowing the reliability of the tests and the average number of years elapsed between the lottery and the observations used in the models, it is impossible to say exactly how large the measurement-error-induced bias may be, but it clearly biases the estimates away from zero, making average charter school effects appear larger than they are.

Computation of the Cumulative Effects of Charter Schooling

Do charter schools significantly close the "Scarsdale-Harlem gap"? The report claims

that the effects of attending a charter school from kindergarten through eighth grade are sufficiently large that they would close the gap between the average student in Harlem and the average student in Scarsdale by 66% in English and 86% in Math. There are three reasons to doubt this claim. First, as described above, several aspects of the data and the statistical models could lead the estimated effects to be biased. If the true annual effects are smaller than the reported effects, then the true cumulative effects must likewise be smaller.

Second, the calculation of the cumulative effects of charter schooling in grades 4 through 8 is erroneous. Recall that one set of models attempts to estimate the *annual* effects of charter schooling in each grade from grade 4 to 8. Even if the models did this correctly (and they do not, because of the inclusion of the lagged test scores in the models and because of the presence of measurement error in the test scores), one cannot simply add the annual impacts estimated from these models in order to obtain the cumulative impact over a number of years. The reason for this is that achievement gains do not persist perfectly from year to year. Students who have a good year one year don't always have quite as good a year the next. Existing data from New York City indicates that about 76%-80% of a student's achievement gains (relative to his or her grade peers) in one year persist to the next year.[9] This means that annual achievement effects must be "discounted" before summing them up to compute a cumulative effect. The study does not appropriately discount the estimated effects. This error, in conjunction with the measurement error bias described above, likely results in the cumulative five-year effect of charter schooling from grades 4 through 8 being overestimated by as much 50% (see the technical appendix to this review for details).

Third, roughly two-thirds of the students in the study participated in lotteries to enter a charter school in the later years of 2004-05 or 2005-06 (see technical report, Table 5), meaning that most of the charter school students in the study have been in charter schools for only three or four years. Moreover, it does not appear that any students in the study could have participated in lotteries and been in a charter school for nine years from kindergarten through eighth grade, because the first lotteries in the study are from 2000-01. Even the number of students in the study who have been enrolled in charter schools for seven or eight years is likely relatively small, and most such students will have attended one of only a few charter schools in existence by 2001-02. This means that the only information about the long-term cumulative effects of attending a charter schools comes from a relatively small subset of students who enrolled in one of the very few charter schools operating before 2002.

Even if the effects of these few early-established charter schools were as strong as is claimed, it is far from obvious that such charter schools are typical of all the charter schools in operation today. Claims regarding the cumulative effects of attending a charter school for nine years are therefore based on an unwarranted extrapolation of the available data.[10] The report would be much more informative if it simply reported the cumulative effects of attending a charter school for a given number of years for the subset of students who attended for that number of years. This would allow a reader to see clearly how the charter school effects accumulate over multiple years, without relying on unwarranted extrapolation.

Inclusion of Sufficiently Detailed Technical Information

Data and sample. The report is based on analysis of test score data and charter school

lottery data for approximately 30,000 students[11] who applied to a New York City charter school in the years 2000-01 through 2005-06 and whose admission was determined by participation in one of 725 charter school lotteries. According to the report, roughly 94% of charter school students were admitted through randomized lotteries (p. vii). Of the 47 charter schools operating in New York City as of 2005-06, 43 are included in the report (more detail below). The report relies on standardized test score data (New York State tests and Regents Examinations) and Regents Diploma data that were available for the years 2000-02 through 2007-08. Because the large sample of students and charter schools, the large number of charter school lotteries, and the availability of test score data spanning grades 3-12 over eight years, the report aims to provide unbiased estimates of the effects of charter schools that are generalizable to most charter schools and their students in New York City.

The report contains very little in the way of detailed descriptive information about the data used in the estimation. There is no detailed information about (1) how many students participated in lotteries in each year and grade; (2) what proportion of lottery participants were lotteried-in and lotteried-out in each year and grade; and (3) what proportion of lotteried-in and -out students are observed in the data, how many years they are observed, and in what years and what grades they are observed. It is thus unclear where the bulk of the information that drives the estimates comes from. Moreover, the lack of such information makes it difficult to assess the extent to which schools and students the report's estimates apply or the extent to which there may be bias in the estimated charter school effects.

To which schools and students do the report's estimates apply? The report states that the estimates are "representative of New York City's charter school *students*; the more students a school has enrolled, the more influence it will have on the results of the study" (p. IV-2). More precisely, however, the estimates are representative of New York City's charter school students *who were admitted via lottery and who were in third grade or higher by 2007-08*; the more students in test-taking grades that a school has admitted via lottery, the more influence it will have on the results of the study.

There were 47 charter schools operating in New York City as of fall 2005; data from 43 of them are included in this study (two charter schools declined to participate; one closed in 2005-06; and one serves a special population of students). The report implies, but never states clearly, that each of these schools was oversubscribed in at least one grade and year and so admitted some students by lottery. However, some schools may have a very small number of students admitted via lottery and who were in test-taking grades by 2007-08.

If the proportion of students in a school admitted via lottery is correlated with the effectiveness of the school, then the effect estimates will be biased in the direction of the effects of the schools with the largest shares of students admitted via lottery. If better schools are more likely to be oversubscribed, and therefore likely to have higher proportions of their students admitted by lottery, they will be disproportionately overrepresented in the charter school effect estimates. Because the report includes no descriptive data regarding the proportion of students admitted via lottery in each school, we have no information that allows an assessment of whether the estimates provided here generalize to the population of New York City charter school students. It would be helpful for readers and users of this study if the study's authors provided estimates of the as-

sociation between charter schools' estimated effects and the proportion of their students who were admitted via lottery. This would be informative not only regarding the generalizability of the results, but also regarding the extent to which parental choices to apply to charter schools are related to school quality (another important policy question).

IV.   ADDITIONAL CONCERNS AND ISSUES

In addition to the issues described above, several issues are not adequately discussed in the report. These are briefly discussed below. Some appear to pose little threat to the validity of the report; in other cases, there is insufficient evidence in the report to assess these issues.

Potential bias due to differential matching rates. Depending on the year of lottery, between 9% and 21% of charter school applicants in the study who participated in lotteries were not able to be matched to the New York City Department of Education data (Table 5, technical report), meaning that test scores were unavailable for these students. The technical report argues (see p. 12) that the primary reason some applicants were not able to be matched is that they were not enrolled in a New York City public school prior to applying to a charter school, and then never attended one (because they subsequently attended a private school, a school outside New York City, or were home-schooled).

The fact that outcome data are not observed for all students who participated in lotteries may or may not lead to bias in the estimated effects of charter schools. For example, if the more academically able of the lotteried-out students disproportionately chose to go to private schools, then the study would be comparing charter school students to a less academically able comparison group of students in

traditional public schools. This would make the charter school students' later test scores appear better, relative to the comparison group, even if charter schools had no effect on student achievement.

The report does not indicate whether the matching rates differ between lotteried-in and lotteried–out students, which would have been helpful in determining the extent to which missing data of this sort may bias the estimates. The report would be more useful and informative if it provided more detailed information on the extent to which student records could not be matched to achievement data and how this non-matching varies by lottery status, cohort, grade of lottery, and student demographic characteristics. Because the report contains very little detail about the extent of non-matching, it is unclear to what extent this may bias the estimated results. The technical appendix to this review contains a brief discussion of the extent to which differential matching may result in biased estimates.

Reliance on balanced lotteries. The report relies for its effect estimates only on the subset of lotteries that appear to be "balanced"—that is, those where the characteristics of lotteried-in and lotteried-out students appear to be similar. These include 94% of students in lotteries. Under the presumption of random assignment within each lottery, standard practice would be to include all lotteries in the analysis rather than a subset of the lotteries. If the lotteries are indeed random (as they appear to be),[12] then estimates based on all lotteries will be unbiased estimates of the average effect of charter schools on all students who apply to charter schools and participate in lotteries. Despite this, the main report includes only estimates based on balanced lotteries, and notes that results based on all lotteries are "similar" (p. III-5) to those based only on

balanced lotteries. Because there is no evidence suggesting the lotteries (even the unbalanced ones) were not conducted via random assignment, it is unclear why the estimates from the balanced lotteries are preferred. At a minimum, the report should include a presentation of results based on balanced and unbalanced lotteries, as did the technical report (albeit using data from 2005-06).[13]

Description of charter school student population. The report describes how students who apply to and enroll in charter schools compare with students in New York City public schools as a whole, which is useful. Equally useful would be a description of how these students compare with the students in the subset of traditional public schools that the charter school students would have attended if they were lotteried-out. That is, we would like to know if charter schools attract students who are disproportionately high- or low-achieving relative to the public schools they come from. This would directly answer the concern of some who argue that charter schools "cream" the highest achieving students from traditional public schools.[14]

The technical report contains some information relevant to this question. It shows that students who applied to charter schools had higher average prior test scores than the average student who attends the traditional public schools from which the charter school students come (see Table 9 of the technical report). That is, within a given traditional public school, the students who apply to charter schools tend to be higher achieving (by 0.15-0.30 standard deviations) than those who do not. However, Hoxby told me that she has redone the analyses reported in Table 9 of the technical report, and that these new analyses show that charter school students do not have higher average prior achievement than their peers at the traditional public schools from which they come. Because these new analyses are not included in the report, however, a reader cannot assess their validity or implications.

Lottery oversubscription rates. The statistical models (described on page III-6) include what statisticians call "lottery fixed effects." The inclusion of the fixed effects ensures that charter school students are implicitly compared only to traditional public school students who participated in the same lottery. One consequence of using a fixed-effects model, however, is that the model implicitly weights students more in the estimation if they participated in a charter school lottery that was highly oversubscribed than if they participated in a lottery that was less oversubscribed (see Appendix for a more technical discussion). This means that if individual charter schools' effects are related to the extent to which they are oversubscribed, then more effective charter schools may be systematically over- or under-weighted in the estimation, leading to bias in the results. There is no information of lottery oversubscription rates in the text, though Hoxby reported to me that "all NYC charter schools have approximately half of the students in the lotteried-in group and half in the lotteried-out group."[15] This implies no correlation between lottery oversubscription rates and charter school effects, and so implies that there is no resulting bias in the estimates due to this source. This should be clearly documented in the text.

The reported finding that all lotteries are roughly equally oversubscribed is somewhat surprising, given that the lotteries include both new schools and more mature schools and span a range of years and grades. This merits discussion in the text. In particular, it raises an additional set of important questions. Market competition theory would sug-

gest that if parents have information about either the quality of their local traditional public schools or the available charter schools, they will be more likely to apply to the charter schools that are highly effective relative to the local traditional public schools. This would lead, all else being equal, to more applicants to the most effective charter schools (or more applicants among families whose alternative traditional public school is ineffective). In fact, part of the rationale for charter schools (and for private school vouchers and other forms of school choice) is that competition among schools will force schools to improve or to lose enrollment. If almost all charter schools are not only over-subscribed (according to the report, 43 of 47 charter schools in New York City rely on lot-teries for admission), but are all equally over-subscribed, it suggests parents may not be able to differentiate among charter schools in terms of quality (though it does suggest that there is excess demand for charter schools in New York City). Although this study is not designed to address the reasons for this, the authors (or others) are encouraged to consider it in future analyses. At a minimum, this study should report detailed data on lottery oversubscription rates that would inform fu-ture studies.

Compliance-effect correlation bias. The re-port relies on instrumental variables (IV) models to provide estimates of the average effect of charter schooling. One feature of IV models of the sort used in the report is that they implicitly weight students by how long they remain in a charter school. If stu-dents for whom charter schools have larger positive effects are more likely to stay in char-ter schools if lotteried-in than are students for whom charter schools have smaller effects, then those students for whom charter schools are more effective will tend to get more weight in the estimates, meaning the estimated effects will be overstated. The New York

City report states that very few (8%) lotte-ried-in students who enroll in a charter school for at least one year ever return to the traditional public schools. Given these low rates of attrition from charter schools, the threat to internal validity posed by using the IV model appears minimal. Nonetheless, the low attrition rate relative to that reported in other studies also suggests the uniqueness of the New York City charter school context.

Existing research from other studies pro-vides some evidence both of high attrition rates in charter schools and that continued charter school enrollment may indeed be correlated with the magnitude of charter school effects. An analysis of attrition from KIPP charter schools in the San Francisco Bay area, for example, found that students who left the KIPP schools prior to eighth grade had experienced smaller gains in the KIPP schools in fifth grade than their coun-terparts who stayed in the KIPP schools.[16] A study of Boston charter schools found high rates of attrition from middle and high school charter schools, though it did not ex-amine whether those who left charter schools had different patterns of achieve-ment gains in charter schools than those who stayed.[17] The low attrition rate in New York City relative to that reported in other studies suggests the possible uniqueness of the New York City charter school context.

Are the estimated effects of charter schooling statistically significant? As described above, the report estimates the effects of charter schooling on four sets of outcomes: (1) math and English tests in third through eighth grade; (2) science and social studies tests in fourth through eighth grade; (3) Regents Ex-aminations in high school; and (4) graduation with a Regents diploma by age 20. Of these, the reported effects on the math and English tests in third through eighth grade and on the Regents Examinations are generally statisti-

cally significant at the conventional $p<.05$ level, meaning that we can confidently conclude that the true effects are not equal to zero (though even this conclusion is dependent on the assumption that none of the biases discussed above are present). The other two sets of reported effects (in science and social studies tests in fourth to eighth grade and on graduation with a Regents diploma) have p-values of roughly 0.15, well above conventionally accepted standards of statistical significance. The report's description of these findings as "marginally statistically significant" does not conform to standard practice in social science research, and is therefore potentially misleading.

V   REVIEW OF CHAPTER V: ASSOCIATING CHARTER SCHOOLS' EFFECTS WITH THEIR POLICIES

The final sections of the report describe how much the effects of charter schools vary and investigate how the magnitude of charter school effects are associated with various characteristics of the schools. While this section of the report is primarily descriptive and exploratory, several issues should be considered in interpreting the results.

Distribution of achievement effects. The report finds a moderate amount of variation among charter schools in their effects on achievement. This is shown in Figures IVg and IVh, which illustrate the distribution of charter schools' effects on math and English achievement, respectively. According to the text, the figures include only charter schools whose effects were large or whose effects were relatively precisely estimated.[18] This means that some (unreported number of) charter schools are omitted from the figures. Schools whose effects are imprecisely estimated are generally schools with small numbers of students who (a) have participated in lotteries and (b) are in grades with

test scores. That means newer schools, schools serving primarily early elementary grades, small schools, and schools in which only a small proportion of students were admitted via lottery are more likely to be excluded, compared with older schools, large schools, those serving middle school grades, and those in which most students are admitted via lottery. It is unclear how many schools are omitted from the figure, but it is possible that the figure misrepresents the actual distribution of charter school effects by excluding some schools. More information would be helpful in this regard.

In addition, it would be helpful if the report were to present the distribution of all estimated effects (as was done in the technical report, albeit using data from fewer years), rather than just the most precisely estimated effects. In the technical report, the precisely estimated effects have a distribution that is much more positive, on average, than the distribution of all effects (see Figures 5 and 6 in the technical report). This suggests that the exclusion of the imprecise estimates causes some distortion in the figures, perhaps giving the impression that the average effects of charter schools are more positive than they in fact are—a distortion in addition to overestimation due to the bias issues described above.

Factors associated with variation in charter school effects. In addition to describing the extent of variation in the charter school effects, the report describes how the magnitudes of the charter school effects are related to a set of charter school policies. Recall that the factors highlighted in the report are a longer school year, more time devoted to English instruction, academic-based mission statements, performance pay for teachers, and a reward/penalty-based disciplinary policy. The report is appropriately careful to acknowledge that these associations cannot be

interpreted as causal relationships. For example, the existence of a correlation between the use of performance-based teacher pay and the effectiveness of charter schools does not imply that if a school adopts performance-based pay its students will learn more.

In addition to the policies of charter schools themselves, there are two potentially important reasons for variance in achievement effects among charters schools, neither of which is explored in the report. First, different charter schools enroll different populations of students. If New York City charter schools as a whole are more effective for some types of students than others, then charter schools enrolling more of those types of students will appear more effective than those enrolling fewer such students, even if the two schools are otherwise identical. For instance, New York City charters may, relative to traditional public schools in the City, take more advantage of involved parents. If this is the case, then charters that enroll more students with such parents might be expected to show better outcomes than those that do not.

Second, it is worth noting that each charter school (each lottery, actually) has a somewhat different counterfactual. What is described as the effect of a given charter school is more accurately described as the effect of attending that charter school rather than the traditional public school that a student would have attended had she or he not been lotteried-in to the charter school. Because the traditional public schools attended by lotteried-out students differ across lotteries (students applying to a charter school in the Bronx are unlikely to attend a traditional public school in Brooklyn if lotteried-out, and vice versa), the comparison traditional public schools differ across charter schools. As a result, the variation in the charter school effects reported in the study may result as much from variation in the quality of

the traditional public schools that are alternatives to each charter school as from variation in the quality of charter schools themselves. A given charter school of moderate quality may appear very effective if compared with a low-quality traditional public school that the students would otherwise have attended. But that same charter school would appear ineffective if the students' alternative school happened to be a high-quality traditional public school. This does not in any way undermine the lottery-based design of the study; it is meant only to point out a potential reason for variation among charters that was not discussed in the report.

None of the above necessarily disqualifies the descriptive results reported regarding the association between charter school policies and their effectiveness, but it should make clear that we cannot unambiguously attribute variation in charter school effectiveness to charter school policies alone. Nor, as the report notes, can we be sure that adopting policies such as a longer school day, performance pay for teachers, or academic-based mission statements would result in improved student achievement.

VI.   USEFULNESS OF THE REPORT FOR GUIDANCE OF POLICY AND PRACTICE

The data used in the New York City charter schools report have the potential to be very valuable for informing policymakers and educators regarding the effectiveness of oversubscribed charter schools in New York City. The data include information on more than 40 charter schools, 725 distinct lotteries, and thousands of students observed over multiple years and grades.

Nonetheless, the analyses in the study contain several possible sources of bias. The most significant of these stem from the in-

clusion of students' prior year's test scores in the statistical models estimating the effects of charter schooling for all grades above third grade. This both destroys the randomization that is the potential strength of the study's design and introduces measurement-error bias in the estimates. In addition, the estimated cumulative effects of charter schools are based on an incorrect computation and on an extrapolation beyond the limits of the data. As a result of these flaws, the report considerably overstates the effects of New York City charter schools on students' cumulative achievement.

The likely bias in the estimates could be easily eliminated by using a more appropriate statistical model, one that does not condition on lagged test scores. Estimates from such models would be more defensible, more transparent, and would avoid relying on untenable assumptions and invalid extrapolation. Given the prominence of charter schools in current education policy discussions, it is important that policymakers, parents, educators, and scholars have access to the most accurate possible evidence regarding the effects of charter schools.

That said, it is worth noting that the estimates of the effects of charter schooling in kindergarten through third grade do not suffer from the same statistical flaws as the estimates of effects in fourth grade and higher. These estimates indicate that students who attend charter schools in the early elementary grades have higher achievement levels (0.13-0.14 standard deviations higher by third grade) than they would have had if they had attended a traditional public school.

This is important evidence. In addition, despite the sources of bias in the estimates for the higher grades, there may still be positive average effects of charter schooling in grades 4-8. A more appropriate statistical analysis would provide a clear answer.

In addition, the report contains very little in the way of detailed information regarding lottery participation and oversubscription rates. As a result it is unclear how generalizable the results are across charter schools, grades, and time. Moreover, it is unclear whether there may be additional sources of bias, because the report does not always include enough information for the reader to assess the methods and results. The report would be far more useful to scholars seeking to understand if and how charter schools affect student achievement, for example, if it presented a detailed accounting of application, enrollment, retention, and attrition patterns among charter school students.

As a result of the potential sources of bias and the lack of detailed information in the reports to assess the extent of such bias, it is not possible for this reviewer or other readers to determine the degree to which the estimated charter school effects in grades 4 and above are valid. Policymakers and educators should therefore not rely on these estimates until the bias issues have been fully investigated and the analysis has undergone rigorous peer review. Given the quality of the data, however, a revised version of the analysis could provide a more definitive answer regarding the effectiveness of New York City's oversubscribed charter schools.

Technical Appendix

A. Problems With the Statistical Model Used in the Report

A stylized example. Before discussing the statistical model, it is useful to simplify matters by considering a single lottery, in which students are randomly assigned to a treatment or control group. Suppose all students comply with their treatment assignment for the duration of the time we observe them (that is, if lotteried-in, students attend a charter school for the duration of the study; if lotteried-out, they attend a traditional public school for the duration of the study). Suppose also that we observe a test score for each student prior to randomization and at the end of each year after randomization. We standardize all test scores to have a mean 0 and a standard deviation 1 in the control group (standardizing based on the control group will make all estimated effects expressible in terms of effect sizes relative to the control group distribution; it also make the derivations below a bit simpler).

We are interested in knowing the cumulative effect of the treatment after N years. That is, we want to know how much higher a student's score would be in year N if he attended a charter school from year 1 to N instead of a traditional public school. A straightforward way to do this would be to fit either the model

$$Y_{iN} = \beta_N + \delta T_i + \epsilon_{it}, \quad \epsilon_{it} \sim [0, \sigma^2]$$

or the model

$$Y_{iN} = \beta_N + \gamma Y_{i0} + \delta T_i + \epsilon_{it}, \quad \epsilon_{it} \sim [0, \sigma^2].$$

Either would yield an unbiased estimate of the average cumulative effect of attending a charter school for N years, measured in standard deviation units. If we had access to test scores measured prior to randomization (Y_{i0}), the second model would be preferable because it would likely yield more precise estimates because of the inclusion of pre-randomization test scores.

What model does the report fit? The report uses a model of the first type above in estimating the average cumulative effects of charter schooling by third grade. To estimate the annual effect of charter schooling in grades 4 and above, however, the report relies on a model of this form (see page III-6):

$$Y_{it} = \beta_t + \gamma Y_{i,t-1} + \delta T_i + \epsilon_{it}, \quad \epsilon_{it} \sim [0, \sigma^2]$$

This model says that a student's observed test score in year t depends linearly on his or her score in the prior year plus an effect δ of attending a charter school, plus some random, i.i.d, mean-zero error. Because both Y_{it} and $Y_{i,t-1}$ are standardized, $\beta_t = 0$, $\gamma = corr(Y_{it}, Y_{i,t-1})|T$ and $\sigma^2 = 1 - \gamma^2$. This model is fit using the stacked data; that is, if student test scores are observed for N years following randomization, each student will have N observations in the data. As is standard in models like this, the standard errors are corrected for the fact that there are multiple observations per student. Note that the model used in the report is actually more complex than this—it also includes lottery fixed effects, grade fixed effects, year fixed effects, and student covariates, and is fit using a lottery assignment as an instrument for enrollment in a charter school

in each year. Nonetheless, this additional complexity does not affect the issues described below.

Problem 1: Controlling for test scores measured *after randomization* destroys the benefits of randomization. In the model described above and used in the report, OLS will yield an unbiased estimate of δ if treatment assignment is independent of the error, conditional on $Y_{i,t-1}$ (that is, if $T_i \perp \epsilon_{it} | Y_{i,t-1}$). Because $Y_{i,t-1}$ is measured *after* randomization (except for the very first observation, when $t-1 = 0$), however, the randomization does not ensure that $T_i \perp \epsilon_{it} | Y_{i,t-1}$) for all $t > 1$. Thus, including lagged test scores measured after randomization destroys the primary benefit of randomization—the guarantee that treatment status is uncorrelated with all characteristics of students, observed or unobserved.

It is possible, under some mild assumptions, to determine the likely direction and magnitude of the bias in models that include lagged scores measured after randomization. Suppose each student's potential outcomes in years 1 and 2 are defined by

$$Y_{i1} = Y_{i0} + \alpha_{i1} + \delta_{i1} T_{i1}$$
$$Y_{i2} = Y_{i0} + \alpha_{i1} + \delta_{i1} T_{i1} + \alpha_{i2} + \delta_{i2} T_{i2}$$

where Y_{i0} is the student's score prior to randomization; α_{i1} and α_{i2} are the gains that a student would make in years 1 and 2, respectively, if he or she attended a traditional public school; T_{i1} and T_{i2} are dummy variables indicating whether a student attended a charter school in years 1 and 2, respectively; and δ_{i1} and δ_{i2} are the effects of attending a charter school in years 1 and 2, respectively. Note that the above implicitly assumes that the gain a student would make in year 2 is independent of whether or not she or he attended a charter or traditional public school in year 1. Suppose, moreover, that test scores of the control group are standardized at each wave.

In order to estimate the effect of charter schooling in year 2 on the students who attended charter schools, we must estimate what the test scores of the charter school students would have been in year 2 had they attended a traditional public school in year 2. It is tempting to use the observed year 2 test scores of public school students who had similar test scores as the charter school students in year 1. However, this assumes that the gains that charter school students would have made in year 2 had they gone to a traditional public school (α_{i2}) are the same as the gains made by traditional public school students who had similar scores in year 1. Had we randomized students to charter or traditional public schools at the end of year 1, we could be sure that the outcomes of traditional public school students in year 2 could stand as a valid counterfactual for the charter school students' test scores in year 2. However, it we randomized students earlier than the end of year 1, it is not clear that this would be a valid counterfactual. To assess whether this is valid, consider the following.

Suppose the effect of charter schooling in year 1 is positive. Then it will be true that

$$E[Y_{i0} | T = 1, Y_1] < E[Y_{i0} | T = 0, Y_1]$$

That is, among students with the same scores at time 1, the charter school students will have had lower average scores at time 0 than the traditional public school students. If there is a non-zero correlation between students' time 0 scores and the gains they would make in year 2 in a tradi-

tional public school (that is, if $cov(Y_{i0}, \alpha_{i2}) \neq 0$), then traditional public school students' scores at time 2 cannot stand as valid counterfactuals for charter school students who had the same scores as them at time 1. Moreover, the direction of bias will depend on the sign of $cov(Y_{i0}, \alpha_{i2})$. If $cov(Y_{i0}, \alpha_{i2}) < 0$,

$$E[\alpha_{i2}|T = 1, Y_1] > E[\alpha_{i2}|T = 0, Y_1]$$

which implies that the students in charter schools would have had higher scores, had they been in traditional public schools, than the traditional public school students who had the same scores at time 2. As a result, the estimates based on comparing charter and traditional public school students with the same time 1 scores will be biased upwards. If, on the other hand, $cov(Y_{i0}, \alpha_{i2}) > 0$, the same logic implies that the estimated charter school effects in year 2 will be biased downward. Because the test scores are standardized within each wave, it is straightforward to write the covariance of Y_{i0} and α_{i2} as

$$\begin{aligned} cov(Y_{i0}, \alpha_{i2}) &= cov(Y_{i0}, Y_{i1} + \alpha_{i2}) - cov(Y_{i0}, Y_{i1}) \\ &= cov(Y_{i0}, Y_{i2}) - cov(Y_{i0}, Y_{i1}) \\ &= corr(Y_{i0}, Y_{i2}) - corr(Y_{i0}, Y_{i1}). \end{aligned}$$

In general, the correlation between student test scores at time 0 and time 2 will be lower than the correlation between scores at time 0 and time 1.[19] Thus, $cov(Y_{i0}, \alpha_{i2}) < 0$. This implies that, if the true effect of charter schools is positive, the likely direction of bias due to conditioning on lagged standardized scores is upward (because charter school students will, on average, have lower initial scores than public school students with the same scores at time 1; and so the fact that $cov(Y_{i0}, \alpha_{i2}) < 0$ implies they would have gained more, on average, in a traditional public school in year 2 than did the students who were in a traditional public school). If δ_{t-1} is the cumulative effect by time $t - 1$, then the magnitude of the bias will be

$$bias = \frac{cov(Y_{i0}, \alpha_{i2})}{cov(Y_{i0}, Y_{i1})} \cdot \delta_{t-1}.$$

That is, the annual effects will be overstated by an amount that is proportional to the cumulative effect of charter schooling over the average number of years students have been in charter schools when their lagged scores are measured.

Problem 2: Measurement error in test scores will lead to biased estimates of the charter school effect. Let us suppose the statistical model is correct. That is, suppose treatment assignment is ignorable conditional on the lagged score. In that case, OLS will yield an unbiased estimate of δ so long as $Y_{i,t-1}$ is measured without error. But test scores do not measure achievement without error. Therefore, it is important to consider if and how measurement error in $Y_{i,t-1}$ may affect the estimate of δ.

To understand the impact of measurement error on the estimate of δ, it is useful to begin by considering a model estimated on a single year's observation for each student. Suppose lagged test scores are measured without error and the true data generating model is given by

$$Y_{it} = \beta_t + \gamma Y_{i,t-1} + \delta T_i + \epsilon_{it}, \quad \epsilon_{it} \sim N[0, \sigma^2]$$

Now, suppose the lagged test score includes independent random measurement error with variance τ. Because the observed test scores are standardized, this implies that the true score has variance $1 - \tau$ within treatment groups. That is,

$$Y_{i,t-1}^* = Y_{i,t-1} + \omega_{i,t-1}, \quad \omega_{i,t-1} \perp Y_{i,t-1}, \omega_{i,t-1} \sim N[0, \tau], var(Y_{i,t-1})|T = 1 - \tau$$

Now define the reliability of the lagged score to be

$$r = \frac{var(Y_{i,t-1})|T}{var(Y_{i,t-1}^*)|T}$$
$$= 1 - \tau$$

Now if we fit the model to the observed data using OLS – that is, if we fit

$$Y_{it} = \beta_t^* + \gamma^* Y_{i,t-1}^* + \delta^* T_i + \epsilon_{it}^*, \quad \epsilon_{it}^* \sim N[0, \sigma^2],$$

then OLS will yield

$$plim \; \hat{\gamma}^* \Rightarrow \frac{cov(Y_{i,t-1}^*, Y_{it})|T}{var(Y_{i,t-1}^*)|T}$$
$$\Rightarrow \frac{cov(Y_{i,t-1} + \omega_{i,t-1}, Y_{it})|T}{var(Y_{i,t-1}^*)|T}$$
$$\Rightarrow r \frac{cov(Y_{i,t-1}, Y_{it})|T}{var(Y_{i,t-1})|T}$$
$$\Rightarrow r\gamma$$

and

$$plim \; \hat{\delta}^* \Rightarrow [\bar{Y}_t|T = 1] - [\bar{Y}_t|T = 0] - \hat{\gamma}^*([\hat{Y}_{t-1}^*|T = 1][\hat{Y}_{t-1}^*|T = 0])$$
$$\Rightarrow \delta + \gamma([\hat{Y}_{t-1}|T = 1][\hat{Y}_{t-1}|T = 0]) - r\gamma([\hat{Y}_{t-1}^*|T = 1][\hat{Y}_{t-1}^*|T = 0])$$
$$\Rightarrow \delta + \gamma(1 - r)([\hat{Y}_{t-1}^*|T = 1][\hat{Y}_{t-1}^*|T = 0])$$
$$\Rightarrow \delta + \gamma^* \frac{(1-r)}{r} \Delta Y_{t-1}$$

where ΔY_{t-1} is the difference in average test scores at time $t - 1$ between charter school students and traditional public school students. The bias will depend on γ^* (the observed correlation between students' scores at times $t - 1$ and t), the reliability of the observed scores, and ΔY_{t-1}. If the reliability is close to 1, $\frac{1-r}{r} \approx 0$, so the bias will be small. Likewise, if the difference in prior scores is small, the bias will be small.

Because the observed test scores are standardized, γ^* is the correlation between scores at time t

and $t-1$. This will be relatively high. Data from a study of the test scores of New York City public school students show that the grade-to-grade correlations of student math scores range from 0.76 to 0.80 across grades 3 to 8, with an estimated average of 0.79.[20] For the sake of illustration, let us assume $\gamma^* = 0.79$. The reliability must be $\gamma^* \leq r \leq 1$. Boyd et al. estimate the reliability of the New York state tests to be 0.831.[21] Then the bias will be

$$bias = \gamma^* \frac{(1-r)}{r} \Delta Y_{t-1} = 0.79 \frac{(1-0.83)}{0.83} \Delta Y_{t-1}$$

We know that, at the end of third grade, $\Delta Y_{t-1} = 0.14$ (TABLE IVc). This suggests that an estimate of the annual effect of charter schooling on achievement during fourth grade will be biased upwards by $0.79 \cdot 0.20 \cdot 0.14 = +0.022$. This is one-quarter to one-fifth of the estimated annual effects on reading and math. Because the model is fit on the stacked data, the amount of bias will depend on the magnitude of ΔY_{t-1} when pooled over all observations used in the estimation. Because information regarding ΔY_{t-1} is not available in the report, it is not clear how large the measurement error bias will be. Nonetheless, the bias will clearly inflate the magnitude of the true effects.

<u>Problem 3: The cumulative effect of charter schooling is not equal to the sum of the estimated annual effects</u>. In order to compute the estimated cumulative effect of charter schooling over a number of years, one cannot simply multiply the estimated annual effect by the number of years. This is because students' achievement gains do not persist perfectly from year to year. To see this, consider the following. Suppose the model used in the report represents the true data generating process. That is, suppose the data are generated by the following structural model:

$$Y_{i1} = \gamma Y_{i0} + \delta T_i + \epsilon_{i1}, \quad \epsilon_{i1} \sim [0, 1-\gamma^2]$$
$$Y_{i2} = \gamma Y_{i1} + \delta T_i + \epsilon_{i2}, \quad \epsilon_{i2} \sim [0, 1-\gamma^2]$$
$$\vdots$$
$$Y_{iN} = \gamma Y_{i,N-1} + \delta T_i + \epsilon_{iN}, \quad \epsilon_{iN} \sim [0, 1-\gamma^2]$$

where $[var(Y_{in})|T] = 1 \; \forall n \in \{0, 1, \dots N\}$. Note that, because each of the test scores Y_{in} are standardized, $\gamma = corr(Y_{in}, Y_{i,n-1})|T \; \forall \in \{1, 2, \dots N\}$.

Then, substituting Y_{i1} into the equation for Y_{i2}, we get

$$Y_{i2} = \gamma^2 Y_{i0} + (1+\gamma)\delta T_i + v_{i2}, \qquad v_{i2} = (\gamma \epsilon_{i1} + \epsilon_{i2}); v_{i2} \sim N[0, 1-\gamma^4]$$

Likewise, substituting this into the equation for Y_{i3}, we get

$$Y_{i3} = \gamma^3 Y_{i0} + (1+\gamma+\gamma^2)\delta T_i + v_{i3}, \qquad v_{i3} = (\gamma v_{i2} + \epsilon_{i3}); v_{i3} \sim N[0, 1-\gamma^6]$$

and so on. The Nth equation will be

$$Y_{iN} = \gamma^N Y_{i0} + \left(\sum_{n=1}^{N} \gamma^{n-1}\right) \delta T_i + v_{iN}, \qquad v_{iN} = (\gamma v_{i,N-1} + \epsilon_{iN}); v_{iN} \sim N[0, 1-\gamma^{2N}]$$

Thus, the cumulative effect of attending a charter school for N years is equal to $\Delta_N = (\sum_{n=1}^{N} \gamma^{n-1})\delta$, where δ is the annual effect. Unless $\gamma = 1$, $\Delta_N < N\delta$. Because γ is the correlation of test scores in adjacent years, it will be less than 1 unless all students have identical growth rates in test scores and there is no measurement error in the test scores. Any estimate of the cumulative effect of attending a charter school that is based on simply summing the estimated annual effects with therefore overestimate the true cumulative effect.

By how much will using $N\delta$ overstate Δ_N? As noted above, data from a study of the test scores of New York City public school students show that the grade-to-grade correlations of student math scores range from 0.76 to 0.80 between across grades 3 to 8. For the sake of illustration, let us take $\gamma = 0.8$. If the estimated value of the annual effect of charter schooling is 0.12 in math, the cumulative effect over the five years from grades 4-8 will be

$$
\begin{aligned}
\Delta_5 &= \left(\sum_{n=1}^{5} (0.8)^{n-1} \right) 0.12 \\
&= (1 + 0.8^1 + 0.8^2 + 0.8^3 + 0.8^4)0.12 \\
&= 3.36 \cdot 0.12 \\
&= 0.40
\end{aligned}
$$

Compare this to the value we would get if we simply sum the annual effects. If we simply multiply the estimated annual effect (0.12) by 5 years, we get an estimated cumulative effect of 0.60; an estimate that is 50% larger than the correct cumulative effect implied by the structural model and the correlations between annual test scores.

Note that the derivation above assumes that test scores contain no measurement error. If there is measurement error in the test scores (as there certainly is), then the observed γ is due to a combination of measurement error and the extent of persistence in true achievement gains over time. The more measurement error there is, the less upward bias due to the incorrect accumulation of gains over time there will be (because if γ is primarily due to measurement error, then the correlation of true scores will be higher than 0.8), but the more upward bias due there will be due to measurement error, and vice versa. In other words, the combination of measurement error bias and the accumulation error will result in upward bias of the estimates, though the exact amount will depend on the extent of measurement error in the test scores. Both sorts of bias could be readily eliminated by estimating cumulative effects using a model that does not control for post-randomization test scores.

B. Potential Bias Due to Non-Matching of Students to Department of Education Records

In an experiment, the estimated treatment effect is given by

$$
\hat{\delta} = \bar{Y}_T - \bar{Y}_C
$$

where $\bar{Y}_T$ is the average outcome among students assigned to the treatment condition (charter schools in this case) and $\bar{Y}_C$ is the average outcome among those assigned to the control condi-

tion. Let π_T and π_C indicate the proportions of students assigned to the treatment and control conditions, respectively, whose outcomes are observed (some students' outcomes are not observed because they could not be matched to the Department of Education records, for example). Then we can write

$$
\begin{aligned}
\hat{\delta} &= \left[\pi_T \bar{Y}_T^{obs} + (1 - \pi_T)\bar{Y}_T^{miss}\right] - \left[\pi_C \bar{Y}_C^{obs} + (1 - \pi_C)\bar{Y}_C^{miss}\right] \\
&= \left[\bar{Y}_T^{obs} + (1 - \pi_T)(\bar{Y}_T^{miss} - \bar{Y}_T^{obs})\right] - \left[\bar{Y}_C^{obs} + (1 - \pi_C)(\bar{Y}_C^{miss} - \bar{Y}_C^{obs})\right] \\
&= \bar{Y}_T^{obs} - \bar{Y}_C^{obs} + \left[(1 - \pi_T)(\bar{Y}_T^{miss} - \bar{Y}_T^{obs})\right] - \left[(1 - \pi_C)(\bar{Y}_C^{miss} - \bar{Y}_C^{obs})\right]
\end{aligned}
$$

where $\bar{Y}_T^{obs}$, $\bar{Y}_C^{obs}$, $\bar{Y}_T^{miss}$, and $\bar{Y}_C^{miss}$ are the average outcomes among students in the treatment and control groups who are observed and missing, respectively. Because we can only base our estimates of the charter school effect on the cases for whom we observe outcomes, we will have

$$
\begin{aligned}
\hat{\delta}^{obs} &= \hat{\delta} + \left[(1 - \pi_C)(\bar{Y}_C^{miss} - \bar{Y}_C^{obs})\right] - \left[(1 - \pi_T)(\bar{Y}_T^{miss} - \bar{Y}_T^{obs})\right] \\
\hat{\delta}^{obs} &= \hat{\delta} + bias
\end{aligned}
$$

So the bias will be given by

$$
bias = \left[(1 - \pi_C)(\bar{Y}_C^{miss} - \bar{Y}_C^{obs})\right] - \left[(1 - \pi_T)(\bar{Y}_T^{miss} - \bar{Y}_T^{obs})\right]
$$

The report does not describe what proportions of lotteried-in and -out students are matched, so we do not know if π_C is larger, smaller, or equal to π_T. If students who are lotteried-out are more likely to subsequently enroll in a private school or a school outside the NYC public school system than are lotteried-in students, then lotteried-out students will have lower matching rates than lotteried-in students, so $\pi_C < \pi_T$. Moreover, we cannot observe $\bar{Y}_C^{miss}$ or $\bar{Y}_T^{miss}$, by definition, but the authors may be able to provide information on some characteristics of those who were not matched to DOE data (because some demographic information is contained on the lottery application forms and thus is available for students even if they are not matched). An analysis of these factors could help to assess whether non-matching may produce upward or downward bias in the estimates.

C. Weighting Due to Inclusion of Lottery Fixed Effects

In a model that includes lottery fixed effects, as do the models used in the report, the estimated effect will be

$$
plim(\hat{\delta}) \Rightarrow \sum_l \frac{N_l \pi_l (1 - \pi_l)}{\sum_l N_l \pi_l (1 - \pi_l)} \delta_l,
$$

where N_l is the number of students participating in lottery l, π_l is the proportion of students in the lottery who are lotteried-in in lottery l, and δ_l is the average effect of charter schooling among students participating in lottery l. That is, the estimand from the fixed effects model is a weighted average of the lottery-specific effects, where the weights are proportional to $N_l \pi_l (1 - \pi_l)$. The advantage of this weighting is that it produces a statistically efficient estimate—lotteries with the most information are weighted most, yielding precise estimates. The disadvantage of this weighting is that if the lottery-specific effects are correlated with the weights, the estimates

will be biased. To understand the bias, note that the number of students admitted via lottery l will be equal to $n_l = N_l \cdot \pi_l$, meaning that the weights are proportional to $w_l = n_l(1 - \pi_l)$. Thus, the fixed effects estimators weight each lottery not by the number of students admitted via lottery, but rather by both the number of students admitted via lottery and the proportion of students who are lotteried out $(1 - \pi_l)$. As a result, charter schools that have lotteries that are more highly oversubscribed will receive larger weight in the estimation. If charter schools that are more effective (or whose local traditional public schools are less effective) are, on average, more oversubscribed than those that are less effective, the lottery fixed effects estimates will be upwardly biased. Conversely, if charter schools that are less effective (or whose local traditional public schools are more effective) are, on average, more oversubscribed than those that are more effective, the lottery fixed effects estimates will be downwardly biased. As noted above, Hoxby reported to me that all the lotteries in New York City have oversubscription rates of roughly 50%, indicating that the weighting of the fixed effects estimator does not result in any appreciable bias.

Notes and References

[1] Maxwell, Lesli A. (2009, June 17). "Obama Team's Advocacy Boosts Charter Momentum," *Education Week*, 28: Issue 35, pp. 1,24-25; Maxwell, Lesli A. (2009, November 2). "Charter Schools Steadily Growing," *Edweek.org*, Retrieved November 3, 2009, from http://blogs.edweek.org/edweek/District_Dossier/2009/11/charters_continue_their_march.html.

[2] For recent studies, see, for example

Abdulkadiroglu, A., Angrist J., Cohodes S., Dynarski, S., Fullerton, J., Kane, T., & Pathak, P. (2009). *Informing the debate: Comparing Boston's charter, pilot and traditional schools*. Boston, MA: The Boston Foundation

Woodworth, K.R., David, J.L., Guha, R., Wang, H., & Lopez-Torkos, A. (2008). *San Francisco Bay Area KIPP schools: A study of early implementation and achievement. Final report.* Menlo Park, CA: SRI International

Center for Research on Education Outcomes (CREDO) (2009, June). *Multiple Choice: Charter School Performance in 16 States*. Palo Alto: CREDO, Stanford University

Hoxby, C. M. & Rockoff, J. E. (2004). *The impact of charter schools on student achievement*. Unpublished manuscript, Harvard University and Columbia University

Dobbie, W., & Fryer, R. G. (2009). *Are High-Quality Schools Enough to Close the Achievement Gap? Evidence from a Bold Social Experiment in Harlem*. Unpublished manuscript, Harvard University.

For reviews of the charter school effects literature, see

Betts, J. R. & Tang, Y. E. (2009, April). *Charter School Achievement: What We Know*, *5th Edition.* Washington, DC: National Alliance for Public Charter Schools

Betts, J. R. & and Tang, Y. E. (2008). *Value added and experimental studies of the effect of charter schools on student achievement*. Seattle, WA: National Charter School Research Project, Center on Reinventing Public Education, University of Washington, Bothell

Carnoy, M., Jacobsen, R., Mishel, L., & Rothstein,R. (2006). Worth the Price? Weighing the Evidence on Charter School Achievement. *Education Finance and Policy,* Winter 2006, Vol. 1, No. 1, Pages 151-161;

Hill, P.T., Angel, L., & Christensen, J.. Charter School Achievement Studies. *Education Finance and Policy,* Winter 2006, Vol. 1, No. 1, Pages 139-150

Miron, G., Evergreen, S., & Urschel, J. (2008). *The impact of school choice reforms on student achievement*. Boulder and Tempe: Education and the Public Interest Center & Education Policy Research Unit Retrieved November 10, 2009, from http://epicpolicy.org/files/CHOICE-10-Miron-FINAL-withapp22.pdf.

[3] Hoxby, C. M., Murarka, S. & Kang, J. (2009, September)."How New York City's Charter Schools Affect Achievement." Second report in series. Cambridge, MA: New York City Charter Schools Evaluation Project, September 2009. Retrieved October 1, 2009, from http://www.nber.org/~schools/charterschoolseval/.

[4] More specifically, the report provides estimated achievement impacts among students admitted to New York City charter schools via lottery during the years 2000-01 to 2005-06. Because almost all charter schools in New York City were oversubscribed, the report covers almost all charter schools in New York City during this period. More detail is provided on this point later in this review.

[5] Hoxby, C. M., Murarka, S.. (2009, April). "Charter Schools in New York City: Who Enrolls and How They Affect Their Students' Achievement." National Bureau of Economic Research Working Paper 14852, April 2009. Cambridge, MA: National Bureau of Economic Research. Retrieved October 1, 2009, from http://www.nber.org/~schools/charterschoolseval/.

[6] Note also that these estimates are based on only two charter schools: those that enrolled students in grades 9-12 in 2005-2006 (only students in these schools could have both attended a charter high school and be observed at age 20 in 2007-08). See Table IVa in main report and footnote 24 in the technical report.

[7] Abdulkadiroglu, A., Angrist J., Cohodes S., Dynarski, S., Fullerton, J., Kane, T., & Pathak, P. (2009). *Informing the debate: Comparing Boston's charter, pilot and traditional schools*. Boston, MA: The Boston Foundation

Hoxby, C. M. & Rockoff, J. E. (2004). *The impact of charter schools on student achievement*. Unpublished manuscript, Harvard University and Columbia University

Dobbie, W., & Fryer, R. G. (2009). *Are High-Quality Schools Enough to Close the Achievement Gap? Evidence from a Bold Social Experiment in Harlem*. Unpublished manuscript, Harvard University.

[8] Greene, W. H. (2003). *Econometric Analysis*. Fifth Edition. Upper Saddle River, NJ: Prentice-Hall.

[9] Boyd, D., Grossman, P. Lankford, H., Loeb, S., & Wyckoff, J. (2008). "Measuring Effect Sizes: the Effect of Measurement Error." Paper prepared for the National Conference on Value-Added Modeling University of Wisconsin-Madison April 22-24, 2008. Retrieved October 20, 2009, from http://www.teacherpolicyresearch.org/portals/1/pdfs/Measuring%20Effect%20Sizes%20%20the%20Effect%20of%20Measurement%20Error%20Boyd%20et%20al%20%2026Jun2008.pdf

[10] In order to check whether this extrapolation is valid (or at least to check whether the small subset of students who have been in charter schools for many years have benefited as much as is claimed), the report includes two figures (Figures IVa and IVb) that compare the regression-based cumulative gain estimates with the "raw data" for the subset of students who attended charter schools from third through eighth grade (or perhaps from kindergarten through first grade; the report includes contradictory statements regarding what students are included in the "raw data" lines in Figures IVa and IVb; see p. IV-8 and footnote 7, page A-1). It is unclear what is meant by the statement that the figures use "raw" data from students who attended charter school students. It may mean that the figure reports average annual gains for the subset of lotteried-in charter school students who are observed for the entire period from third through eighth grade. However, if this is the case, the comparison is problematic. The relevant check for the extrapolation would be to estimate the *effects* of charter schooling using only the subset of students who are observed from kindergarten through 8th grade (if there are any such students who participated in lotteries; the report does not appear to include students who participated in lotteries earlier than 2000-01) and to compare this estimate to the estimate based on the full sample of students. Figures IVa and IVb do not provide such a comparison. All they show is what the test scores of a subset of charter school students were from grades 3 to 8; this is not the same as reporting what the *effect* of attending a charter school is from grades 3 to 8 or grades K to 8. Accordingly, Figures IVa and IVb do not provide the necessary information to determine whether the extrapolation is valid even for the small subset of students who have attended charter schools for many years.

[11] Neither the main report nor the technical report includes clear and detailed information regarding sample sizes, but Table 5 in the technical report implies that roughly 33,000 students applied to charter schools and participated in lotteries from 2000-01 through 2005-06, of whom roughly 85-90% were matched to New York City Department of Education data.

[12] Because each charter school controls its own lotteries (that is, the charter school lotteries are not run by an external agency, unlike the New York City public school choice program), it is impossible to verify that the lotteries were random. Nonetheless, the report provides sufficient evidence to suggest that the lotteries were in fact random (and no evidence that suggests they were not). First, Table 10 in the technical report shows no statistically significant differences in demographic characteristics or prior test scores between those lotteried-in and those lotteried-out, when averaged over all lotteries in the data (see columns 1 and 4, Table 10, technical report). Second, the researchers examined each individual lottery to see if the lotteried-in and lotteried-out students were similar in each lottery. They report that the vast majority (94%) of lotteries appear balanced (if all lotteries were truly random, 95% would appear balanced according to the criterion used in the report; sampling variation will result in some lotteries yielding unbalanced samples). Thus, there is nothing in the reported data to suggest that the lotteries were not conducted through actual randomization. Moreover, the "unbalanced" lotteries as a group appear to be collectively "balanced." Table 10 in the technical report shows that the overall balance on observable characteristics is just as good, if not better, when all lotteries are included as when only the balanced lotteries are included. This implies that the 6% of lotte-

ries that are individually unbalanced are not systematically unbalanced in any one direction. They are just as likely to be unbalanced in favor of high-achieving as low-achieving students, for example.

[13] Although the September 2009 report does not include estimates based on data from all lotteries, the technical report includes estimates based on balanced lotteries alone as well as estimates based on all lotteries (albeit using 2005-06 data). Based on those earlier data, the technical report shows that the estimated effects are smaller (40% smaller in math, 15% smaller in English) when all lotteries are included (see columns 5 and 8 of Tables 12 and 13, technical report).

[14] Evidence on the extent to which charter schools attract disproportionately more or less academically skilled students varies, and likely depends on the specific charter school and context. For some discussion of these issues and related evidence, see

Carnoy, M., Jacobsen, R., Mishel, L. & Rothstein, R. (2005). *The charter school dust-up: Examining the evidence on enrollment and achievement*. New York and Washington DC: Teachers College Press, and the Economic Policy Institute.

Hoxby, C. M. & Murarka, S. (2006) "Comprehensive Yet Simple: Florida's Tapestry of School Choice Programs." In Paul Peterson (Ed.), *Reforming Education in Florida*. Stanford: Hoover Institution Press.

Woodworth, K.R., David, J.L., Guha, R., Wang, H., & Lopez-Torkos, A. (2008). *San Francisco Bay Area KIPP schools: A study of early implementation and achievement. Final report.* Menlo Park, CA: SRI International

Center for Research on Education Outcomes (CREDO) (2009, June). *Multiple Choice: Charter School Performance in 16 States*. Palo Alto: CREDO, Stanford University.

[15] Caroline M. Hoxby, personal communication, 8 October, 2009 and 28 October, 2009.

[16] Woodworth, K.R., David, J.L., Guha, R., Wang, H., & Lopez-Torkos, A. (2008). *San Francisco Bay Area KIPP schools: A study of early implementation and achievement. Final report.* Menlo Park, CA: SRI International

[17] Abdulkadiroglu, A., Angrist J., Cohodes S., Dynarski, S., Fullerton, J., Kane, T., & Pathak, P. (2009). *Informing the debate: Comparing Boston's charter, pilot and traditional schools*. Boston, MA: The Boston Foundation

Skinner, K.J. (2009). Charter school success or selective out-migration of low achievers. Boston: Massachusetts Teachers Association. Retrieved October 25, 2009, from http://www.massteacher.org/news/headlines/charterschools0909.pdf.

[18] According to the report, the figures exclude charter schools whose estimated effects are so imprecise that an effect of 0.10 standard deviations per year would be statistically insignificant at the $\alpha = 0.15$ level *unless* the effects are large enough that they are nonetheless statistically significant at the $\alpha = 0.15$ level (see notes for Figure IVg, p. IV-21 and footnote 8, p. A-1). Hoxby, however, informed me that the text is incorrect: the figures exclude charter schools whose estimated effects are so imprecise that an effect of 0.10 standard deviations per year would be statistically insignificant at the $\alpha = 0.15$ level, regardless of the size of the estimated effect.

[19] Boyd, D., Grossman, P. Lankford, H., Loeb, S., & Wyckoff, J. (2008). "Measuring Effect Sizes: the Effect of Measurement Error." Paper prepared for the National Conference on Value-Added Modeling University of Wisconsin-Madison April 22-24, 2008. Retrieved October 20, 2009, from http://www.teacherpolicyresearch.org/portals/1/pdfs/Measuring%20Effect%20Sizes%20%20the%20Effect%20of%20Measurement%20Error%20Boyd%20et%20al%20%2026Jun2008.pdf. See Table 1.

[20] Boyd, D., Grossman, P. Lankford, H., Loeb, S., & Wyckoff, J. (2008). "Measuring Effect Sizes: the Effect of Measurement Error." Paper prepared for the National Conference on Value-Added Modeling University of Wisconsin-Madison April 22-24, 2008. Retrieved October 20, 2009, from http://www.teacherpolicyresearch.org/portals/1/pdfs/Measuring%20Effect%20Sizes%20%20the%20Effect%20of%20Measurement%20Error%20Boyd%20et%20al%20%2026Jun2008.pdf. See Tables 1 and 2.

[21] Boyd, D., Grossman, P. Lankford, H., Loeb, S., & Wyckoff, J. (2008). "Measuring Effect Sizes: the Effect of Measurement Error." Paper prepared for the National Conference on Value-Added Modeling University of Wisconsin-Madison April 22-24, 2008. Retrieved October 20, 2009, from http://www.teacherpolicyresearch.org/portals/1/pdfs/Measuring%20Effect%20Sizes%20%20the%20Effect%20of%20Measurement%20Error%20Boyd%20et%20al%20%2026Jun2008.pdf. See p.17.