**Ron Zimmer**
**Vanderbilt**

**John Engberg**
**RAND**

# Can broad inferences be drawn from lottery analyses of school choice programs? An exploration of appropriate sensitivity analyses

May 15, 2014

Abstract:

School choice programs continue to be controversial, spurring a number of researchers into investigating and evaluating them. When possible, researchers evaluate these program using randomized designs to eliminate possible selection bias. Randomized designs are often thought of as the gold standard for research, but this approach can have limited inferences in the context of evaluating school choice programs. In this paper, we examine whether these limitations apply to previous evaluations of voucher, charter schools, magnet, and open enrollment programs. We establish the legitimacy of these concerns of inferences and then look at data from an anonymous district to examine whether students admitted to magnet middle schools via lottery have similar student characteristics (including prior achievement and achievement growth) as students admitted outside of a lottery. The point of the analysis is not so much whether these groups are different in our particular case, but that they could be and there are simple sensitivity analyses that researchers could conduct to see if there are reasons to be cautious about the breadth of inferences researcher can make.

Key words: School Choice, Inferences, Sensitive Analyses

Financial support for this research is provided by the Institute of Education Sciences (IES R305A070117 and R305D090016)

## I. Introduction

One of the more controversial topics in education in recent years has been the use of school choice programs to improve the performance of students. Because of this controversy, a number of researchers have tried to examine whether students attending schools of choice outperform students in traditional public schools. However, in any estimate based on students actively making a choice to enroll in a school, researchers worry about selection bias. Some researchers argue that the only way to get valid estimates of these effects is to use students who are randomly assigned to a treatment (i.e., choice school) or control group (i.e., traditional public schools) through a randomized control trial (RCT).

For voucher programs, this often means randomly issuing a voucher to a student and then a student choosing an eligible private school to attend (Greene, 2000; Mayer et al., 2000; Howell and Peterson, 2002; Kruger and Zhu; 2002; Bettinger and Slonim, 2006). For magnet, open enrollment, or charter programs, it often means using a lottery to admit students to *individual oversubscribed schools* (Hoxby & Rockoff, 2004; Dobbie & Fryer, Forthcoming; Hoxby et al., 2009; Abdulkadiroglu et al., 2009; Gleason, et al., 2010; Cullen and Jacob, 2009; Hastings, Kane, and Staiger, 2006; Engberg et al, 2014). Using lottery data from oversubscribed schools has the potential to provide unbiased estimates, but there are questions of whether broad inferences can be drawn from the select set of students entering these programs via lotteries or from the select set of schools that are oversubscribed. For instance, if preferences are based on school quality, it would follow that charter or magnet schools with wait lists would be the best schools.[1] If this is the case, the results would offer limited insight into the performance of undersubscribed schools (Abdulkadiroglu et al., 2009; Bifulco, Cobb, & Bell, 2009; Zimmer,

[1] Cullen, Jacob, and Levitt (2006, p. 1199) show a relationship between the quality of magnet programs and the demand of those programs.

Guarino, and Buddin, 2010; Tuttle, Gleason, & Clark, 2012).[2] [3] In addition, students entering an oversubscribed school because of exemptions (e.g., sibling, neighborhood preference) may be different from students entering a school through a lottery in ways that are related to student outcomes and may limit the inferences within the school. Stakeholders in these debates often make broad inferences from lottery-based studies despite these external validity concerns.

In this paper, we first discuss whether these concerns are relevant to various school choice programs and examine whether these concerns apply to the existing literature. Once we establish the legitimacy of these concerns, we then look at data from an anonymous district to examine whether students admitted to magnet middle schools via lottery have similar student characteristics (including prior achievement and achievement growth) as students admitted outside of a lottery. The point of the analysis is not so much whether these groups are different in our particular case—i.e., we are not trying to make broad inferences from our data. Rather, we are trying to make the conceptual point that oversubscribed schools and undersubscribed schools, as well as students admitted inside and outside of lotteries, could be inherently different; we want to illustrate simple sensitivity analyses that researchers could conduct to see if there are reasons to be cautious about the breadth of inferences researchers can make.

[2] Similarly, families applying for these schools may be more engaged in learning or have high aspirations for their children, so outcomes for these students may provide little information about outcomes for students from families with different motivations. This raises issues of the inferences that can be made from an analysis using lottery data of school choice program if the program expands in scale or scope over time. For instance, voucher programs in Milwaukee, Indiana, and Louisiana had caps on the number of students that could take advantage of these programs (Alliance for School Choice, 2012). Similarly, states often implemented charter schools with caps on the number of schools, but often increase over time (Levin and Belfield, 2002). In some cases, legislation enabling these school choice programs requires an evaluation when these programs are relatively small in scale and raises questions of what inferences could be made to the school choice policies as they expand in size and scope. While this issue is beyond the scope of the current study, it is an issue that needs better recognition as a limitation of these studies and further examination.

[3] Furthermore, questions have been raised about the lotteries themselves. Tuttle et al., (2012) noted that students can be admitted into oversubscribed schools with exemption rules outside of a lottery and it could be difficult to know this in administrative data if strong documentation is not maintained.

II. How Prevalent Are These Issues

Currently, researchers have evaluated a growing but small fraction of school choice programs using an RCT approach. To examine the breadth of inferences from these studies, we first discuss the *policy relevance* of the two external validity threats across the various school choice programs—(1) whether students admitted via randomization have inferences to other students in the same school admitted outside of randomization and (2) whether students admitted to oversubscribed school via randomization have inferences to students in undersubscribed schools. (Hereafter, we refer to these threats as external validity threats (1) and (2)).

**Review of the relevance of external validity threats across school choice program**

For voucher programs, which provide in-kind financial support to families rather than the school directly, RCTs have been used to evaluate both U.S. and international voucher programs (Greene, 2000; Mayer et al., 2000; Howell and Peterson, 2002; Kruger and Zhu; 2003; Bettinger and Slonim, 2006; Wolf, et al., 2010; Angrist et al., 2002; Angrist, Bettinger, & Kremer, 2006; Bettinger, Kremer, and Saavedra, 2006). In these cases, neither external validity threats (1) nor (2) highlighted above apply, as policymakers are primarily interested in the effect voucher use has on students *using the voucher*, not the impact of all students in private schools. As a result, RCT voucher analyses are estimating the impact to the population of greatest interest to policymakers.

In the case of charter schools and magnet programs, which are both public schools of choice (charter schools are autonomous from a district while magnet programs are not), researchers have exploited lottery admission procedures for oversubscribed schools to gain unbiased estimates (Tuttle et al., 2013; Dobbie & Fryer, 2011; Gleason, et al., 2010; Hoxby et al., 2009;

Abdulkadiroglu et al., 2009; Engberg et al, 2014; Ballou, Goldring, and Liu, 2006; Cullen, Jacob, & Levitt, 2006; Hoxby & Rockoff, 2004). However, in these cases, policymakers are interested in the impact of all students in these programs, not just the students attending these schools as result of a lottery admission. Therefore, external validity threats (1) and (2) are policy-relevant concerns.

Like charter schools and magnet programs, researchers have used lottery admission procedures to evaluate open-enrollment programs (Hastings, Kane, & Staiger, 2006; Cullen, Jacob, & Levitt, 2006; Cullen and Jacob, 2009). Open-enrollment programs allow families to choose among different district-run schools, regardless of whether the school is a neighborhood school or not. In this case, policymakers are most interested in students who chose to attend a school other than their assigned school. Therefore, like the voucher studies, the fact that an RCT study of open enrollment program cannot make inferences to students in oversubscribed schools that gain access to the school outside of a lottery is not a major concern from a policy perspective. Therefore, external validity threat (1) does not apply. However, some students can gain access to undersubscribed open-enrollment schools outside of lottery, which raises external validity concerns of the inferences. As a result, the external validity threat (2) is a relevant policy inference concern.

Together, this suggests that external inference threats (1) and (2) are policy-relevant concerns for charter schools and magnet programs; external validity threat (2) is a policy-relevant concern for open enrollment programs; and neither external validity threat (1) nor (2) is a policy relevant concern for voucher programs. However, our discussion so far does not tell us whether the threats are *actually* important in the existing literature. It could be that the vast majority of students enter these schools via lotteries and the vast majority of schools are oversubscribed. To address this question, we examine the existing charter, magnet, and open enrollment literature using lottery data to see how prevalent these concerns are. If there is little evidence that these validity threats are of actual concern, then raising these issues could be more theoretical than practical. Alternatively, if we find that these issues are prevalent, then these threats are both theoretical and practical.

It should be noted that almost any study of school choice programs is going to have an issue of external inferences to other geographic locations because almost all studies of school choice programs are restricted to a single state or district. However, in our review of the existing literature, we focus on whether studies have been able to make inferences to the population of students and schools of choice within the locations the researchers studied. In the two cases in which authors try to make more national claims, we examine whether the study has external validity threats to a national sample. Finally, in our review, we do not include papers examining a single charter school. While these studies can provide important information to a specific charter school and their operational features, they are generally not viewed as important papers for policy debates.

**Existing Studies**

To date, the broadest lottery-based geographic study of charter schools is a national Mathematica evaluation of 36 charter middle schools in 15 states (Gleason, et al., 2010). Although it is nearly impossible to collect up-to-date accurate counts of charter middle schools across the nation, it is safe to assume that 36 charter middle schools only represent a fraction of all charter middle schools across the country. Therefore, external validity threat (2) applies. In examining whether external validity threat (1) applies, we not only examine the original report, but also examine a follow-up study (Tuttle, et al, 2012), which estimates that the average oversubscribed charter school in the study admitted about a third of students through lottery exemptions. In addition, tables in the original report suggest that lottery winners are disproportionally white, and that black students are underrepresented relative to the general population within the same school the student applied: Lottery winners are 60% white, 26% Hispanic, 11% black, and 5% other. In contrast, the school-wide averages of these same schools are 53% white, 26% Hispanic, 16% black, and 5% other. The authors note in an email correspondence that the proportion of lottery winners who score proficient or higher on the state assessment is a bit lower than non-lottery winners in the same schools—about 59-60% of lottery winners scored at the proficient level or higher in math in the first and second follow-up years compared with 65-67% of students in the school as a whole (email correspondence with Phil Gleason, April 4, 2013). Overall, the differences in the racial makeup as well as the proficiency levels suggest that lottery students are different in observed ways and likely in unobserved ways from students in the same oversubscribed schools and therefore, external validity threat (1) applies.

Mathematica also produced a national evaluation of charter middle schools managed by the charter management organization Knowledge is Power Program (or KIPP for short). The authors of the study were able to include 13 out of the 43 KIPP charter middle schools in existence at the beginning of the 2010-11 school year in a lottery-based analysis (Tuttle et al, 2013, p. 5).[4] Therefore, external validity threat (2) applies. In addition, a fair number of students entered oversubscribed KIPP schools outside of the lottery as 389 students gained access to an

[4] Some of the remaining 30 schools were oversubscribed and did have lotteries. A few of these schools declined to participate and for other schools, the research team could not get consent for students to participate in time from their parents. These schools, along with the undersubscribed schools, were evaluated using a matching strategy.

oversubscribed school (Table A.9, p84). This is relative to 535 students the analysis counted as "treatment lottery winners".[5] Therefore, external validity (1) applies.

In the first study to examine charter schools using a lottery data, Hoxby and Rockoff (2004) examined Chicago International Charter School, which had five campuses at the time. The study focused on three of the five campuses as two of these campuses did not have sufficient wait lists to use a randomized design. Also, at that time, 13 other charters had been granted (Illinois State Board of Education, p.7), some with multiple campuses. Together, this suggests that external validity threat (2) would apply. The authors also note that they limited their analysis to students who applied to the charter school from a CPS school, dropping students who applied from non-CPS schools, which is about 35 percent of the sample (Hoxby and Rockoff, 2004, p. 13). In addition, both spring and summer lotteries were held for these schools. Because these populations could be different, the authors included only the spring lottery students in their analysis, which represents about 90 percent of the school slots (p. 11). [6] No other means of admission, such as sibling exceptions, was mentioned by the authors. Assuming the study truly included 65 percent of all students, then external validity threat is only somewhat of a concern and only partially applies. However, if there were other students gaining admission not described in the paper, then the threat would be of a greater concern.

In one of the most cited studies of charter schools, Abdulkadiroglu and colleagues (2009) use lottery data from Boston. In this case, the authors were able not able to evaluate any of the four elementary charter schools, but were able to evaluate five out of 11 middle charter schools and

[5] The definition of lottery winner student is somewhat complicated in the study. Students were considered lottery winners if their parents provided consent to participate in the study and if they were offered admission at the time of the lottery on the basis of the lottery draw. There were some students who won the lottery, but were not counted as lottery winners in the study if their parents did not provide consent. Therefore, some students were not part of the lottery analysis, but also did not gain access to the oversubscribed school outside of the lottery.

[6] They note that their analysis is not sensitive to this restriction.

four out of eight charter high schools within the district (p. 39). Therefore, external validity threat (2) applies, which the authors acknowledge in the paper. From the paper, we cannot discern what proportion of students attend an oversubscribed charter school outside of a lottery, but the authors do note “some lottery losers end up in charter school later, either because they enter the admissions lottery in a future year, gain sibling preference when a sibling wins the lottery, or move off a wait list after the offers coded by our instrument” (p.14). This quote suggests that the external validity threat (1) could apply as well.

In another often cited study of charter schools, Hoxby, Kang, and Murarka, (2009) examined New York City charter schools. Unlike many of these other studies, external validity threats (1) and (2) do not seem to be a problem as the authors note that the study includes 93 percent of all charter students tested within the city (Hoxby, Kang, & Murarka, 2009, p. 6) as 94 percent of all students in charter schools are admitted via a lottery (p 3). In addition, the authors note that nearly all charter schools are routinely oversubscribed, and that only a few charter schools serving disabled students are not routinely oversubscribed (p.2).

In a study of magnet middle school program in an anonymous district using lottery data (and these same data are later used in this study), less than 10 percent of students in magnet middle school programs are admitted via a lottery, and of the twelve middle school magnets operating during this period, two were oversubscribed virtually all the time, and four more were oversubscribed some of the time (Engberg, et al, 2014). However, due to priority status for siblings and certain elementary feeder programs, even the oversubscribed schools admitted only a portion of their students through lottery. Together, this suggests that external validity threats (1) and (2) both apply.

In an often-cited study of 19 magnet high school programs in Chicago by Cullen, Jacob, and Levitt (2006), the authors note that 171 of 375 observed lotteries (schools have more than one lottery per school)[7] did not have any lottery losers (p. 1225), which means that nearly half of the lotteries are for programs not oversubscribed, and suggests that many students may gain access to oversubscribed magnet school outside of a lottery and many magnet schools may not have oversubscribed lotteries. This suggests that both external validity threats (1) and (2) could apply. In a follow-up study of both open enrollment and magnet programs in Chicago elementary using lottery data by Cullen and Jacob (2009), the authors noted that:

> "Given our research design—which involves comparing students who won a lottery with their peers who lost the same lottery—our analysis is necessarily limited to the set of lotteries where there were at least some winners and losers. Among applications to lottery schools, 50.2 percent were to lotteries with both winners and losers, 42.0 percent were to lotteries with no winners and 7.8 percent were to lotteries with no losers. A lottery will not have any winners if the campus is unable to accept applications to a specific grade due to overcrowding. Since we cannot estimate any treatment effects, we exclude applications to both types of degenerate lotteries from our analysis" (Cullen and Jacob, 2009, p. 53).

Again, this suggests that in this analysis, external validity threat (1) (and the only external validity concern for open enrollment program) applies.

A series of studies of Charlotte-Mecklenburg's open enrollment program have received significant attention (Deming, Hastings, Kane, & Staiger, 2011; Hastings and Weinstein, 2008; Hastings, Kane, & Staiger, 2006). Again, because this is an open enrollment policy, only external validity threat (1) is of possible concern. Across the papers, we could not identify the number or proportion of students who gain access to an open-enrollment school outside of a lottery, but some of the papers note that siblings of students already utilizing the open enrollment policy are exempt from the lottery admission process (Deming, et al, 2011, p.5), and only about a

[7] The authors note that there can be multiple lotteries for each magnet high school as there are separate lotteries by year, race, gender, and grades (Cullen, Jacob, and Levitt, 2006, p. 1196).

third of schools across the district are oversubscribed (Hastings, Kane, & Staiger, 2006, p. 7), suggesting that many students would be able to enter a school of choice through the open enrollment program outside of a lottery. Together, this suggests that the external validity threat (1) may apply.

In research evaluating Connecticut's inter-district magnet program, Bifulco, Cobb, and Bell (2009) examined two oversubscribed schools out of 54 magnet schools in operation at the time, which suggests that external validity threat (2) applies (p. 326). Again, while the exact number or proportion of students who gain access to an oversubscribed school is not provided within the study, the authors note that siblings of students already in these schools can gain access to these schools outside of an admission process. In addition, like many admission lotteries, the lotteries for these schools were held by year and by grade and not all grade-year lotteries were oversubscribed and the authors used 14 out of 22 possible grade-year lotteries. Again, this suggests that at least some students gained access to these schools outside of lottery admission process (p. 332). Furthermore, they noted that students enrolling in the magnet schools from non-public schools and students from Hartford were dropped from their analysis, as students who lost the lotteries from private schools were less likely to have observed outcomes later and the Hartford students were all offered admission to the program (p. 332). Together, these suggest that external validity threat (1) applies.

Finally, in an evaluation of magnet middle schools in anonymous southern city, authors Ballou, Goldring, and Liu (2006) note that the percentage of students entering magnet schools outside of a lottery range from 0 to 33 percent, with three out of the four schools having at least some students entering the programs outside of a lottery. This suggests that the external validity threat (1) could be an issue. However, the evaluations included four out of five middle magnet schools within the district, which suggests that external validity threat (2) may be less of an issue.

Table 1 summarizes the external validity threats that apply across the various papers. As a whole, the summary suggests that external validity threats (1) and (2) apply to most studies, including some of the more influential papers.

Table 1 Here

While most of the authors of these studies acknowledge these limitations, Abdulkadiroglu et al. (2009), Bifulco et al (2009), and Tuttle et al. (2013) tried to address the challenge. In these cases, the authors employed an observational analysis to the same set of students included in the lottery-based analysis. Given that they found a fair amount of overlap between the results of the two approaches, they applied the observational approach to non-lottery students with some degree of confidence that it is not producing bias results. While this is an appropriate approach to deal with the two external validity concerns we lay out in our paper, not all studies will overlap in results. In this paper, we lay out an alternative approach to test whether there are “red flags” in broadly interpreting results from an RCT. Again, the point of our analysis is not to make broad inferences from our data of a single district, but to illustrate simple sensitivity analyses researchers could apply when evaluating relevant school choice policies.

## III. Data

To illustrate possible sensitivity analyses that can be utilized to address the two possible external validity threats, we use longitudinal student-level data from the 1997-1998 school year through 2006-2007 school years from an anonymous district. The database includes a unique student ID, which provides a way to track students over time. In addition, we have the students' test scores, schools of attendance, addresses, and demographic data, including race, gender, and free-and-reduced lunch status, as well as whether the student participated and won a binding lottery. Using the addresses, we can include census tract information for each student such as the poverty and adult education levels of their neighborhood. The community characteristics are measured at the census tract level. Poverty is the percentage of adults in each student's census tract with an income level below the poverty line. Education is the percentage of adults in the student's census tract without a high school degree. The database also contains the outcomes of the magnet lotteries and the mechanism through which a student entered a magnet program. To track student performance, the district not only administers the state accountability test, but also other national tests in some grades in which students are not tested for state accountability. In all, we use tests from grades five through eight for students who are in grade eight in 2001-02 through 2006-07, as well as tests from grades five and eight for students who are in grade eight in 2000-01. Because students are administered different tests across grades, we standardized the outcomes by converting all scores into standard Z-scores with a mean of zero and standard deviation of one by test by grade by year.

For our analysis, we focus on students in middle school magnet programs because we eventually want to compare the performance of students entering a magnet program via a lottery versus all other magnet students. Therefore, we need to see students' growth in achievement while attending magnet programs, relative to a baseline test score prior to entering a magnet program.[8] At the elementary level, we would not be able to observe a baseline test score. At the

[8] It should be noted that researchers often use middle schools when examining the effectiveness of school choice programs because elementary schools often have non-random attrition between lottery winners and losers and because there is no baseline test scores for students in elementary schools (Engberg et al., forthcoming; Tuttle et al., 2013).

high school level, there are a limited number of magnet high schools, and we cannot observe student test scores in consecutive years.

Before proceeding to our analysis, we first provide greater details of the lottery admission process that was in effect at the time of our study. Interested students and their parents who lived within the district could submit an application for one magnet program of their choice during a registration period. The number of available slots for magnet students differed both by program and year. Some of the magnet programs had more openings than applications. In these cases, all applications were accepted contingent on the student's eligibility for the grade to which he or she applied. However, if the number of applications submitted during registration for any magnet program exceeded the number of available spaces, a lottery was held to determine the order in which applicants would be offered program placement. In the case of oversubscription, the administration used a computerized random selection algorithm to determine each student's lottery number.

However, before students were admitted based on lottery numbers, students were placed into preference groups. Table 2 lists the preference groups in order. If there were a sufficient number of spaces for all students in the top preference group, then all students in this group would be offered program placement. The district would proceed through the preference groups until it reached the group that contained more applicants than remaining spaces. Within this preference group, spaces would be offered to the students with the lowest lottery numbers. If students rejected offers, then the district would offer spaces to students in an order determined first by their preference group and then by their lottery number. In order to preserve racial balance in the magnet programs, the district would reserve one half of the spaces in each magnet for black students and half for other students. Only when the list of applicants of one race was exhausted would spaces beyond the initial one half of total projected enrollment be offered to the other race. If spaces remained available after the all applicants who had applied on time were offered spaces, then they were filled on a first come, first served basis with applicants from after the registration period.

A lottery analysis of each magnet program would only be able to use students who were in the preference group that contained the last student who was accepted a space based on his or her lottery number. All students in more-preferred groups were offered spaces, and no students in less-preferred groups were offered spaces, eliminating any randomized comparison within these groups.

In Table 2, we describe each preference code. We also give the total number of students in each preference group that entered a magnet program and the number in that group who were subject to a binding lottery. While some preference codes virtually guarantee a spot in a magnet program, others do not. This explains why we observe no students participating in binding lotteries for some preference codes. It should also be noted that the vast majority of students enter a magnet program outside of the lottery admission process, which in itself raises concerns about inferences that could be made from a lottery analysis in our particular case, but again, we are not trying to make inferences from our data. Nevertheless, the previous literature review, illustrates that it is quite possible that a large number of students could enter a school choice program outside of a lottery process; this underscores the care one should take when making inferences from a lottery-based analysis. We explore the inferences both descriptively and with formal analyses in the next section.

Table 2 Here

## IV. Analysis and Results

To address the question of what inferences can be made from an analysis using lottery data, we need to gain a better understanding of whether students admitted or rejected to a choice program via randomization are different from other students in ways that may affect student outcomes. Ideally, across these different populations, we would have a complete set of student characteristics, including measures of student motivation and effort. Unfortunately, like researchers in general, we have a limited set of student characteristics in our administrative data set. Nevertheless, we can get a sense of whether there are differences between these populations by examining the observable characteristics of students, which is presented in Table 3. In the table, we do pairwise comparisons between students included in a lottery analysis versus students who enter a magnet school through alternative means. These comparisons are analogous to randomized design studies that do "balance checks" of observable characteristics of students in order to provide insight into whether the samples of treatment and control students have been randomly assigned (Hoxby & Rockoff, 2004; Hoxby et al., 2009; Abdulkadiroglu et al., 2009; Cullen and Jacob, 2009; Engberg et al, 2014; Bifulco, 2012). While finding no significant difference among observable characteristics cannot exclude the possibility of unobservable differences in populations, finding statistically significant differences among a number of different observable student characteristics would suggest that these populations are likely to have unobservable differences as well.

Table 3 Here

The pairwise comparisons indicate that two out of seven characteristics are statistically different—more than one would expect by chance. Students entering magnet programs via a lottery are disproportionally black and from neighborhoods with fewer high school dropouts relative to other magnet students. Together, these observable differences raise some concerns that using an analysis with a lottery design might have limited inferences to other magnet students.

However, to more fully address our original question of whether we can make inferences from a lottery analysis of a magnet program relative to other magnet students, we examine differences in this set of students who won admission into a magnet program via a lottery relative to other magnet students while including multiple student characteristics simultaneously. One of the key factors we want to examine is whether there are differences in achievement patterns of lottery winners versus all other magnet students. If there are differences among these populations, caution is warranted in making broad interpretations of these results.

To formally examine whether lottery students are unique from non-lottery students, we restrict the sample to magnet students only and run the following probit model:

$$\Pr(y_{it}=1) = \mathrm{F}(\, TS_{i0}\ \delta + X_{it}\ \beta + \varphi_t + \gamma_{it})$$

where the dependent variable is a dichotomous indicator of whether a student won a binding lottery (i.e., a lottery that determines admission into the program) where $y_i = 1$ if the i$^{th}$ student won a binding lottery in the t$^{th}$ year and $0$ otherwise; $TS_{i0}$ is vector of baseline test scores in math and reading of the i$^{th}$ student where baseline is defined as the year before a student enters a magnet program; the vector $X_{it}$ is student characteristics for the i$^{th}$ student in year t including whether the student is black interacted with free-and-reduced lunch (FRL) status, non-black interacted with FRL status, gender, census tract poverty education and poverty measure; and year and school fixed effects are $\varphi$ and $\gamma$, respectively. For the analysis, we restrict the sample to the entry grade of 6th grade to avoid double counting of students across years.

We chose to use black and non-black as our only racial categories because the district has separate lotteries for these two groups of students. In addition, the district effectively has only two racial groups—black (49 percent) and white students (46 percent)—with only 5 percent falling outside of these racial groups (and nearly 3 percent of these students are mixed race students). Therefore, we lump all other students in with white students to create the non-black variable. We included the interactions of race and FRL status to gain a more fine-grained depiction of the likelihood of winning a binding lottery and student characteristics. With these interactions, the omitted category is non-black, non-FRL magnet students and all estimates are relative to this group. To simplify interpretation of the probit estimates, we convert the coefficient outcomes into marginal probabilities with changes in probabilities at the mean values for continuous variables and the change from 0 to 1 for discrete variables.

The results are presented in Table 4. First, it is important to highlight that the baseline test scores of students do not significantly predict whether a student wins admission into a magnet program via binding lottery controlling for other observable characteristics. However, relative to all non-FRL, non-black magnet students, black students (regardless of whether the student is a FRL student) are about 5 to 6 percent more likely to be a winner in a binding lottery. In addition, neighborhood poverty rate is negatively associated with winning a lottery.

Table 4 Here

While the insignificant differences in baseline achievement reduces concerns about limitations in achievement inferences from magnet students included in our lottery analysis to broader set of magnet students, the differences in socioeconomic characteristics of students suggests that the students included in the lottery analysis are not a completely random set of students and the lottery students could be different in ways that may affect an analysis of *achievement growth* while attending a magnet program. To further explore whether there are differential value added of students who are admitted via a binding lottery versus other magnet students, we conduct two additional analyses.

First, we examine the gains of students in math and reading using a school-fixed effect model. The model is a *within school analysis* in which we compare the value added growth of students who won lottery admission into an oversubscribed school to other magnet students within these same schools. This model allows us to gain insights into whether students included in a lottery analysis in oversubscribed schools could produce inferences that generalize to other students who enter oversubscribed schools through an alternative route. We first run this analysis without controls for student observable characteristics (but do include grade and year fixed effects to control for nuanced differences in test scores across years and grades). We run the model without controls to see whether students who enter a magnet program via a lottery have differential achievement growth, regardless of whether the difference is explained by observable or unobservable characteristics of the student. Fundamentally, we want to know whether students attending a magnet school via a lottery are different from other students—in this case, students within the same school. If they are different, regardless of whether the differences are because of observable or unobservable characteristics, it suggests that inferences to a broader population are limited.

However, we also run a "full" school fixed effect model, which includes observable student characteristics, including the same set of observable student characteristics as equation 1 as well as grade and year fixed effects. This model examines whether any differences observed in the school fixed effect model without controls can be explained by observable characteristics. In both the model with and without controls, we run the analyses separately for math and reading. Across the models, current year test scores are the dependent variables and achievement growth for the student is measured by including baseline math and reading test scores where baseline is defined as the year before a student enters a magnet program. In the full model, a vector of student characteristics are included, which includes whether the student is black interacted with free-and-reduced lunch (FRL) status, non-black interacted with FRL status, gender, census tract poverty and dropout measures.

The results of the analyses are presented in Table 5. In examining the results *without the controls* for observable characteristics first (columns 1 and 3), we see that that there are differences in the achievement growth of lottery winners versus all other students within the same school. Lottery winners have smaller growth rate of .07 and .09 of a standard deviation in math and reading, respectively, both of which are significant at the 5 percent level. This provides some evidence that students entering a choice school via a lottery may not have strong inferences for non-lottery students in these same schools. In fact, since lottery students represent less than a majority of students within the oversubscribed schools, the lottery analysis will have limitations in comprehensiveness of the inferences for these schools.

In examining the results controlling for student characteristics (columns 2 and 4), we can see whether controlling for observable characteristics explains away these differences. The coefficients do shrink with lottery winners having an estimated smaller achievement growth of .05 and .06 of a standard deviation for math and reading, respectively. The coefficient in the reading regression remains statistically significant at the 10 percent level, suggesting that there are likely to be differences in achievement growth beyond observable characteristics. These differences are most likely the result of unobservable student characteristics as it unlikely that these achievement differences could be explained by differences in the impact of schooling since the analysis is comparing students within the same schools receiving the same instruction, curriculum, etc. These analyses suggest that there are both observable and unobservable differences between magnet students included in a lottery analysis relative to other students within the same schools and raises serious concerns about the inferences that can be made from a lottery analysis to the school as a whole, at least for this district.

Table 5 Here

To address the question of whether oversubscribed choice schools could have inferences to undersubscribed choice schools, we conduct a second set of analyses in which we drop the school fixed effects, and exclude students who enter an oversubscribed school through a mechanism other than a lottery. By dropping the school fixed effect, we are no longer conducting a within school analysis, which allows us to examine whether students in under and oversubscribed schools have differential achievement growth. As with the within-school analyses, we run the models with and without student observable controls. As shown in Table 6, whether we control for student observable characteristics or not, the results suggest that lottery winners attending oversubscribed magnet programs do not have significantly different achievement growth than students in undersubscribed schools. These results do not raise any red flags of the inferences from oversubscribed schools to undersubscribed schools.

Table 6 Here

## V. Conclusions

Over the last decade or so, there has been a strong push by government funding agencies and foundations like the Institute of Education Sciences and the William T. Grant Foundation to use randomized control trials to evaluate education policies, including school choice programs such as vouchers, charter, and magnet schools. A number of researchers have embraced this approach and argue that randomized design provides the best guard against selection bias in school choice programs (Hoxby and Murarka, 2006). While it is true that randomized designs such as using lottery data from oversubscribed schools can provide strong internally valid outcome estimates for students attending schools via a lottery, it may have limitations in the inferences to external students. First, it may have weak inferences to students within the same school that were admitted outside of the lottery. These students could be admitted for a variety of reasons including having a sibling at the same school, neighborhood preferences, or feeder patterns exemptions. These students may be different in both observable and unobservable ways from lottery students and could lead to differential outcomes. Second, while we did not find evidence for this being true in our case, an analysis of lottery students in oversubscribed schools may have weak inferences for students in undersubscribed schools, as these schools may differ in their operational and instructional practices, and the students attending these schools could once again be different in both observable and unobservable ways.

This study sheds light on this issue by not only highlighting these issues as possible concerns in the existing literature and by examining whether there are differences among these populations, but also illustrates some sensitivity analyses that researchers could employ. Specifically, researchers could examine differences in achievement growth: (1) between students attending a school of choice via a lottery versus other students within the same schools; and (2) between students attending an oversubscribed school versus undersubscribed school. Neither of these analyses can definitively address whether broad inferences can be made, but if achievement differences are found, it may raise red flags on the breadth of such inferences.

References:

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,” NBER Working Paper, 15549.

Alliance for School Choice (2012). School Choice Now: The Year of School Choice School Choice Yearbook 20111-12. Washington, D.C.

Angrist, J., Bettinger, E., Bloom, E., Kremer, M., & King, E. (2002). The Effects of School Vouchers on Students: Evidence from Colombia. *American Economic Review 92*(5), 1535–1558.

Angrist, J., Bettinger, E., & Michael Kremer (2006), "Long-Term Educational Consequences of Secondary School Vouchers: Evidence from Administrative Records in Colombia," *American Economic Review 96*(3): 847-862.

Ballou, D., Goldring, E., & Liu, K. (2006). “Magnet Schools and Student Achievement” Working Paper, Vanderbilt University, retrieved May 1, 2013 at http://ncspe.org/publications_files/OP123.pdf

Bettinger, E., Kremer, M., & Saavedra, J. (2006) "How Educational Vouchers Affect Students – The Role of Vocational High Schools in Colombia." Case Western Reserve mimeo.

Bettinger, E. & Slonim, R. (2006), “Using Experimental Economics to Measure the Effects of a Natural Educational Experiment on Altruism," *Journal of Public Economics 90*(8-9): 1625-1648.

Bifulco, R. (2012). “Can Nonexperimental Estimates Replicate Estimates Based on Random Assignment in Evaluations of School Choice? A Within-Study Comparison. *Journal of Policy Analysis and Management*, 31(3): 729-751.

Bifulco, R., Cobb, C., & Bell, C. (2009). “Can Interdistrict Choice Boost Achievement: Evidence from Connecticut’s Interdistrict Magnets”. *Educational Evaluation and Policy Analysis*, 323-345.

Cullen, J.B. & Jacob, B.A (2009). “Is Gaining Access to Selective Elementary Schools Gaining Ground? Evidence from Randomized Lotteries,” in The Problems of Disadvantage Youth: An Economic Perspective, Jonathan Gruber, editor, University of Chicago Press, 43-84.

Cullen, J.B, & Jacob, B.A, & Levitt, S. (2006). “The Effect of School Choice on Participants: Evidence From Randomized Lotteries” *Econometrica*, 74(5), 1191-1230.

Deming, D., Hastings, J., Kane, T., & Staiger, D. (2011). "School Choice, School Quality and Academic Achievement. Retrieved July 5, 2013 at http://aefpweb.org/sites/default/files/webform/DHKS_March2011.pdf

Dobbie, Will and Roland G. Fryer. (2010). "Are High-Quality Schools Enough to Close the Achievement Gap? Evidence from a Bold Social Experiment in Harlem." *American Economic Journal: Applied Economics*, forthcoming.

Engberg, J, Epple, D., Imbrogno, J., Sieg, H., & Zimmer, R. (2014) "Identification and Estimating of Treatment Effects when Program Participation is Partially Determined by Lotteries: The Case of Magnet Programs", *Journal of Labor Economics*, Vol. 32(1): 27-63.

Gleason, P., Clark, M., Tuttle, M.C., & Dwoyer, E. (2010). *The Evaluation of Charter School Impacts: Final Report* (NCEE 2010-4029). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education, June 2010.

Greene, Jay P. (2000). *The Effect of School Choice: An Evaluation of the Charlotte Children's Scholarship Fund*. New York: Center for Civic Innovation at the Manhattan Institute.

Hastings, J.S., & Weinstein, J.M. (2008). "Information, School Choice, and Academic Achievement: Evidence from Two Experiments." *Quarterly Journal of Economics*, 123(4): 1373-1414

Hastings, J.S., Kane, T.J.,, & Staiger, D.O. (2006). "Gender and Performance: Evidence from School Assignment by Randomized Lottery," *American Economic Review Papers and Proceedings*, 96(2):232-236

Hastings, J.S. Kane, T., & Staiger, D (2006). "Preferences and Heterogeneous Treatment of Effects in a Public School Choice Lottery," NBER Working Paper 12145.

Howell, W. G., & Peterson, P. E. (2002). *The education gap: Vouchers and urban schools.* Washington, DC: Brookings Institute.

Hoxby, C.M., & Rockoff, J. (2004). "The Impact of Charter Schools on Student Achievement," Working Paper, Harvard University.

Hoxby, C.M., & Murarka, S. (2006). *Methods of Assessing the Achievement of Students in Charter Schools*, delivered at the National Conference on Charter School Research, Vanderbilt University, Nashville, Tenn., As of November 5, 2008: http://www.vanderbilt.edu/schoolchoice/conference/papers/Hoxby-Murarka_2006-DRAFT.pdf

Hoxby, C., Kang, J., & Murarka, S. (2009). Techinical Report: *How New York City's charter schools affect achievement*, New York City Charter Schools Evaluation Project. Retrieved July 5, 2013 at:

http://users.nber.org/~schools/charterschoolseval/how_nyc_charter_schools_affect_achievement_technical_report_2009.pdf

Illinois State Board of Education, 2009-2010 and 2010-2011 Illinois Charter School Biennial Report. Retrieved July 3, 2013 from http://www.isbe.state.il.us/charter/pdf/biennial_rpt_09-10_10-11.pdf

Krueger, A.B., & Zhu, P. (2002). "Another Look at the New York City School Voucher Experiment," Working Paper 9418 Cambridge, MA: National Bureau of Economic Research.

Levin, H., & Belfield, C. (2002) "The Effects of Competition on Educational Outcomes: A Review of U.S. Evidence," with C. Belfield, *Review of Educational Research,* 72, (2): 279-341.

Meyers, D., Peterson, P., Mayer, D., Chou, J, & Howell, W.G. (2000). *School Choice in New York City After Two Years: An Evaluation of the School Choice Scholarship Program .* Washington, DC: Mathematica Policy Research.

Tuttle, C., Gleason, P., & Clark, M. (2012). "Using lotteries to evaluate schools of choice: Evidence from a national study of charter schools. *Economics of Education Review*, 31: 237-253.

Tuttle, C., Gill, B., Gleason, P., Knechtel, V., Nichols-Barrer, I., & Resch, A. (2013). *KIPP Middle Schools: Impacts on Achievement and Other Outcomes*. Mathematica Policy Research.

Wolf, P., Gutmann, B., Puma, M., Kisida, B., Rizzo, L., Eissa, N., & Carr, M. (2010). *Evaluation of the DC Opportunity Scholarship Program: Final Report*, NCEE 2010-4018, U.S. Department of Education, Institute for Education Sciences, National Center for Education Evaluation and Regional Assistance (Washington, DC: U.S. Government Printing Office) http://ies.ed.gov/ncee/pubs/20104018/pdf/20104018.pdf

Zimmer, R., Guarino, C., & Buddin, R. (2010). "School Choice: Options and Outcomes," book chapter in *Urban and Regional Policy and Its Effect*, *Vol. 3,* edited by N. Pindus and H. Wolman, Brookings Institution.

Table 1
Summary of the Literature

| Study | External Validity Threat (1): | External Validity Threat (2): |
|---|---|---|
| National Charter School Study (Gleason et al., 2010) | Yes | Yes |
| KIPP Study (Tuttle et al., 2013) | Yes | Yes |
| Chicago Charter Schools (Hoxby and Rockoff, 2004) | At least somewhat of a concern | Yes |
| New York Charter Schools (Hoxby, Kang, & Murarka, 2009) | No | No |
| Boston Charter Schools (Abdulkadiroglu, et al., 2009) | At least somewhat of a concern | Yes |
| Anonymous district magnet program (Engberg et al., 2014) | Yes | Yes |
| Chicago magnet program (Cullen, Jacob, & Levitt, 2006) | Yes | Yes |
| Chicago open enrollment program Cullen and Jacob (2009) | Yes | NA |
| Charlotte-Mecklenburg open enrollment (Deming, Hastings, Kane, & Staiger, 2011; Hastings and Weinstein, 2008; Hastings, Kane, & Staiger, 2006) | At least some evidence it might apply | NA |
| Conneticut interdistrict magnet program (Bifulco, Cobb, & Bell, 2009) | Yes | Yes |
| Anonymous district magnet program (Ballou, Goldring, & Liu, 2006) | Not a major concern | Yes |

Table 2: Different Ways for Entering a Magnet Program

| Mechanism for Entering a Magnet Program | Total Number of Entering Students | Number of Students Entering through Lottery |
|---|---|---|
| Students in an elementary magnet program that feeds into the magnet program for which they apply | 704 | 0 |
| Sibling of a student currently in the same magnet program | 120 | 0 |
| Neighborhood student (i.e., student within 1.5 miles range of a school) | 7 | 0 |
| Feeder Pattern (i.e., students who live in the magnet's traditional residential feeder pattern) | 11 | 0 |
| Regional (i.e., students who live in the third of the district that contains the magnet) | 492 | 75 |
| General Registration (i.e., students that apply to a magnet program but don’t have any special priority for the magnet to which they are applying) | 371 | 133 |
| Post-Registration students (i.e., students that submit their application after the deadline) | 152 | 0 |
| Magnet applicants who attend a different magnet | 1,457 | 0 |
| Total magnet students | 3,314 | 208 |

Table 3

| Observable Characteristic | Lottery Winner Magnet Students | All Other Magnet Students |
|---|---|---|
| Baseline Math Test Scores | 0.07 | 0.08 |
| Baseline Reading Scores | 0.07 | 0.09 |
| Free-and-Reduced Lunch | 0.34 | 0.47 |
| Female | 0.55 | 0.53 |
| Black | 0.72 | 0.54*** |
| Neighborhood Poverty | 0.22 | 0.22 |
| Neighborhood Dropout Rate | 0.19 | 0.20** |
| N | 208 | 3,314 |

***indicates statistical difference at the 1 percent level in the pairwise comparison between lottery magnet students and non-lottery magnet students

**indicates statistical difference at the 5 percent level in the pairwise comparison between lottery magnet students and non-lottery magnet students

*indicates statistical difference at the 10 percent level in the pairwise comparison between lottery magnet students and non-lottery magnet students

Table 4: Examination of differences of magnet students participating in a binding lottery versus all other magnet students

| Observable Characteristics | Marginal Probability (Standard Error) |
|---|---|
| Math Baseline Score | .006 (.006) |
| Reading Baseline Score | -.005 (.006) |
| Female | .005 (.006) |
| Non-Black Free and Reduced Lunch Student | -.013 (.019) |
| Black non Free and Reduced Lunch Student | .057*** (.011) |
| Black Free and Reduced Lunch Student | .054*** (.012) |
| Neighborhood Poverty Rate | -.068* (.039) |
| Neighborhood Dropout Rate | -.019 (.056) |
| Grade Fixed Effect | Yes |
| School Fixed Effect | Yes |
| Sample Size | 3,502 |
| Pseudo R2 | 0.20 |

***Indicates significance at the 1 percent level
*Indicates significance at the 10 percent level

Table 5: Examination of the value added scores of lottery magnet and non-lottery magnet students within schools

| Variable | Math | | Reading | |
|---|---|---|---|---|
| | Without Controls | With Controls | Without Controls | With Controls |
| | (1) | (2) | (3) | (4) |
| Lottery Winners | -.07** (.03) | -.05 (.03) | -.09** (.04) | -.06* (.03) |
| Math Baseline Score | .66*** (.01) | .63*** (.01) | .28*** (.013) | .25*** (.01) |
| Reading Baseline Score | .22*** (.01) | .20*** (.01) | .54*** (.01) | .50*** (.01) |
| Female | | -.03* (.02) | | .09*** (.01) |
| Non-Black Free and Reduced Lunch Student | | -.11*** (.03) | | -.14*** (.03) |
| Black non Free and Reduced Lunch Student | | -.19*** (.02) | | -.20*** (.02) |
| Black Free and Reduced Lunch Student | | -.20*** (.02) | | -.22*** (.02) |
| Neighborhood Poverty Rate | | -.07 (.08) | | -.07 (.08) |
| Neighborhood Dropout Rate | | -.47*** (.11) | | -.46*** (.12) |
| Grade Fixed Effect | YES | YES | YES | YES |
| Year Fixed Effect | YES | YES | YES | YES |
| School Fixed Effect | YES | YES | YES | YES |
| Sample Size | 10,480 | 10,480 | 10,475 | 10,475 |
| R2 | .64 | .65 | .61 | .60 |

Robust Standard Errors in Parentheses
***Indicates significance at the 1 percent level
**Indicates significance at the 5 percent level
*Indicates significance at the 10 percent level

Table 6: Examination of the value added scores of lottery magnet and non-lottery magnet students across schools

| Variable | Math | | Reading | |
|---|---|---|---|---|
| | Without Controls | With Controls | Without Controls | With Controls |
| | (1) | (2) | (3) | (4) |
| Lottery Winners | -.02<br>(.04) | -.01<br>(.04) | -.02<br>(.05) | -.02<br>(.04) |
| Math Baseline Score | .66***<br>(01) | .63***<br>(.02) | .28***<br>(.02) | .26***<br>(.02) |
| Reading Baseline Score | .19***<br>(.02) | .18***<br>(.02) | .55***<br>(.02) | .52***<br>(.02) |
| Female | | -.02<br>(.02) | | .12***<br>(.02) |
| Non-Black Free and Reduced Lunch Student | | -.06*<br>(.03) | | -.15***<br>(.03) |
| Black non Free and Reduced Lunch Student | | -.14***<br>(.03) | | -.15***<br>(.03) |
| Black Free and Reduced Lunch Student | | -.13***<br>(.03) | | -.20***<br>(.03) |
| Neighborhood Poverty Rate | | -.13<br>(.10) | | -.02<br>(.10) |
| Neighborhood Dropout Rate | | -.20<br>(.14) | | -.49***<br>(.15) |
| Grade Fixed Effect | YES | YES | YES | YES |
| Year Fixed Effect | YES | YES | YES | YES |
| School Fixed Effect | NO | NO | NO | NO |
| Sample Size | 6,356 | 6,356 | 6,354 | 6,354 |
| R2 | .60 | .61 | .56 | .57 |

Robust Standard Errors in Parentheses
***Indicates significance at the 1 percent level
**Indicates significance at the 5 percent level
*Indicates significance at the 10 percent level