Technical Report

THE EFFECTS OF THE LOUISIANA SCHOLARSHIP PROGRAM ON STUDENT ACHIEVEMENT AFTER THREE YEARS

EDUCATION RESEARCH ALLIANCE FOR NEW ORLEANS

SCHOOL CHOICE DEMONSTRATION PROJECT

Jonathan N. Mills & Patrick J. Wolf

June 26, 2017

EducationResearchAllianceNOLA.org
UAedreform.org/school-choice-demonstration-project

THE EFFECTS OF THE LOUISIANA SCHOLARSHIP PROGRAM ON STUDENT ACHIEVEMENT AFTER THREE YEARS

Jonathan N. Mills
Department of Education Reform
University of Arkansas
Fayetteville, AR 72701
jnm003@uark.edu

Patrick J. Wolf
Department of Education Reform
University of Arkansas
Fayetteville, AR 72701
pwolf@uark.edu

Louisiana Scholarship Program Evaluation Report #7
June, 2017

School Choice Demonstration Project, University of Arkansas, Fayetteville, AR
Education Research Alliance for New Orleans, Tulane University, New Orleans, LA

Acknowledgments
We thank the Smith Richardson Foundation for financial support for this research. We also thank Jay P. Greene, Gema Zamarro, Brian Kisida, Joshua Cowen, David Figlio, Brian Gill, Celeste Carruthers, and Nathan Barrett for their comments on previous drafts as well as the Louisiana Department of Education for their cooperation and assistance with providing the necessary data to conduct the analyses. Douglas Harris has been particularly helpful in the development and organization of this report. The content of the report is solely the responsibility of the authors and does not necessarily represent the views of the Smith Richardson Foundation, University of Arkansas, Tulane University, or the Louisiana Department of Education.

* Corresponding author. Address 201 Graduate Education Building, Fayetteville, AR 72703; Tel.:1-479-575-3172; *E-mail address*: jnm003@uark.edu

Abstract

The Louisiana Scholarship Program (LSP) offers publicly-funded vouchers to students in low-performing schools with family income no greater than 250 percent of the poverty line, allowing them to enroll in participating private schools. Established in 2008 as a pilot program in New Orleans, the LSP was expanded statewide in 2012. This report examines the experimental effects of using an LSP scholarship to enroll in one's first choice private school on student achievement in the three years following the program's expansion. Large negative achievement effects in the first year of the program appear to have been followed by improvement in the second and third years. Based on our primary analytic sample, the effects of the LSP on English Language Arts (ELA) are positive and math are negative in Year 3, but neither is statistically significant. These results are partially reflective of declining statistical power and appear to be influenced by the return of students to public schools for whom the program was not working. Subgroup analyses indicate that students with lower ELA scores at baseline realized statistically significant achievement gains in ELA from the program, while students applying to the earlier elementary grades experienced large achievement losses from the program in math.

Keywords: school vouchers, school choice, student achievement, randomized control trial

The Effects of the Louisiana Scholarship Program on

Student Achievement after Three Years

The Louisiana Scholarship Program (LSP) is a statewide school voucher initiative providing public funds for low-income students in underperforming public schools to attend participating private schools.[1] Originally piloted in New Orleans in 2008, the statewide expansion of the LSP in 2012-13 allowed almost 5,000 low- to moderate-income students across Louisiana to transfer out of their traditional public schools and into private schools. The evidence presented here examines how the LSP has impacted student achievement for the 2012-13 application cohort three years after the statewide expansion.

Our analysis uses oversubscription lotteries for nearly 10,000 eligible applicants to estimate the achievement impacts of LSP as a randomized control trial (RCT). Admission lotteries are used as instrumental variables to estimate the effect of using an LSP scholarship to enroll in one's first-choice, or top-ranked[2], private school for applicants induced to attend a private school as a result of winning the lottery. Our analysis uses student-level data obtained via a data-sharing agreement with the state of Louisiana. Achievement is measured by student performance on the criterion-referenced tests mandated by the state for public school accountability purposes.

Our analysis indicates:

- The immediate impact of participating in the LSP was large negative achievement effects, especially in math, in the first year after random assignment;

[1] Originally called the Student Scholarships for Educational Excellence.

[2] Eligible LSP applicants were allowed to submit up to five rank-ordered private school preferences. We focus on first-choice school lotteries to ensure independence of treatment assignment, as whether or not a student won a lottery for placement in a lower-choice school likely was influenced by factors such as the number and popularity of non-first-choice schools listed which could bias comparisons of "any-lottery winners" with "no-lottery winners." Given evidence suggesting over-subscribed schools tend to be better performing (Abdulkadiroglu, Angrist, Dynarski, Kane, & Pathak, 2011) as well as the likelihood that first-choice schools are popular schools, it is likely the effects presented here are upper bound estimates of the impact of LSP scholarship usage.

- Those initial negative test score effects attenuated somewhat in the second year, especially in math, but remained statistically significant;
- Three years after random assignment, the average test scores of program participants were statistically similar to those of the experimental control group when controlling for baseline achievement, with small positive impact estimates for English Language Arts (ELA) achievement and negative effects for math;
- The statistical similarity between the average test scores of LSP participants and control group students in the third year of the evaluation is partly due to a reduction in the gap between the average scores of the two groups and partly due to an increase in the variability surrounding those average scores;
- The variability, or statistical noise, surrounding our estimates of the test score impacts of the LSP increased in the third year of our analysis due to a smaller sample size, as more students in our study panel aged out of the grade range for testing, and because Louisiana changed the outcome test used in ELA and math for accountability purposes from the LEAP/iLEAP to the PARCC.

Our study indicates that the immediate effects of the Louisiana Scholarship Program on student test scores was negative but that the intermediate effects, after three years, are inconclusive and might reasonably be null or even positive given the high level of statistical uncertainty involved. These effects are not differentiated by gender or race; however, we find evidence of positive ELA impacts among the lowest performers at baseline. While not conclusive, the pattern of results from our study suggests the initial negative impacts of the program may be dissipating over time, especially in math.

The report proceeds as follows. In the next section, we provide a brief background on vouchers as a policy instrument in K-12 education and summarize the evidence of their effects on student achievement drawn from prior random assignment studies. We then describe the LSP and the lottery process that enabled the experimental analysis. Next we discuss the data and analytical strategy used to estimate the participant effects of the first two years of the statewide expansion of the LSP. We then present the results of our analysis and conclude with a discussion of our findings.

School Vouchers and K-12 Education.

School vouchers provide government resources to families to attend a private school of their choosing (Wolf, 2008). While voucher programs can be universal, most are limited to disadvantaged students. Strictly speaking, a private school choice initiative is only a "voucher" program if the government funds the program directly through an appropriation. Other private school choice programs are funded indirectly, through tax credits provided to businesses or individuals who contribute to nonprofit scholarship-granting organizations, or privately through charitable contributions. Since these tax-credit and privately funded scholarship programs accomplish the same general purpose as voucher programs we treat all types of private school choice programs as functionally equivalent in this report, although we label specific initiatives appropriately when discussing them.

While economist Milton Friedman (1955) introduced the idea of education vouchers in the U.S., the theoretical support for their desirability dates back to political philosophers Thomas Paine (1791) and John Stuart Mill (1962 [1869]). School voucher theory holds that government should provide funds supporting compulsory education but need not necessarily deliver the schooling itself (Friedman, 1955). Vouchers are expected to benefit individual students by better facilitating matches of school programs and student academic needs and by increasing the competitive pressures schools face in the broader education system (Moe, 2005). The extent to which students benefit from vouchers, however, is an empirical question (Doolittle & Connors, 2001). Experimental design is critical in school voucher evaluations as the potential for motivated and able families to self-sort into private schools generates concerns of selection bias (Murnane, 2005). Fortunately, much of the research on school vouchers in the U.S. has been experimental.

Prior Experimental or Rigorous Quasi-Experimental Evaluations of School Vouchers

Prior rigorous empirical studies of the effects of school vouchers on participants' achievement have not produced a scholarly consensus on how vouchers impact students' academic outcomes (Wolf, 2008; Barrow & Rouse, 2008). A total of 17 analyses have applied experimental, regression discontinuity design (RDD), or reliable student matching methods to data from voucher and voucher-type scholarship programs in Charlotte, Dayton, the District of Columbia, Florida, Milwaukee, New York, and Louisiana to determine their impacts on student achievement. Test-score results from experimental and rigorous quasi-experimental voucher studies are almost equally divided between findings of modest positive effects and findings of no statistically significant difference. A recent meta-analysis of the experimental evaluations of U.S. programs reports that the average effect of private school choice on student test scores is a gain of .08 standard deviations in reading and .07 standard deviations in math, neither of which is statistically significant with 95% or greater confidence (Shakeel, Anderson & Wolf, 2016).

Some studies report significant positive findings of vouchers overall. Both analyses of the Charlotte data find that the privately-funded scholarship program produced positive and statistically significant achievement impacts (Greene, 2001; Cowen, 2008). Two early experimental evaluations of the Milwaukee Parental Choice (voucher) Program report statistically significant gains in mathematics (Greene, Peterson, & Du, 1999; Rouse, 1998). Greene et al. (1999) additionally report modest positive reading effects.

Program effects often vary over time. An evaluation of the privately-funded Washington Scholarship Fund in DC found that initial achievement gains disappeared in the third and final years of the study (Howell & Peterson, 2006). A later evaluation of the District of Columbia Opportunity Scholarship (voucher) Program, reported significant positive impacts in reading after three years (Wolf et al. 2009, p. 36) that were only significant at a 94 percent level of confidence in the fourth and final year of the study (Wolf et al., 2013). A recent evaluation of the Milwaukee voucher program concluded that a combination of the choice program and a high-stakes testing policy generated test score gains in reading only in the study's fourth and final year (Witte et al. 2014). A more recent experimental analysis of the DC Opportunity Scholarship Program finds, however, statistically significant negative impacts in mathematics one year after receiving a scholarship (Dynarski, Rui, Webber, & Gutmann, 2017).

Most experimental evaluations report evidence of effect heterogeneity though the source of variation in effects is not consistent. Wolf et al. (2013) find that students with higher previous performance, students applying from public schools not classified as "in need of improvement", and females disproportionately benefitted from voucher receipt. A study of the privately-funded Parents Advancing Choice in Education Scholarships in Dayton, OH, reports positive findings for African American students. Similarly three of five evaluations of the New York City voucher program report significant positive effects for African American students (Barnard, Frangakis, Hill, & Rubin, 2003; Howell & Peterson, 2006; Jin, Barnard, & Rubin, 2010). A fourth study by Krueger and Zhu (2004), which uses a unique method for classifying students as African

American, finds no evidence of significant achievement gains, overall or for any participant subgroup. A fifth study concludes the New York City program had no clear effects for subgroups along the achievement distribution (Bitler, Domina, Penner, & Hoynes, 2015). Finally, a regression discontinuity design (RDD) analysis of the tax-credit scholarship program in Florida finds that students near the income eligibility cutoff experienced clear achievement gains in reading, but not necessarily in mathematics, due to the program (Figlio, 2011).

The pattern of results from previous experimental, RDD, and rigorous quasi-experimental evaluations of voucher programs outside of Louisiana has ranged from neutral to positive, with few studies reporting significant negative impacts on student achievement.[3] In contrast, two recent evaluations of the Louisiana Scholarship Program report statistically significant negative impacts of voucher usage on student achievement in reading, math, science, and social studies (Abdulkadiroglu, Pathak & Walters, 2016; Mills, 2015). Both studies only examine student outcomes in the first year of statewide implementation of the Louisiana voucher program, with students tested eight months after switching to a participating private school. The present study, in contrast, includes two additional years of student achievement outcome data, thereby allowing for a more comprehensive picture of the effects of the program on short-run outcomes.

[3] The lone exception is Dyrnarski et al.'s (2017) recent experimental evaluation of the DC Opportunity Scholarship Program, which finds significant negative impacts on mathematics achievement one year after receiving a scholarship randomly via lottery.

Description of the Intervention

The Louisiana Scholarship Program (LSP) is a statewide school voucher initiative available to moderate- to low-income students in low-performing public schools. The program is limited to students (1) with family income at or below 250 percent of the federal poverty line attending a public school that was graded C, D, or F for the prior school year according to the state's school accountability system, (2) entering kindergarten, or (3) enrolled in the Recovery School District, which includes most of the public schools in the city of New Orleans, several in Baton Rouge, and a single school in Shreveport, Louisiana. In the program's first year, 9, 736 students were eligible applicants, a majority of them outside New Orleans.

The LSP was created by Act 2 of the 2012 Regular Session of the Louisiana Legislature and Senate. The voucher size is the lesser of the amount allocated to the local school system in which the student resides or the tuition charged by the participating private school that the student attends. Average tuition at participating private schools ranges from $2,966 to $8,999, with a median cost of $4,925, compared to an average total minimum foundation program per pupil amount of $8,500 for Louisiana public schools in the 2012-13 school year.[4]

Private schools must meet certain criteria to participate in the program involving enrollment, financial practice, student mobility, and the health, safety and welfare of students. A survey of participating and non-participating private schools in Louisiana suggests that the program's regulatory requirements have influenced schools' choices to participate (Kisida, Wolf, & Rhinesmith, 2013), potentially explaining why only a third of eligible private schools opted

[4] Tuition data collected for the 2014-15 school year indicate private schools choosing to participate in the LSP have lower tuitions on average relative to non-participating Louisiana private schools, as well as lower variation in tuition. The latter finding suggests a similar group of private schools chose to participate in the program. Sude, DeAngelis, and Wolf (2017) find LSP participating private schools generally have lower enrollment, are more likely to be Catholic, and serve higher percentages of minority students than non-participating schools in the same geographic locale.

into the program in 2012-13, although school participation in the LSP has increased slightly since.[5]

Research Methodology

Experimental Design

When the LSP was expanded statewide in 2012, the Louisiana Department of Education also changed the allocation process determining scholarship awards. While the New Orleans pilot program allowed families to request only one private school for admission, the revised application process allowed individuals to offer up to five private school preferences. This new allocation process is similar to the deferred acceptance lottery used in New York City to assign students to schools through the city's public school choice program (Abdulkadiroglu, Pathak, & Roth, 2005). The algorithm prevents gaming, incentivizing families to reveal their true school preference rankings.

Eligible LSP applicants are allowed to submit up to five private school preferences and the LSP lottery algorithm places students into schools while taking into account lottery priorities. First, students with disabilities and "multiple birth siblings" – siblings with the same birthdate such as twins, triplets, etc. – are manually awarded LSP scholarships if there is available space at their preferred school. Remaining students are assigned one of six priorities:

- **Priority 1** – Students who received LSP scholarships in the prior school year who are applying to the same school

[5] There are currently four private school choice programs in operation in Louisiana, including the Louisiana Scholarship Program (Friedman Foundation for Educational Choice, 2015). The Louisiana Elementary and Secondary School Tuition Deduction program was implemented in 2008 to offer tax deductions to individual tax payers seeking to cover some of their private school expenses. The Louisiana School Choice Program for Certain Students with Exceptionalities initially launched in 2011 serving students with disabilities. Lastly, the Louisiana Tuition Donation Rebate Program, a tax-credit scholarship program, was implemented in 2012. All Louisiana private schools are eligible to participate in the Tuition Deduction program, since it is a partial tax rebate program for parents of students in private schools. Private schools can decide to participate in all, any, or none of the other three private school choice programs.

- **Priority 2** – Non-multiple birth siblings of Priority 1 awardees in the current round
- **Priority 3** – Students who received LSP scholarships in the prior school year who are applying to a different school
- **Priority 4** – New applicants who attended public schools that received a "D" or "F" grade in Louisiana's school accountability system at baseline
- **Priority 5** – New applicants who attended public schools that received a "C" grade
- **Priority 6** – New applicants who are applying to kindergarten[6]

The first stage of the LSP award process is summarized in Figure 1. The process begins by attempting to place all Priority 1 students into their first-choice school. The algorithm first groups all Priority 1 students applying to the same school and grade combination and then checks the number of available seats for that grouping. If there are more seats than applicants, all students receive a scholarship. If there are no seats available, no students receive a scholarship. If there are more applicants than seats, students are awarded LSP scholarships through a lottery.

Once the process is complete for all Priority 1 students, the algorithm attempts to place Priority 2 students into their first-choice school using the same decision rules. After cycling through all remaining priority categories, the LSP algorithm moves to the second stage of the allocation process by attempting to place remaining students in their second choice schools. The LSP algorithm continues until all eligible applicants have either been awarded or not awarded an LSP scholarship.

<<Figure 1 here>>

Only a subset of eligible applicants participated in a lottery: students in Priority 1 through 6 whose school-grade combination had more applicants than seats. Using data on student

[6] Kindergarten applicants were subject to the family income requirement; however they were not subject to the public school letter grade requirement.

characteristics and school preferences, we identify a lottery as occurring when the percentage of students awarded an LSP scholarship falls between 0 and 100 percent for a given combination of priority category, school, and grade. We focus on this subset of LSP applicants facing lotteries for their first-choice school to estimate the effects of the LSP on student achievement after three years of program participation. This focus on first-choice school lotteries ensures that an individual's own scholarship assignment is independent of other student lottery outcomes. First-choice school lotteries have been used to study the relationship between school choice and post-secondary outcomes (Deming, Hastings, Kane, & Staiger, 2014) as well as the effects of small high schools on student achievement (Bloom & Unterman, 2014).

Nevertheless, our reliance on oversubscription lotteries occurring in first-choice schools suggests our analysis may be capturing the most favorable estimates of the program's effectiveness. First, the schools in our sample are more likely to be popular among applicants, as over-subscription lotteries can only occur where there are more applicants than seats available. Moreover, higher quality schools are often more likely to be oversubscribed than lower quality schools (Abdulkadiroglu et al., 2011). These points suggest that the estimates presented here are upper bounds of the program's true effect on student achievement.

Data Description

Most of the data for this study come from student-level datasets provided by the Louisiana Department of Education (LDOE) in accordance with our data agreement with the state. The LDOE provided us with:

- Student Information Systems (SIS) files for 2011-12 and 2012-13 which include data on student enrollment and demographic background;

- LSP eligible applicant file, which includes information on the school choice sets of all eligible applicants as well as the results of the 2012-13 placement lottery[7];
- State assessment files for the 2011-12 (Baseline), 2012-13 (Year 1 Outcome), 2013-14 (Year 2 Outcome), and 2014-15 (Year 3 Outcome) school years, which include data on each student's participation in the annual accountability assessments and their scores.[8]

The Louisiana state accountability system places a strong emphasis on test-based accountability. This study uses student performance on the Louisiana state assessments in grades three through eight as our primary outcome measure of interest.[9] All students participating in the LSP are required to be tested by their private schools, using the state accountability assessments, for any grade in which the public school system also tests its students. The 2011-12, 2012-13, and 2013-14 assessment data in our study contain student scores on the LEAP and iLEAP exams, criterion-referenced tests aligned to Louisiana state education standards. The 2014-15 (Year 3 Outcome) data in our analysis instead provide student scores in ELA and math on the PARCC, a criterion-referenced test aligned with the Common Core standards, and scores on science and social studies on the Louisiana Accountability Assessment, a continuation of the LEAP/iLEAP exams aligned with state standards.[10]

It is unclear how this change in assessment regime may impact our analysis. The newness of the state test for both private and public school students may have leveled the playing field and produced a more valid gauge of the impact of the LSP on student achievement in the third

[7] Less than 1 percent of the applicant data include records with missing ID variables. These records are dropped from our analysis because we cannot link them to other data files. The applicant file also includes 20 duplicate records for which we resolve either by cross-referencing with other files or randomly keeping a single record.
[8] When possible, we have resolved duplicates by keeping records with the most complete data on LSP participants. For the remaining observations, we have randomly kept one record and dropped the other. These records represent no more than 1 percent of LSP applicants in any given year.
[9] The Louisiana program of assessments offers two alternative assessments for students with disabilities. Performance on these assessments is excluded from our analysis.
[10] PARCC assessments were administered as paper and pencil tests in grades 3 through 8 for both ELA and math. While students could receive testing accommodations, the PARCC assessments do not offer modified or alterative versions. The spring 2015 administration of PARCC assessments was considered a transition period by LDOE, with no summer retest period was made available.

year. On the other hand, LDOE considered the spring 2015 PARCC administration to be a transition period for the state's school accountability system which may have led to reduced variation in school performance if schools no longer considered the exams to be high stakes. Nevertheless, schools were not held harmless in the 2014-15 school year: student performance on PARCC assessments factored into performance scores for both public and private schools (with a sufficient number of test-takers); however the schools were graded on a curve.

The state-provided assessment data files also include information on student demographics, disability status, and participation in school initiatives such as the free- and reduced-price lunch (FRL) program and special education. Our analysis controls for these baseline covariates in order to improve effect estimate precision.

Sample Selection Process

The student-level data provided by the LDOE indicate an initial sample of 9,736 eligible LSP applicants in the first year of the program's statewide expansion. Of these, 5,296 students received LSP scholarship placements in a specific private school and 4,440 did not receive a voucher-supported placement. Our analysis relies on a sample of this original population who did not list a special education designation on their application and who were not multiple birth siblings applying for grades 1 through 6 (totaling 5,194 students). Of these, 2,746 students have outcome data in Year 3 and participated in over-subscription lotteries for their first-choice school, with 48 percent receiving placement. When we focus further on students with baseline achievement data in grades three through five, our analytical sample drops to 1,206 students. Of these, 514 – or 43 percent – won LSP scholarships to their first-choice school. This final sample of students – those with baseline achievement in grades three through five – represent our primary analytical sample of interest.

Analytical Strategy

We begin with a description of our primary analyses, which uses the results of eligible applicants' first school choice lotteries to estimate the impact of LSP scholarship usage on student achievement in a two-stage least squares (2SLS) framework. We then outline a series of subgroup analyses conducted to examine possible effect heterogeneity of the LSP.

Local Average Treatment Effect estimation. The fact that LSP scholarships are awarded through a deferred acceptance algorithm complicates our attempt to estimate the program's impact on student achievement because assignments to lower-ranked schools depend on the outcomes of earlier lotteries. We can, however, still leverage the random assignment of first-choice school lotteries to estimate the program's effect. In this design, the treatment group consists of students who receive a scholarship to attend their first-choice school, with all other students participating in LSP lotteries, including those placed in non-first-choice private schools and those not placed in any private schools, allocated to the control group. With treatment defined as winning a scholarship to attend one's first-choice school, the traditional intent-to-treat (ITT) estimator has little policy relevance, as students can participate in multiple lotteries in a deferred acceptance award process (Bloom & Unterman, 2014).

Instead, we estimate the impact of LSP scholarship usage on student achievement – also known as the Local Average Treatment Effect (LATE) (Angrist & Pischke, 2009, Cowen, 2008) – by using the result of one's first-choice school lottery to instrument for scholarship usage in a 2SLS framework. The lottery is an ideal instrumental variable as the high placement take-up rate for this program ensures that it is a strong predictor of private schooling while the random nature of the lottery process assures that scholarship receipt is uncorrelated with unobserved factors related to student achievement (Murray 2006). Because the lottery is the only way a student could receive an LSP scholarship to attend their most preferred private school, we can be confident that the variable only influences student outcomes through the private schooling that it enables.

We use the following 2SLS model to estimate the effects of LSP scholarship usage on student achievement after two years:

1. $E_i = \sum \pi_j R_{ji} + \delta T_i + \boldsymbol{X_i \beta} + u_i$

2. $A_i = \sum \alpha_j R_{ji} + \tau \widehat{E_i} + \boldsymbol{X_i \gamma} + \epsilon_i$

Where i denotes student and j denotes lottery:

- E_i indicates if a student used an LSP scholarship to enroll in an LSP-participating private school in the 2014-15 school year[11]
- R_i is a fixed effect for a student's first-choice school lottery[12]
- T_i indicates if a student received an LSP scholarship to their first-choice school via lottery in 2012-13
- A_i is standardized student mathematics or English Language Arts achievement in Year 3 of the program (2014-15)[13]
- X_i is a vector of student characteristics – including achievement – collected either at baseline (2011-12) or from the student's LSP application form

The 2SLS procedure uses one's treatment status to first predict scholarship usage and then uses this predicted value to produce an unbiased LATE effect estimate ($\hat{\tau}$) for the program.

[11] Prior evaluations of school voucher programs have examined enrollment effects in several ways. For example, Mayer et al. (2002) define enrollment as being "consistently enrolled in a private school", Wolf et al. (2013) define enrollment as "ever attending a private school", and Rouse (1998) defines enrollment as the number of years enrolled in an attempt to capture potential dosage effects. By defining enrollment as enrolling in an LSP private school in Year 3, our study falls in line with the Howell et al. (2002) evaluation of voucher programs in New York, Dayton, and Washington, D.C.

[12] We include a fixed effect for first school choice lottery to account for differing probabilities of success across lotteries (Gerber & Green, 2012). By using fixed effects, we are essentially comparing lottery winners and losers within the same strata to calculate unbiased estimates of the effect of being randomly offered an LSP scholarship. The approach is comparable to analyzing the impact of hundreds of "mini-experiments" and aggregating the results across them.

[13] Student achievement scores are standardized using distributional parameters of outcomes from the control group.

The 2SLS procedure effectively treats students who lose their first choice lottery but go on to win an LSP to a lower school preference as control-group crossovers. The result is an unbiased estimate of the effect of using a LSP scholarship to attend one's first-choice school for those who both faced and complied with their lottery assignment for placement in their first-choice school (Bloom & Unterman, 2014).

We account for nesting of students within lotteries using bootstrapped standard errors (Angrist & Pischke, 2009).[14] In addition, we may be concerned by clustering of students within their post-treatment schools or within family units (Wolf et al., 2013). The results presented here do not account for these types of nesting due to the complex nature of multi-level clustering. Clustering on lottery should capture a large amount of the nesting of individuals within current school because lottery includes the student's school of application. Moreover, we do not believe our results are strongly influenced by sibling clustering, as siblings constitute only 7 percent of our analytical sample.

Subgroup analysis. We examine if LSP impacts are differentiated by gender, race, and baseline achievement category. These comparisons are motivated by prior evaluations of school choice programs. Analyses of the New York Children's Scholarship Program, for example, find significant achievement effects for African Americans, but insignificant effect estimates overall (Mayer et al., 2002; Barnard et al., 2003). Similarly, Wolf et al. (2013) report significant improvement in reading for female participants in the DC OSP evaluation, but no significant differences for males. Wolf and colleagues also note positive achievement effects for students who were already performing well at baseline.

Treatment-Control Contrast

---

[14] The standard errors are based on 400 replications using random draws with replacement to produce cluster data with sample sizes equal to the original sample (Cameron & Trivedi, 2010).

While eligible applicants were randomly assigned to receive or not receive an LSP scholarship to their most-preferred private school, participating families were not required to use the scholarship. It is important, therefore, to verify that treatment assignment is strongly correlated with school sector enrollment. Table 1 describes the patterns of enrollment for student applicants for the 2012-13 LSP cohort that received and did not receive LSP scholarships to their first-choice schools for the three years following their initial application to the program. The analytical sample presented in Table 1 are students who did not list a special education classification on their LSP application, who were not multiple birth siblings, with baseline achievement data in grades three through five. Because our LATE analysis focuses on the results of first-choice school lotteries, the control group includes students who were never awarded a scholarship and students who received a scholarship to one of their non-first choice private school preferences. The latter group, accounting for 103 students in 2014-15, are control-group crossovers in our LATE analysis.

While the majority of lottery winners used their scholarships to attend private schools, over 75 percent of students who did not receive scholarships attended public-sector schools in all years of our study. Attrition represents no more than 12 percent of either group across all three years of data. The difference in attrition rates between treatment and control groups is disconcertingly large (What Works Clearinghouse, 2014) in Year 1, with more attrition in the control than the treatment group in Year 1 (9% versus 3%). Unfortunately, our reliance on administrative data does not allow us to distinguish the causes behind these missing data. While our primary effect estimates do not account for differential attrition, we examine the estimates' sensitivity to differential attrition using Lee's (2009) effect bounding exercise. In general, the bounding analysis does not suggest differential attrition strongly influences our primary LATE estimates.

<<Table 1 here>>

Baseline Equivalence

As a final step, we check if the LSP lottery process effectively randomized the treatment and control groups. While we cannot know if members of the treatment and control group differ on unobservable characteristics, we can get a good idea of the success of the lottery process by testing for equivalence in observable characteristics at baseline. The results of this analysis are presented in Table 2, which displays t-tests for differences in means on key baseline covariates between members of the treatment and control groups included in our analytic sample. All analyses include fixed effects for one's first school choice lottery to account for different probabilities of selection. As before, the analytical sample consists of students with baseline testing data in grades three through five who did not list a special education classification on their application, were not multiple birth siblings, who experienced lotteries for their first-choice school.

<< Table 2 here >>

The results are favorable for our analysis, as nearly all of the estimated differences between lottery winners and losers are statistically insignificant, suggesting that we have adequately identified random lotteries in our analytic sample. The lone exception is that lottery winners provided significantly fewer school preferences on average than lottery losers in both samples. Given this difference, our preferred models include controls for the full set of variables examined in Table 2.

Results

The sections that follow present our estimates of the LSP's impact on student achievement after three years. We begin with our preferred estimation model, which controls for baseline achievement and other demographics. We next explore impact heterogeneity over time and among different subgroups of students. Two robustness checks follow, which examine the LSP's impact in a broader sample of students and the extent to which differential attrition between treatment and control group members impacts our estimates.

In contrast to our prior work, which finds negative achievement impacts associated with two years of scholarship usage (Mills & Wolf, 2017), the impacts estimates presented here indicate students using an LSP scholarship to attend their most preferred private school are not generally outperforming or underperforming their control group counterparts. More specifically, the results indicate small positive effects for ELA and negative effects for mathematics. These effects are, however, not statistically significant; which is partially reflective of a decrease in statistical power due to a small sample size. As before, we find limited evidence of differential impacts across gender and race; however our results indicate positive effects among students initially performing in the bottom third of the ELA achievement distribution at baseline. For the most part, these findings are robust to alternative specifications; however we do find evidence of statistically significant negative effects on mathematics achievement when we do not require baseline achievement for inclusion in the analytical sample.

Primary Estimates of the Impact of Using an LSP on Student Achievement

We begin with results from our preferred analytical model, which requires baseline achievement for sample inclusion. This model is preferred as research continues to demonstrate the importance of controlling for baseline achievement (Peterson & Howell, 2004). By requiring baseline achievement, however, we are restricted to an analytical sample of students taking the Louisiana state assessments grades three through five in 2011-12, which effectively decreases our statistical power. Moreover, we note that the findings presented below may not necessarily be representative of all members of the 2012-13 LSP scholarship cohort. We explore this further along in the report with a robustness check performed on a larger sample of students.

The results of our primary LATE analyses are presented in Table 3. Column 1 displays coefficient estimates for first stage regressions using scholarship award to predict the likelihood of usage in fully specified models including both student demographics and baseline achievement. Students who received an LSP scholarship to their most preferred school are between 31 percentage points more likely to still be enrolled in an LSP scholarship school three years later. Column 2 presents LATE estimates for simple models controlling only for lottery fixed effects. The models presented in column 3 additionally control for baseline achievement and an indicator for if the student ever re-took a subject test in two consecutive years. Column 4 presents LATE estimates for a fully specified model controlling for baseline achievement, test re-taking, student demographics, number of school preferences listed, and an indicator capturing if a student applied to the program from a New Orleans public school. This model represents our preferred analytical model due to the predictive power of lagged dependent variables.

<< Table 3 here >>

The results presented in Table 3 indicate students using LSP scholarships to attend their first choice private school were preforming slightly better in ELA and slightly worse in mathematics and science than their control group counterparts after three years; however all effect estimates are not statistically significant. In contrast, scholarship users appear to be performing nearly 40 percent of a standard deviation behind their control group counterparts in social studies after three years. These findings stand somewhat in contrast to our earlier work, which indicates significant negative achievement impacts in the first two years of program participation (Mills & Wolf, 2017).

Next, we examine how the LSP effects vary over time. Our prior research indicates large negative impacts on ELA and mathematics in the first year of participation that appear to diminish slightly by Year 2 (Mills & Wolf, 2017). Figure 2 presents LATE estimates for ELA and mathematics for Years 1 through 3 for a consistent sample of students contributing to the analyses presented in Table 3. The dashed lines in Figure 2 represent 95 percent confidence intervals. As expected, we observe differences in student achievement between the two groups that are not statistically significant at baseline (Spring 2012). Similar to our prior work, the results presented in Figure 2 indicate LSP scholarship usage is associated with strong negative impacts in both ELA and mathematics achievement after one year, with scholarship users performing about 20 percent of a standard deviation behind in ELA and over 60 percent of a standard deviation behind in mathematics.[15] The magnitude of both effects diminish by Year 2, with ELA results not significantly different from zero in our analytical sample. By Year 3, the impacts are not statistically significant in both ELA and mathematics.

[15] The negative effect of school transfers are at least partially reflected in the negative treatment effects. Transferring between schools tends to have a disruptive impact on student achievement; which we would expect to observe among the treatment group as scholarship users by switched schools by definition. Nevertheless, the disruptive effect of school transfers cannot fully capture the large negative first year effects, as school transfers tend to be associated with between a .05 and .10 standard deviation decline in achievement (Hanushek, Kain, & Rivkin, 2004).

<< Figure 2 here >>

While the results presented in Figure 2 are in line with our previous research (Mills & Wolf, 2017), it is important to note that the estimates are less negative than those observed in our prior work. For example, we find in a consistent sample of students with baseline achievement in grades three through six significant ELA effects of -.18 standard deviations (Mills & Wolf, 2017). It is possible, therefore, that students in the current analytical sample experienced less negative impacts than students in other grades. Figure 3 examines this claim by not restricting the analysis to a consistent sample. Instead, the point estimates for 2013 are informed by students with baseline testing data in grades three through seven, the estimates for 2014 are informed by students in grades three through six at baseline, and the estimates for 2015 are informed by students in grades three through five at baseline. Consistent with the hypothesis, we see negative point estimates that are larger in magnitude in the unrestricted samples.

<< Figure 3 here >>

The model presented in column 5 of Table 3 examines the dramatic statewide expansion. There is bound to be a degree of instability during the implementation of a new education program, as schools and families learn to adjust to their new environment. These models exclude the subset of LSP applicants applying within New Orleans, home to the original LSP pilot program. New Orleans private schools arguably could be better prepared for the 2011-12 cohort of LSP students because of their experience with the pilot program since 2008. By focusing on LSP applicants outside of New Orleans, we effectively restrict our comparison to schools experiencing the flux typically accompanying the launch of new school voucher programs. In contrast to this hypothesis, the estimated LATE effects are smaller in magnitude – or more positive – when New Orleans students are excluded from the sample, perhaps reflecting the fact that most New Orleans control group students ended up attending relatively high-performing public charter schools after losing the scholarship lottery (Harris, 2015).

Subgroup Impacts

The models presented in Table 4 examine if LSP effects are differentiated by gender, ethnicity, and baseline achievement. The first four columns of Table 4 present results for ELA and the next four present results for math. All models are restricted to an analytical sample of students with baseline achievement data in grades three through five. Models labeled "Simple" include only lottery fixed effects. The models in columns 3 and 7 additionally include baseline achievement and a control for test retaking. Columns 4 and 8 present our preferred models, which are fully specified.

<< Table 4 here >>

In general, we observe limited evidence suggesting differential effects across gender and ethnicity. No estimated differences between these subgroups are statistically significant; however it should be noted that the results for ethnicity are quite noisy.

Finally, we examine if the estimated effects vary by initial achievement, with students divided into three performance categories by their baseline test scores. In contrast to the other subgroup comparisons in Table 4 which are based on interaction models, we run separate regressions by baseline achievement category to allow for the inclusion of baseline test score as a control variable.

These results suggest that the LSP's effects on achievement do differ by student baseline achievement category in ELA. Students performing in the upper and middle thirds of the achievement distribution at baseline have statistically insignificant negative program ELA impacts. In contrast, LSP students initially scoring in the lower end of the performance distribution appear to be significantly outperforming their control counterparts in ELA after three years. In contrast all estimated effects for math by baseline performance subgroup are statistically insignificant.

To summarize, these subgroup analyses largely indicate that the effects on achievement of using an LSP scholarship are not differentiated by gender or race. The estimated effects are, however, quite noisy: the reported standard errors for the difference estimates for gender and race for models requiring baseline achievement can only detect differences of over 60 percent of a standard deviation. In addition, our results indicate that students initially performing in the bottom third of the performance distribution may be strongly benefitting from LSP scholarship usage in ELA achievement by Year 3.

Robustness Checks

In general, our analyses indicate that participation in the statewide expansion of the LSP induced a program impact on Year 3 test scores that are not statistically significant, although impacts were clearly negative after one and two years. These negative findings are somewhat unique among random assignment evaluations of school voucher programs, with all but one study reporting insignificant or positive outcomes in all years examined. The lone exception is the recent evaluation of the DC Opportunity Scholarship Program, which finds negative impacts on math after one year of participation (Dynarski et al. 2017). This section presents analyses examining the sensitivity of our results to our sample restrictions and differential rates of attrition observed between treatment and control group students.

Do our results hold in an expanded sample? Our preferred analytical strategy focuses on a subset of LSP applicants with baseline achievement data in grades three through five to allow us to control for baseline achievement in our primary impact estimate models. This is our preferred strategy given research indicating the importance of controlling for pre-tests in impact evaluations (Bifulco, 2012; Peterson & Hewitt, 2004). The drawback of this strategy, however, is that it restricts our analysis to a small subsample of LSP applicants. Indeed, our analytical sample represents less than a fifth of eligible applicants to the 2012-13 LSP cohort.

Table 5 checks the sensitivity of our results to this sample restriction by dropping the baseline achievement requirement for inclusion in our analytical sample. Instead, the estimates in Table 5 are drawn from a sample of students applying for grades one through six for the 2012-13 school year. While research demonstrates the importance of controlling for baseline achievement (Peterson & Howell, 2004), the estimates presented in Table 5 are unbiased estimates of LSP's effects if we are correctly identifying lotteries in the data.[16]

<< Table 5 here >>

For the most part, the results presented in Table 5 largely align with those presented in Table 3, however they are larger in magnitude. This is consistent with our prior work, which finds larger negative effects among students applying for earlier grades who do not have baseline achievement (Mills & Wolf, 2017). The point estimates for ELA and science are all statistically non-significant, indicating students using LSP scholarships to attend their most preferred private school are performing no worse or better than their control group counterparts by Year 3. The results for mathematics and social studies, in contrast, indicate statistically significant negative impacts for LSP scholarship users. The contrasting results for mathematics between Table 3 and 5 are particularly striking, with the latter indicating scholarship users as much as 41 percent of a standard deviation behind their control group counterparts by Year 3.

The key difference between the primary analytical sample (Table 3) and the expanded sample (Table 5) is the requirement of baseline achievement data. This requirement—which we

[16] Baseline imbalance tests for the expanded sample are presented in Appendix Table A2.

implemented in our initial evaluation of the LSP (Mills, 2015; Mills, Sude, & Wolf, 2015) due to the importance of baseline achievement as a predictor and have maintained in follow-up studies (Mills & Wolf, 2017)—restricts us to a subsample of LSP applicants with baseline achievement in grades three through five. The observed differences in point estimates, especially for math, suggests the possibility of effect heterogeneity across grades. The analysis presented in Table 6 explores this possibility by examining if students applying for earlier grades experienced larger negative impacts to their achievement in response to LSP scholarship usage relative to other students in our expanded sample.

The rows of Table 6 are divided into two sections. The first section replicates the expanded sample LATE estimates presented in Table 5. The next section presents LATE estimates, standard errors, as well as lower and upper bounds of the associated 95% confidence interval (in block parentheses). Columns 1 and 3 present results from simple models that control only for lottery fixed effects. Columns 2 and 4 additionally control for student demographics, test re-taking, number of school preferences listed on the LSP application, and whether or not a student applied within New Orleans.

<< Table 6 here >>

Consistent with expectations, the results presented in Table 6 indicate significant negative LATE estimates in mathematics among students applying for lower grades. Treatment compliers who applied for first grade in the 2012-13 school year are over a quarter of a standard deviation behind their control group counterparts by 2014-15.[17] The most striking findings are the large negative effects experienced among those applying to third grade for both ELA and mathematics. These students also do not contribute to the preferred sample findings presented in Table 3, as these students would be in second grade at baseline and therefore do not have test scores. In

[17] Assuming standard grade progression, these students would be in third grade by 2013-14.

contrast, the LATE estimates for students applying for grades four through six—all of whom contribute to the analyses presented in Table 3—are all statistically non-significant. These results confirm the presence of effect heterogeneity by grade and that the difference in effects are driven more by sample than statistical model.

Does differential attrition impact the estimates? Table 1 indicates differential rates of attrition between treatment and control group members across all three years of our study. This difference, if driven by non-random factors, raises the concern of biased effect estimates (Gerber & Green, 2012; Lee, 2009; What Works Clearinghouse, 2014). If, for example, those in the treatment group with lower expected outcomes both in public and private school leave the sample with higher probability, our LATE estimates will be positively biased.

As a second check of our findings, we examine how much differential attrition may impact our estimates using a bounding procedure developed by Lee (2009). If one knows non-random attrition is concentrated solely in either the treatment or control group, Lee (2009) shows that the true, unbiased program effect lies between two bounds created by parsing away the top and bottom performers from the non-affected group.[18] In our case, we are concerned about disproportionately high levels of attrition among low performers in the treatment group. Using Lee's method, we produce an upper bound of the true effect by re-estimating the LATE effects on a subsample that excludes the lowest Year 3 performers in the treatment group. Similarly, a lower bound is created by parsing away the highest performers.

[18] Lee's (2009) bounding method relies on two assumptions: the assignment mechanism is random and sample selection is a monotonic function of treatment status. The first assumption is satisfied by the LSP lottery process. The second assumption requires that there are no lottery outcome "defiers" in our sample of LSP applicants. A defier would be a student who opts on their own accord to enroll in an LSP private school if they do not receive a scholarship, but who would not enroll in that same private school if they did win a scholarship. While we cannot validate this assumption empirically, it seems unlikely such defiers exist in our data – especially given the program's income threshold.

Table 7 presents both the original LATE estimates produced in Table 3 – included as a reference – as well as results from the bounding exercise described above. The analytical sample is restricted to students with baseline achievement data in grades three through five. As expected, the Lee bounds presented in Table 5 are quite large, with differences between lower and upper bounds of over 80 percent of a standard deviation in achievement. Despite the magnitude of these gaps, the results are consistent with LSP scholarship usage having a null or possibly negative effect on achievement. The effects for Year 3 in particular indicate null effects overall.

<< Table 7 here >>

Discussion: How Non-Compliance Affects the LATE

Our estimates in our preferred model, which controls for baseline achievement, indicate that students using their LSP scholarship to attend their most-preferred private school are performing slightly better than their control group counterparts in ELA and slightly worse in mathematics three years after initial randomization; however all estimates are not statistically significant. While the robustness check in Table 7 do not indicate these findings are sensitive to differential attrition between treatment and control group students, the comparisons presented in Figures 2 and 3 indicate the subsample of original applicants contributing to our primary program effect estimates have historically experienced less negative impacts than other applicants. This latter finding, which is potentially driven by changing rates of lottery assignment compliance, has important implications for the extent to which one can generalize these findings to the general population of LSP participants.

It is important to keep in mind that the results presented here are Local Average Treatment Effects (LATEs), which depend on lottery compliance. More specially, the LATE is the effect of using an LSP scholarship to attend one's most preferred private school *for those* *complying with their original lottery outcome* (Angrist & Pischke, 2009; Gerber & Green, 2012). This group, known as *compliers*, are students whose decision to enroll is driven by the lottery outcome: they enroll in an LSP school if they receive a scholarship via lottery and do not if they are not awarded a scholarship. In contrast, *non-compliers*, are those students whose enrollment outcome differs from their lottery outcome. This latter group includes students initially randomized to receive LSP scholarship who nevertheless enroll in public schools as well as students who do not receive a scholarship but who enroll in n LSP school.[19] Non-compliance is relevant here because the LATE effectively assumes the effect among non-compliers is zero. This section delves more deeply into changing rates of compliance to better understand the estimates presented here.

Table 8 examines compliance and non-compliance for students in our Year 3 analytical sample. The first six columns of Table 8 focus on students randomly assigned to receive an LSP scholarship to their most-preferred private school in 2012-13 ("Treatment") and the next six columns focus on students who did not receive a scholarship to attend their first-choice school via lottery ("Control").

<< Table 8 here >>

We first examine how compliance rates have changed over time in the treatment and control groups. Columns 1 and 7 identify the numbers of treatment and control group members complying and not-complying with their lottery assignment in the three years following random assignment. For example, 78 percent of treatment students contributing to our analysis initially complied with their lottery assignment in 2012-13 by attending an LSP private school. In contrast, only 31, or 5 percent, of control group students did not comply with their lottery

[19] Control group non-compliers include students wining scholarship to their less preferred private schools as well as 2012-13 LSP applicants who do not initially receive a scholarship who reapplied and were successful in later LSP cohorts.

assignment by attending an LSP private school.[20] Non-compliance has generally increased over time in both groups, with larger rates of non-compliance observed in the treatment group. Non-compliance increased from 22 percent of the treatment group in 2012-13 to 39 percent in 2013-14. By 2014-15, we see that half of the treatment group is still complying with their initial lottery assignment and half are not. In general, these results indicate an annual non-compliance rate of roughly 20 percent.

Columns 3 through 6 and 9 through 12 of Table 8 additionally allow us to understand the academic composition of non-compliers. Columns 3 and 9 present average standardized ELA achievement at baseline (2011-12 school year) for compliers and non-compliers in the treatment and control groups, respectively.[21] Treatment and control group compliers in Year 1, for example, were performing 36 and 31 percent of a standard deviation behind the state average in 2011-12, respectively. The next three columns, columns 4 through 6 in the treatment group and columns 10 through 12 in the control group, present the average difference in standardized achievement from baseline in Years 1, 2, and 3. These columns allow us to gauge if student achievement trajectories differed between compliers and non-compliers. Taken together, we see that treatment group students who initially did not comply with their treatment assignment had slightly better performance at baseline and experienced much less of a decline in performance in Year 1 relative to treatment compliers (-.02 percent of a standard deviation compared to -.19 for treatment compliers).

[20] Control group non-compliers in Year 1 are students who received an LSP scholarship to attend a lower ranked private school. As is described in the Analytical Strategy section, these students are treated as control-group crossovers by the LATE.
[21] Table 7 presents results for ELA achievement to explore the positive, yet insignificant point estimates presented in Table 3. A duplicate analysis focusing on mathematics outcomes is presented in the appendix. The results are substantively similar (see Table A3).

Looking at Years 2 and 3, we generally see that treatment group non-compliers generally performed worse than compliers at baseline. Moreover, we observe that new non-compliance appears to be predicated by large declines in achievement in the prior year, as highlighted with boxes in Table 8 for Years 2 and 3. New non-compliers in Year 2, or students initially enrolling in an LSP school in Year 1 but then switching into a public school in Year 2, experienced a decline in standardized achievement of .26 standard deviation units compared to a decline of .17 standard deviation units for compliers. Similarly, new non-compliers in Year 3 experienced a decline in ELA achievement of .22 standard deviation units in Year 2 compared to .20 standard deviation units among compliers. In addition, Table 8 indicates that non-compliance tends to result in achievement gains over time, as new non-compliers tend to experience much smaller declines in achievement in the years following their return to public schools.

Figure 4 examines non-compliance from a slightly different perspective. Here, we divide students by compliance status in the 2014-15 school year and track their achievement over time. State standardized achievement is presented on the vertical axis and test administration year is presented on the horizontal axis. Achievement is tracked for four groups of students: (1) control group students who complied with their lottery assignment in 2014-15 (blue line), (2) control group students who did not comply with their lottery assignment in 2014-15 (dashed blue line), (3) treatment group compliers (red line), and (4) treatment group non-compliers (dashed red line).

<< Figure 4 here >>

Focusing first on a comparison of the two groups of compliers, we see the performance trends that eventually result in the positive, yet statistically nonsignificant, LATE estimates for ELA achievement by Year 3 presented in Table 3. Treatment group compliers initially experience a noticeable decline in achievement in Year 1 (Spring 2013), that trends up slightly by Year 2, and then experience a large increase in achievement such that they are actually outperforming control group compliers by Year 3. In addition, Figure 4 demonstrates that treatment group non-compliers initially are the lowest performing of all four groups at baseline. Similar to treatment group compliers, non-compliers experience a decrease in achievement in Year 1. Interestingly, non-compliers additionally share an upward trajectory in performance after Year 1, however the slope is not as steep as treatment group compliers. Figure 4 additionally indicates evidence of a positive slope for control group non-compliers between Spring 2012 and 2015. While descriptive, these results are consistent with a story that students voluntarily opting out of their initial program assignment eventually recovered their achievement. Moreover, the results presented in Figure 4 indicate that all groups are performing slightly better relative to the state average compared to their initial positions in the achievement distribution in 2011-12.

Figures 5 and 6 additionally examine non-compliance in the LSP by relating school-level non-compliance rates to LSP private-school quality as proxied by school-level value added (VA) estimates based on achievement of those LSP students attending the school.[22] In both figures, school-level non-compliance rates are presented on the vertical axis relative to the same school's VA score in the prior year, with results presented separately for ELA and math. Figure 5 presents the relation between school-level non-compliance rates in Year 2 and school VA estimates in Year 1. The results indicate a negative relationship between school non-compliance and prior

[22] We estimate simple dynamic OLS value-added model (Guarino, Reckase, & Wooldridge, 2012) of the form $y_{ijgt} = \alpha y_{it-1} + \boldsymbol{X_i}\boldsymbol{\beta} + \delta_g + \gamma_j + \epsilon_{ijgt}$ where y is student achievement that has been standardized to the state's test distribution for student *i*, in school *j*, in grade *g*, in time period *t*; $\boldsymbol{X_i}$ is a vector of student characteristics recorded in time *t-1* (as reported on the student's test file); δ_g is a vector of grade fixed effects for time period *t;* γ_j is a vector of school fixed effects; and ϵ is a residual error term. The vector γ_j is our proxy measure of school quality. It is important to note that this measure is based only on student achievement for LSP students attending the school; we do not have access to achievement data for private school students who are not enrolled in the LSP.

year VA, suggesting students in particularly poor performing private schools were more likely to transfer to public schools. This relationship is weaker in mathematics relative to ELA.

<< Figure 5 here >>

<< Figure 6 here >>

Figure 6 plots the same relationship, but for Year 3. The finding of an upward slope for mathematics is somewhat confusing, as it indicates relatively higher performing private schools experienced greater non-compliance in the prior year. However, this result is not statistically different from zero. Similarly, the relationship between non-compliance and prior year ELA VA is not statistically significant. Taken together, Figures 5 and 6 suggest a link between non-compliance and school quality between Years 2 and 1; but not between Years 3 and 2.

Finally, Figure 7 examines how the compliers contributing to the Year 3 analysis have differed in achievement over time. One concern might be that treatment group compliers have historically been the highest achieving among the treatment cohort; suggesting the upward trend observed in Figure 3 is more an artifact of non-compliance than any type of program effect. The exploratory analysis presented in Figure 7 examines how the individuals contributing to the Year 3 analysis have compared over time. Interestingly, treatment group students in our preferred analytical sample who complied with their treatment assignment in Year 3 appear to have experienced very large declines in achievement in Year 1, especially in math. Consistent with the findings presented here, we observe an upward trend following Year 1, such that treatment-group compliers are performing statistically similarly to their control group counterparts in Year 3. While exploratory, the results presented in Figure 7 do not indicate the treatment compliers contributing to the Year 3 analysis have historically experienced small negative impacts. Instead, these students actually experienced larger declines in achievement in Year 1 than the general effect estimates presented in Figure 3. In contrast, Figure 7 suggests that the compliers contributing to our Year 3 estimates have struggled on successfully despite experiencing such large initial drops in ELA and math performance.

<< Figure 7 here >>

In summary, it is important to keep in mind how changing rates of lottery compliance affect our final estimate of the LSP on student achievement. The estimates presented in Table 3 are highly robust causal estimates of the LSP's impact on student achievement for students complying with their initial lottery assignment. We do observe, however, that lottery non-compliance has increased over time, as several initial scholarship recipients have opted to return to public schooling. Moreover, these returns often appear to be predicated by declining achievement in the prior year. Nevertheless, the general trends presented in Figure 4 also suggest that students making such choices experience improved performance on average over time.

Conclusion

This study examines how the statewide expansion of the Louisiana Scholarship Program (LSP) – one of the newest and largest school voucher programs in the U.S. – affected student achievement after three years. This research contributes to the existing literature on the participant effects of publicly funded voucher programs for two reasons. First, it uses a highly rigorous experimental design to estimate treatment effects while avoiding self-selection bias concerns. Second, it is among the first evaluations of a statewide school voucher program as new private school choice initiatives tend to expand from cities to encompass entire states.

The results indicate the statistically significant negative achievement effects of the LSP observed in Years 1 and 2 of our study were no longer statistically significant in Year 3, with slight positive effects estimated for ELA achievement and small negative effects estimated for mathematics. Although the immediate effects of the LSP on student achievement were clearly negative, the intermediate effects are inconclusive. The Year 3 achievement effects of the LSP are not definitively negative because the sizes of those effects have decreased somewhat since the initial disruptive year of program implementation while the variability in the estimates of test score effects has increased. Smaller differences across groups with greater variation around those differences can lead to findings of no significant differences, as is the case here.

The results from our preferred sample and model—which control for baseline achievement—indicate statistically non-significant positive LATE estimates for ELA and negative estimates for math. A robustness check dropping the baseline achievement requirement for sample inclusion, in contrast, indicates statistically non-significant LATE estimates for ELA, but significant negative effects for math. We have further shown that the difference appears to be driven by effect heterogeneity across grades, with statistically significant negative math effects concentrated among students applying for earlier grades.

The researcher preferred sample is, however, the sample that is restricted to students with baseline achievement data in grades three through five who did not list a special education exclusion on their application, were not indicated to be multiple birth siblings, and who we are able to identify as having participated in lotteries. This is our preferred analytical sample for two reasons.

First, this has consistently been our preferred methodology, starting with our study of the program's effects on student achievement after one year (Mills, 2015) and emphasized in our evaluation after two years (Mills & Wolf, 2017). This methodological decision was made initially due to the strong evidence of the importance of controlling for pre-tests in education field experiments (Bifulco, 2012; Peterson & Howell, 2004). Nevertheless, it should be noted that it is the randomization created by over-subscription lotteries, and not the baseline achievement controls, which ensure unbiased estimates of the program's effects on achievement (Murnane & Willett, 2011).

The second justification for emphasizing the results in the sample requiring baseline achievement, is that we do not actually have explicit data on which lotteries occurred. Instead, we must infer cases of lotteries using information on the 2012-13 LSP matching algorithm. We can get a sense of the extent to which our lottery identification process works by testing for baseline imbalance, as we do in Table 2. We cannot do this with the same confidence in the expanded sample because we do not have data on student achievement at baseline for all sample members nor English Language Learner status.[23] Moreover, for the added students in the unrestricted sample, not only do we not know their baseline scores, we also don't know the time trend in their scores. Thus, in the expanded sample, we are unable to be as confident that we have identified lotteries as we are with the researcher preferred sample.

Nevertheless, while we emphasize the results from the researcher preferred sample, which requires baseline achievement among the sample inclusion criteria, the evidence of effect heterogeneity clearly demonstrates that the findings from Table 3 should be generalized to all LSP students. Indeed, the results presented in Table 6 indicate large negative effects of LSP lottery compliance in math for students applying to earlier grades even after three years of program participation. While we cannot definitely identify these effects as causal estimates due to our inability to for baseline imbalance in this sample; we believe these results certainly merit caution.

[23] Indeed, we are only able to compare treatment and control group students on seven pre-treatment demographic variables for the expanded sample, with one difference found to be statistically significant at the 99% confidence level. See Table A2 in the appendix for a baseline imbalance test for the expanded sample.

Moreover, our analyses are based on a small subsample of LSP participants with performance data on the Louisiana state assessments representing approximately 15 percent of the 2012 cohort of eligible applicants. In a real sense, this study is an evaluation of the experiences of students in grades three through five at baseline, who participated in actual lotteries, with testing outcomes in Year 3. The educational impact of the LSP on the many thousands of program participants who do not satisfy those criteria remains unknown.

The purpose of this work is to provide the most rigorous assessment of the effect of the program on student achievement. In this regard, it is clear the LSP had initial negative effects on the achievement of the subset of eligible participating LSP students examined here, as measured by the official state achievement test, and those early negative effects dissipated somewhat over time even as our ability to detect programmatic effects with confidence also decreased. Three years after random assignment, we cannot reject the null hypothesis that the average test scores of LSP students and their control group peers are statistically similar.

References

Abdulkadiroglu, A., Pathak, P. A., & Roth, A. E. (2005). The New York City high school match. *American Economic Review, Papers and Proceedings*, *95*, 364-367.

Abdulkadiroglu, A., Angrist, J. D., Dynarski, S. M., Kane, T. J., & Pathak, P. A. (2011). Accountability and flexibility in public schools: Evidence from Boston's charters and pilots. *The Quarterly Journal of Economics*, *126*(2), 699–748.

Abdulkadiroglu, A., Pathak, P. A., & Walters, C. R. (2016). *School vouchers and student achievement: First-year evidence from the Louisiana Scholarship Program* (NBER Working Paper No. 21839). Cambridge, MA: National Bureau of Economic Research.

Angrist, J. D., & Pischke, J. S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton, NJ: Princeton University Press.

Barnard, J., Frangakis, C. E., Hill, J. L., & Rubin, D. B. (2003). Principal stratification approach to broken randomized experiments: A case study of school choice vouchers in New York City. *Journal of the American Statistical Association*, *98*, 299-323.

Barrow, L., & Rouse, C. E. (2008). *School vouchers: Recent findings and unanswered questions* (Economic Perspectives). Chicago, IL: Federal Reserve Bank of Chicago.

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.

Bitler, M. P., Domina, T., Penner, E. K., & Hoynes, H. W. (2015). Distributional effects of a school voucher program: Evidence from New York City (*Journal of Research on Educational Effectiveness*, *8*(3), 419-450.

Bloom, H. S., & Unterman, R. (2014). Can small high schools of choice improve education prospects for disadvantaged students? *Journal of Policy Analysis and Management*, *33*(2), 290–319.

Cameron, A. C., & Trivedi, P. K. (2010). *Microeconometrics using Stata* (Revised edition). College Station, Texas: Stata Press.

Cowen Institute for Public Education Initiatives (2013, January). Spotlight on choice: Parent opinions on school selection in New Orleans. Tulane University. Retrieved from http://www.coweninstitute.com/wp-content/uploads/2013/01/Choice-Focus-Groups-FINAL-small.pdf

Cowen, J. M. (2008). School choice as a latent variable: Estimating the complier average causal effect of vouchers in Charlotte. *Policy Studies Journal*, 36, 301–315.

Deming, D. J., Hastings, J. S., Kane, T. J., & Staiger, D. O. (2014). School choice, school quality, and postsecondary attainment. *American Economic Review*, *104*(3), 991–1013.

Dynarski, M., Rui,. N., Webber, A., & Gutmann, B. (2017). Evaluation of the DC Opportunity Scholarship Program: Impacts after one year. U.S. Department of Education, Institute for Education Sciences, National Center for Education Evaluation and Regional Assistance, Washington, DC: U.S. Government Printing Office, NCEE 2017-4022.

Figlio, D. (2011) Evaluation of the Florida Tax Credit Scholarship Program participation, compliance, and text scores in 2009-10, Report to the Florida State Department of Education, August.

Friedman, M. (1955). The Role of Government in Education. In R. A. Solo (Ed.), *Economics and the Public Interest* (pp. 123–144). New Brunswick, NJ: Rutgers University Press.

Friedman Foundation for Educational Choice. (2015). School Choice: Louisiana. Retrieved December 27, 2015, from http://www.edchoice.org/school-choice/state/louisiana/

Gerber, A. S., & Green, D. P. (2012). *Field experiments: Design, analysis, and interpretation*. New York, NY: W. W. Norton & Company.

Greene, J. P. (2001). Vouchers in Charlotte. *Education Matters*, *1*, 55–60.

Greene, J. P., Peterson, P. E., & Du, J. (1999). Effectiveness of school choice: The Milwaukee experiment. *Education and Urban Society*, *31*(2), 191–213

Gaurino, C., Reckase, M. D., & Wooldridge, J. M. (2012). *Can value-added measures of teacher education performance be trusted?* (Working Paper NO. 18). East Lansing, MI: The Education Policy Center at Michigan State University.

Hanushek, E. A., Kain, J. F., & Rivkin, S. G. (2004). Disruption versus Tiebout improvement: The costs and benefits of switching schools. *Journal of Public Economics*, *88*, 1721–1746.

Harris, D. N. (2015). Good news for New Orleans: Early evidence shows reforms lifting student achievement. *Education Next*, *15*(4).

Howell, W. G., Wolf, P. J., Campbell, D. E., & Peterson, P. E. (2002). School vouchers and academic performance: Results from three randomized field trials. *Journal of Policy Analysis and Management*, 21, 191–217.

Howell, W. G., & Peterson, P. E. (with Wolf, P. J., & Campbell, D. E.) (2006). *The educational gap: Vouchers and urban schools* (Rev. ed.). Washington, DC: Brookings.

Jin, H., Barnard, J., & Rubin, D. B. (2010). A modified general location model for noncompliance with missing data: Revisiting the New York City School Choice Scholarship Program using Principal Stratification. *Journal of Educational and Behavioral Statistics*, *35*(2), 154–173.

Kisida, B., Wolf, P. J., & Rhinesmith, E. (2015). *Views from private schools: Attitudes about school choice programs in three states*. Washington, DC: American Enterprise Institute.

Krueger, A. B., & Zhu, P. (2004). Another look at the New York City school voucher experiment. *American Behavioral Scientist*, 47, 658–698

Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *Review of Economic Studies*, 76: 1071-1102.

Louisiana Department of Education (2013a, Spring). iLEAP Interpretive Guide, Grades 3, 5, 6, and 7. Retrieved from http://www.louisianabelieves.com/docs/assessment/ileap-interpretive-guide.pdf

Louisiana Department of Education (2013b, Spring). LEAP Interpretive Guide, Grades 4 and 8. Retrieved from http://www.louisianabelieves.com/docs/assessment/leap-interpretive-guide.pdf

Louisiana Department of Education (2014). *Louisiana Scholarship Program: Annual Report 2013-2014*. Accessed on January 24, 2016, from http://www.louisianabelieves.com/docs/default-source/school-choice/2013-2014-scholarship-annual-report.pdf

Louisiana Department of Education (2013). *Louisiana Scholarship Program: Annual Report 201-2013*. Accessed on January 24, 2016, from http://www.louisianabelieves.com/docs/school-choice/scholarship-annual-report-%282013%29.pdf?sfvrsn=4

Mayer, D. P., Peterson, P. E., Myers, D. E., Tuttle, C. C., & Howell, W. G. (2002). School choice in New York City after three years: An evaluation of the school choice scholarships program. MPR Reference No. 8404-045. Cambridge, MA: Mathematica Policy Research.

Mill, J. S. (1962). *Utilitarianism, on liberty, essay on Bentham*. (Warnock, M. ed.) New York: Meridian.

Mills, J. N., Sude, Y., & Wolf, P. J. (2015). An evaluation of cognitive and non-cognitive skills in the Louisiana Scholarship Program. Thirty-seventh Annual Fall Research Meetings of the Association for Public Policy Analysis and Management, November 12-14, 2015, Miami, FL.

Mills, J. N. (2015). *The effectiveness of cash transfers as a policy instrument in K-16 education* (Doctoral Dissertation). University of Arkansas, Fayetteville, AR.

Mills, J. N. & Wolf, P. J. (2017). Vouchers in the bayou: The effects of the Louisiana Scholarship Program on student achievement after two years. *Educational Evaluation and Policy Analysis*. Available at http://journals.sagepub.com/doi/abs/10.3102/0162373717693108

Moe, T. M. (1995). *Private vouchers*. Stanford, CA: Hoover Press.

Murnane, R. J. (2005). The role of markets in American K-12 education. *The limits of market organization* (Nelson, R. R., ed.). New York: Russell Sage.

Murnane, R. J. & Willett, J. B. (2010). *Methods matter: Improving causal inference in educational and social science research*. Oxford University Press.

Murray, M. P. (2006). Avoiding invalid instruments and coping with weak instruments. *Journal of Economic Perspectives*, 20(4), 111-132.

Paine, T. (1791). *The rights of man: Answer to Mr. Burke's attack on the French Revolution*. London: J. S. Jordan.

Peterson, P. E., & Howell, W. G. (2004). Efficiency, bias, and classification schemes: A response to Alan B. Krueger and Pei Zhu. *American Behavioral Scientist*, *47*, 699–717.

Rouse, C. E. (1998). Private school vouchers and student achievement: An evaluation of the Milwaukee Parental Choice Program. *Quarterly Journal of Economics*, 113, 553–602.

Shakeel, M.D., Anderson, K. P., & Wolf, P. J. (2016). The participant effects of private school vouchers across the globe: A meta-analytic and systematic review. Social Science Research Network, EDRE Working Paper 2016-07, May 10, retrieved from: http://www.uaedreform.org/downloads/2016/05/the-participant-effects-of-private-school-vouchers-across-the-globe-a-meta-analytic-and-systematic-review-2.pdf

Sude,. Y., DeAngelis, C. A., Wolf, P. J. (2017). Supplying choice: An analysis of school participation decisions in voucher programs in DC, Indiana, and Louisiana (Louisiana Scholarship Program Evaluation Report #9). New Orleans, Louisiana: Education Research Alliance for New Orleans & School Choice Demonstration Project.

What Works Clearinghouse (2014). *Procedure and Standards Handbook, Version 3.0*. Washington, DC: Institute of Education Sciences.

Witte, J. F., Wolf, P. J., Cowen, J. M., Carlson, D., & Fleming, D. F. (2014). High stakes choice: Achievement and accountability in the nation's oldest urban voucher program. *Education Evaluation and Policy Analysis*, 36(4), 437-456.

Wolf, P. J. (2008). School voucher programs: What the research says about parental school choice. *Brigham Young University Law Review*, 2008, 415–446.

Wolf, P. J., Gutmann, B., Puma, M, Kisida, B., Rizzo, L., & Eissa, N. O. (2009) Evaluation of the DC Opportunity Scholarship Program: Impacts after three years. U.S. Department of Education, Institute for Education Sciences, National Center for Education Evaluation and Regional Assistance, Washington, DC: U.S. Government Printing Office, NCEE 2009-4050, March. http://ies.ed.gov/ncee/pubs/20094050/

Wolf, P. J., Kisida, B., Gutmann, B., Puma, M, Eissa, N. O., & Rizzo, L., (2013) School vouchers and student outcomes: Experimental evidence from Washington, DC. *Journal of Policy Analysis and Management*, 32, 246-270.

Wolf, P. J., & Macedo, S. (eds.) (2004). *Educating citizens: International perspectives on school choice and civic values*. Washington, DC: Brookings Press.

Tables

Table 1.

School enrollment patterns by scholarship award

	Treatment Group (Received LSP to First-Choice School)		Control Group (Did Not Receive LSP to First-Choice School)	
	N	%	N	%
Year 1 (2012-13)				
Non-Public School	442	78.8	49	6.4
TPS or Magnet School	71	12.7	555	72.0
Charter	29	5.2	101	13.1
Unknown/Missing School	19	3.4	66	8.6
Year 2 (2013-14)				
Non-Public School	336	59.9	112	14.5
TPS or Magnet School	132	23.5	469	60.8
Charter	53	9.4	112	14.5
Unknown/Missing School	40	7.1	78	10.1
Year 3 (2014-15) - PARCC data				
Non-Public School	282	50.2	103	13.4
TPS or Magnet School	150	26.7	452	58.6
Charter	82	14.6	137	17.8
Unknown/Missing School	48	8.5	79	10.2
Year 3 (2014-15) - LAA data				
Non-Public School	282	50.2	103	13.4
TPS or Magnet School	153	27.2	448	58.1
Charter	82	14.6	131	17.0
Unknown/Missing School	45	8.0	89	11.5

Notes. All students participated in LSP lotteries. Analysis sample excludes students with disabilities and multiple birth siblings. Year 1 is restricted to students with baseline achievement in grades 3-7. Year 2 is restricted to applicants with baseline achievement in grades 3-6. Year 3 is restricted to applicants with baseline achievement in grades 3-5. *Source*. Authors' calculations.

Table 2.

Baseline equivalence of treatment and control groups on covariates, Year 3

	N	Treatment Avg.	Control Avg.	Adjusted Diff.	*s.e.*
	(1)	(2)	(3)	(4)	(5)
Female	1,216	0.53	0.52	0.01	0.03
Race/Ethnicity					
African American	1,216	0.89	0.89	-0.03	0.02
Hispanic	1,216	0.03	0.02	0.02	0.01
White	1,216	0.06	0.07	0.00	0.02
Other	1,216	0.03	0.02	0.01	0.01
Limited English Proficiency	1,217	0.01	0.01	0.00	0.00
Free-or-Reduced Price Lunch	1,198	0.80	0.93	-0.01	0.02
Number of School Preferences Listed	1,217	1.98	2.45	-0.28***	0.08
Standardized Performance[a]					
ELA Scale Score	1,217	-0.36	-0.34	-0.01	0.06
Math Scale Score	1,216	-0.41	-0.40	0.02	0.06
Science Scale Score	1,216	-0.49	-0.49	0.04	0.06
Social Studies Scale Score	1,216	-0.42	-0.38	0.00	0.06

*** - p<.01, ** - p<.05, * - p<0.10

a. Scores are standardized within grade based on the observed distributions of scale scores across Louisiana.

Notes. Analysis sample excludes students with disabilities and multiple birth siblings. The analysis sample represents LSP applicants to grades one through six in 2012-13 who did not list a special education exclusion on their LSP application and were not multiple birth siblings. The analysis sample is additionally restricted to students with baseline in grades three through five. *Treatment* refers to students receiving LSP scholarships to their first choice private school. All other students comprise the control group. Demographics are drawn from the 2011-12 testing data. *Adjusted Diff* is the difference between Treatment and Control group students, controlling for first-choice school lottery fixed effects. “s.e.” indicates standard error of the difference, which accounts for clustering within lotteries. *Source*. Authors’ calculations

Table 3.

Estimated effects of LSP usage on student achievement after three years

		LATE			
	First Stage	Simple Model	+ Test Retake	Fully Specified	Omitting New Orleans
	(1)	(2)	(3)	(4)	(5)
English Language Arts	0.31***	0.07	0.11	0.08	0.12
	(0.04)	(0.26)	(0.25)	(0.20)	(0.19)
Mathematics	0.31***	-0.08	-0.05	-0.14	-0.03
	(0.04)	(0.25)	(0.24)	(0.24)	(0.23)
Science	0.31***	-0.08	-0.03	-0.16	-0.17
	(0.04)	(0.23)	(0.23)	(0.19)	(0.19)
Social Studies	0.31***	-0.36	-0.31	-0.38*	-0.32
	(0.04)	(0.23)	(0.22)	(0.20)	(0.21)
Controls					
Test Re-take			X	X	X
Baseline Achieve.			X	X	X
Demographics				X	X
# of Sch. Choices				X	X
New Orleans				X	
N	1195 - 1204		1195 - 1204	1175 - 1184	930 - 934
Lotteries	150		150	150	108

*** - p<.01, ** - p<.05, * - p<0.10

Notes. *Simple Model* refers estimations that only controls for risk set fixed effects. *Test Retake* indicates models including an indicator for if a student took the same subject test in 2 consecutive years. *Full Model* refers to models controlling for test retaking, baseline achievement, student demographics, number of school preferences offered, and geography. Column (5) omits students who attended New Orleans public schools in 2011-12, to account for the existence of the New Orleans pilot program. Performance measures are standardized within grade based on control group score distributions. All models include first-choice school lottery fixed effects. Standard errors (parentheses) account for clustering within lotteries. First stage F-statistics all exceed Staiger and Stock's (1997) recommended threshold of 10. *Source*. Authors' calculations.

Table 4.

Differential Effects of the LSP by Gender, Ethnicity, and Baseline Achievement

	English Language Arts				Mathematics			
	N	Simple	+ Test Retake	Full Model	N	Simple	+ Test Retake	Full Model
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Gender								
Female students	625	-0.05	0.14	0.16	625	-0.22	-0.13	-0.14
		(0.30)	(0.25)	(0.25)		(0.26)	(0.25)	(0.25)
Male students	578	0.17	0.01	0.01	576	0.04	-0.16	-0.14
		(0.29)	(0.26)	(0.25)		(0.31)	(0.30)	(0.30)
Difference		-0.21	0.14	0.14		-0.24	0.04	-0.01
		(0.31)	(0.27)	(0.28)		(0.30)	(0.26)	(0.27)
Race/Ethnicity								
Black students	1073	0.05	0.07	0.09	1072	-0.09	-0.14	-0.12
		(0.27)	(0.22)	(0.22)		(0.26)	(0.25)	(0.25)
Other students	130	-0.05	0.02	-0.09	129	-0.27	-0.21	-0.38
		(0.47)	(0.49)	(0.51)		(0.46)	(0.38)	(0.42)
Difference		0.11	0.06	0.18		0.19	0.07	0.26
		(0.47)	(0.49)	(0.53)		(0.43)	(0.36)	(0.42)
Baseline achievement								
Lower Third	408	1.32***	1.19**	1.14**	424	0.65	0.65	0.55
		(0.46)	(0.46)	(0.49)		(0.92)	(0.91)	(0.82)
Middle Third	398	0.03	-0.11	-0.32	393	-0.37	-0.40	-0.47
		(0.49)	(0.43)	(0.38)		(0.34)	(0.35)	(0.35)
Upper Third	398	-0.40	-0.39	-0.29	385	-0.28	-0.21	-0.17
		(0.29)	(0.27)	(0.28)		(0.40)	(0.45)	(0.42)

*** - p<.01, ** - p<.05, * - p<0.10

Notes. Performance measures standardized within grade based on control group score distributions. Standard errors (parentheses) account for clustering within risk sets. First stage regressions indicate the LSP scholarship award result is a good instrument for actual use. *Source*. Authors' calculations.

Table 5.

Estimated effects of LSP usage on student achievement after three years, expanded sample

		LATE			
	First Stage	Simple Model	+ Test Retake	Fully Specified	Omitting New Orleans
	(1)	(2)	(3)	(4)	(5)
English Language Arts	0.35***	-0.07	-0.04	-0.11	-0.12
	(0.03)	(0.13)	(0.13)	(0.15)	(0.14)
Mathematics	0.35***	-0.35**	-0.31**	-0.41***	-0.41***
	(0.03)	(0.15)	(0.14)	(0.15)	(0.14)
Science	0.35***	-0.14	-0.10	-0.17	-0.19
	(0.03)	(0.14)	(0.13)	(0.16)	(0.14)
Social Studies	0.35***	-0.30**	-0.26**	-0.33**	-0.26**
	(0.03)	(0.12)	(0.12)	(0.14)	(0.13)
Controls					
Test Re-take			X	X	X
Demographics				X	X
# of Sch. Choices				X	X
New Orleans				X	
N	2737 - 2746		2737 - 2746	2082 - 2087	1823 - 1827
Lotteries	299		299	253 - 254	210 - 211

*** - p<.01, ** - p<.05, * - p<0.10

Notes. Performance measures standardized within grade based on control group score distributions. All models include risk set fixed effects. Standard errors (parentheses) account for clustering within risk sets. First stage F-statistics all exceed Staiger and Stock's (1997) recommended threshold of 10.

Table 6.

Effect heterogeneity across application grades in expanded sample

	ELA		Mathematics	
	Simple Model	Full Model	Simple Model	Full Model
	(1)	(2)	(3)	(4)
General Model	-0.07	-0.11	-0.35***	-0.41***
	(0.13)	(0.13)	(0.12)	(0.11)
	[-0.33,0.19]	[-0.36,0.13]	[-0.58,-0.11]	[-0.62,-0.19]
Application Grade				
First Grade	0.17	-0.02	-0.11	-0.26*
	(0.24)	(0.32)	(0.24)	(0.16)
	[-0.31,0.65]	[-0.65,0.62]	[-0.57,0.36]	[-0.57,0.05]
Second Grade	0.13	0.16	-0.59**	-0.55*
	(0.19)	(0.22)	(0.24)	(0.33)
	[-0.24,0.49]	[-0.27,0.58]	[-1.06,-0.11]	[-1.19,0.09]
Third Grade	-0.59	-0.86**	-0.72*	-1.23***
	(0.38)	(0.43)	(0.43)	(0.47)
	[-1.33,0.15]	[-1.70,-0.02]	[-1.56,0.12]	[-2.16,-0.31]
Fourth Grade	-0.18	-0.28	-0.29	-0.11
	(0.43)	(0.61)	(0.42)	(0.46)
	[-1.03,0.67]	[-1.48,0.92]	[-1.11,0.54]	[-1.01,0.79]
Fifth Grade	-0.24	0.05	-0.64	-0.66
	(0.38)	(0.57)	(0.39)	(0.56)
	[-0.98,0.51]	[-1.05,1.16]	[-1.4,0.13]	[-1.76,0.44]
Sixth Grade	0.15	0.17	0.15	0.25
	(0.25)	(0.29)	(0.35)	(0.47)
	[-0.33,0.64]	[-0.40,0.74]	[-0.54,0.85]	[-0.67,1.17]
Controls				
Test Re-take		X		X
Demographics		X		X
# of Sch. Choices		X		X
New Orleans		X		X

*** - p<.01, ** - p<.05, * - p<0.10

Notes. Performance measures standardized within grade based on control group score distributions. All models include risk set fixed effects. Standard errors (parentheses) account for clustering within risk sets. Lower and upper bounds for 95% confidence intervals are presented in block parentheses, respectively. First stage F-statistics all exceed Staiger and Stock's (1997) recommended threshold of 10.

Table 7.

Examining effects of differential attrition

	English Language Arts		Mathematics	
	N	LATE w/ covariates	N	LATE w/ covariates
	(1)	(2)	(3)	(4)
Year 1 - Spring 2013				
Primary LATE	1273	-0.40***	1274	-1.18***
		(0.15)		(0.22)
Lee Lower Bound	1242	-0.57***	1246	-1.30***
		(0.16)		(0.22)
Lee Upper Bound	1245	-0.17	1246	-0.88***
		(0.13)		(0.18)
Year 2 - Spring 2014				
Primary LATE	1242	-0.11	1240	-0.44**
		(0.13)		(0.19)
Lee Lower Bound	1226	-0.18	1223	-0.52***
		(0.13)		(0.16)
Lee Upper Bound	1226	0.03	1222	-0.29*
		(0.13)		(0.16)
Year 3 - Spring 2015 - PARCC Test Takers				
Primary LATE	1204	0.08	1202	-0.14
		(0.20)		(0.24)
Lee Lower Bound	1191	-0.11	1189	-0.33
		(0.19)		(0.23)
Lee Upper Bound	1190	0.21	1186	0.03
		(0.18)		(0.22)

*** - p<.01, ** - p<.05, * - p<0.10

Notes. All models control for baseline achievement, test retaking, student demographics, number of school preferences offered, geography, and include first-choice school lottery fixed effects. Performance measures standardized within grade based on control group score distributions. Standard errors (parentheses) account for clustering within lotteries. First stage regressions indicate the LSP scholarship award result is a good instrument for actual use. *Source*. Authors' calculations.

Table 8.

Examining magnitude and composition of non-compliance over time for consistent sample of students contributing to Year 3 analysis, ELA Achievement

	Treatment (Received LSP to First-Choice school)						Control (Did Not Receive LSP to First-Choice School)					
			State Standardized ELA Achievement						State Standardized ELA Achievement			
	N	%	Baseline 2011-12	Year 1 Diff.	Year 2 Diff.	Year 3 Diff.	N	%	Baseline 2011-12	Year 1 Diff.	Year 2 Diff.	Year 3 Diff.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Year 1												
Compliers	315	78%	-0.36	-0.19	-0.15	0.06	644	95%	-0.31	0.02	-0.02	0.10
Non-compliers	90	22%	-0.34	-0.02	0.09	0.09	31	5%	-0.43	-0.07	0.04	0.17
Year 2												
Compliers	246	61%	-0.26	-0.17	-0.20	0.04	578	86%	-0.32	0.03	0.00	0.10
Non-compliers	159	39%	-0.51	-0.14	0.06	0.11	97	14%	-0.30	-0.04	-0.11	0.11
Old Non-compliers	78	19%	-0.38	-0.01	0.14	0.11	29	4%	-0.40	-0.11	0.04	0.21
New Non-compliers	81	20%	-0.63	-0.26	-0.02	0.10	68	10%	-0.25	-0.01	-0.17	0.07
Year 3												
Compliers	202	50%	-0.26	-0.21	-0.20	0.06	584	87%	-0.33	0.02	-0.01	0.10
Non-compliers	203	50%	-0.44	-0.10	0.00	0.08	91	14%	-0.22	0.00	-0.09	0.12
Old Non-compliers	158	39%	-0.49	-0.14	0.06	0.10	82	12%	-0.23	-0.03	-0.10	0.12
New Non-compliers	45	11%	-0.27	0.03	-0.22	0.02	9	1%	-0.16	0.29	0.02	0.11

Notes. Sample includes students contributing to fully specified LATE estimates presented in Table 3, column 4. Sample inclusion requires baseline achievement in grades three through five, Spring 2015 outcome data, and non-missing values for all demographic variables. Achievement estimates are standardized to the state's testing distribution in a given year. *Year 1 Diff* refers to the difference in standardized achievement in 2012-13 from 2011-12. *Year 2 Diff* refers to the difference in standardized achievement in 2013-14 from 2011-12. *Year 3 Diff* refers to the difference in standardized achievement in 2014-15 from 2011-12. New Non-compliers among the treatment group are highlighted in Year 2 and Year 3 for clarity. *Source*. Authors' calculations.

Figures

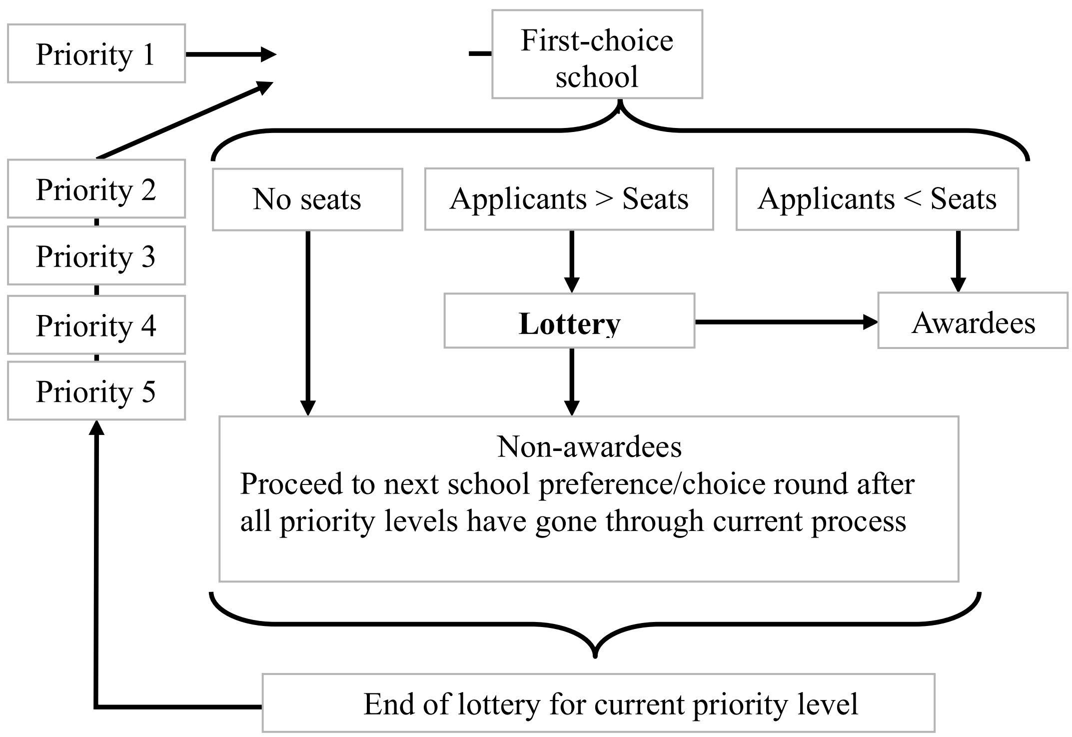

Figure 1. First stage of the Louisiana Scholarship Program award allocation process for the 2012-2013 school year. This figure illustrates the iterative process used to allocate LSP scholarships to students. In addition, this figure highlights the fact that only a subset of students was awarded LSP scholarships via lotteries. Our analysis focuses on isolating lotteries for one’s first-choice school. LSP = Louisiana Scholarship Program.

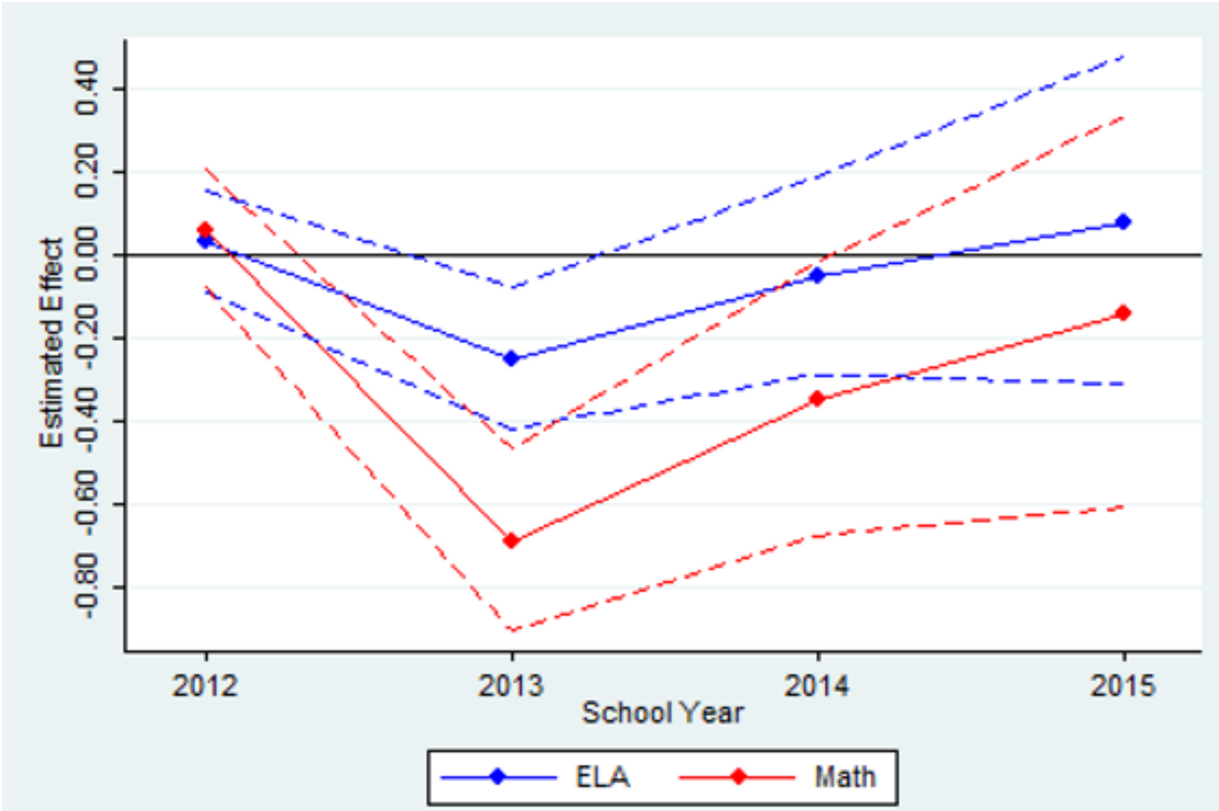

Figure 2. Estimated Local Average Treatment Effects over time. Figure presents point estimates from fully specified models for 2011-12 (baseline) through 2014-15 for ELA and math. Results are presented for a consistent sample of students with Spring 2015 outcome data. ELA and math results are based on student achievement on the Louisiana state assessments (LAA) in 2011-12 through 2013-14, but are based on PARCC assessment performance in 2014-15. Dashed lines represent 90% confidence intervals for the performance averages. σ indicates "standard deviation".

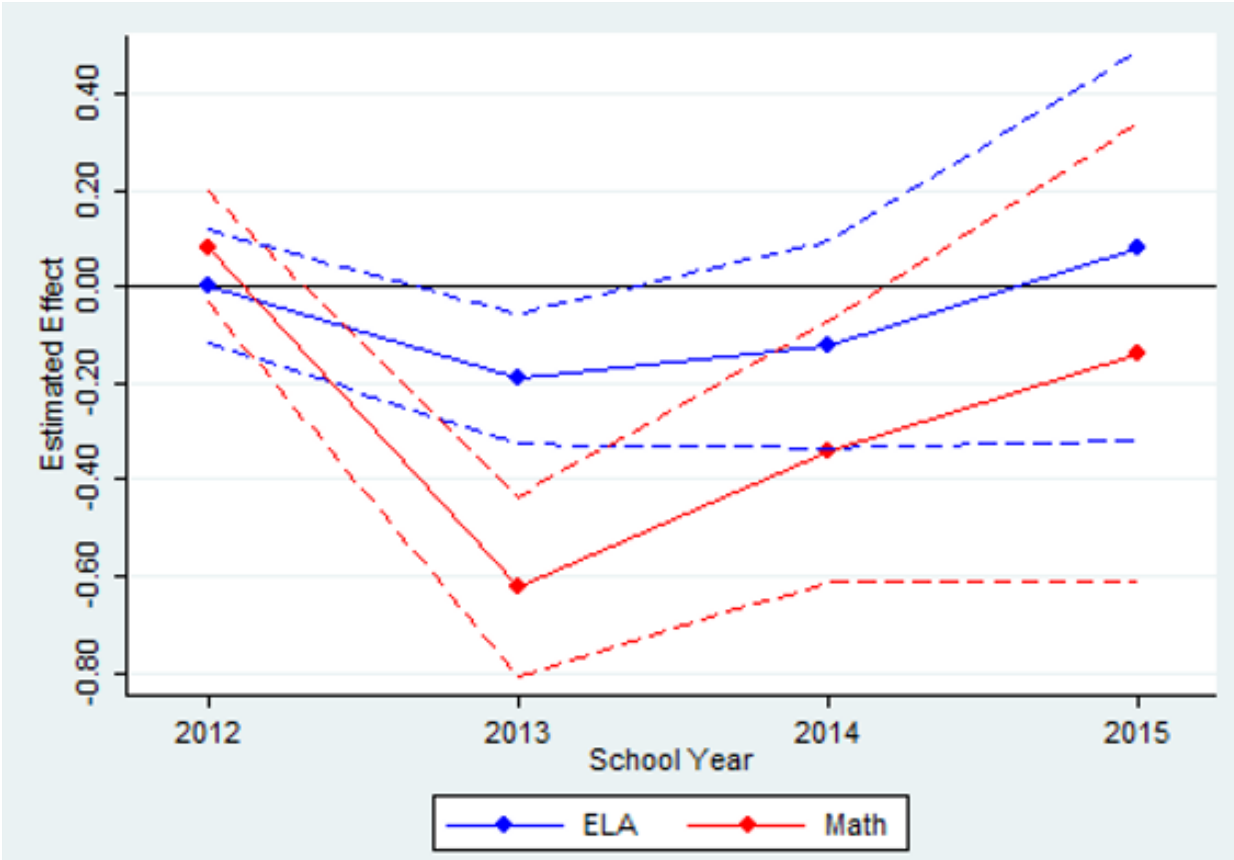

Figure 3. Estimated Local Average Treatment Effects over time. Figure presents point estimates from fully specified models for 2011-12 (baseline) through 2014-15 for ELA and math. Results are presented for a all records that can contribute to analysis of students with Spring 2015 outcome data. ELA and math results are based on student achievement on the Louisiana state assessments (LAA) in 2011-12 through 2013-14, but are based on PARCC assessment performance in 2014-15. Dashed lines represent 95% confidence intervals for the performance averages. σ indicates "standard deviation".

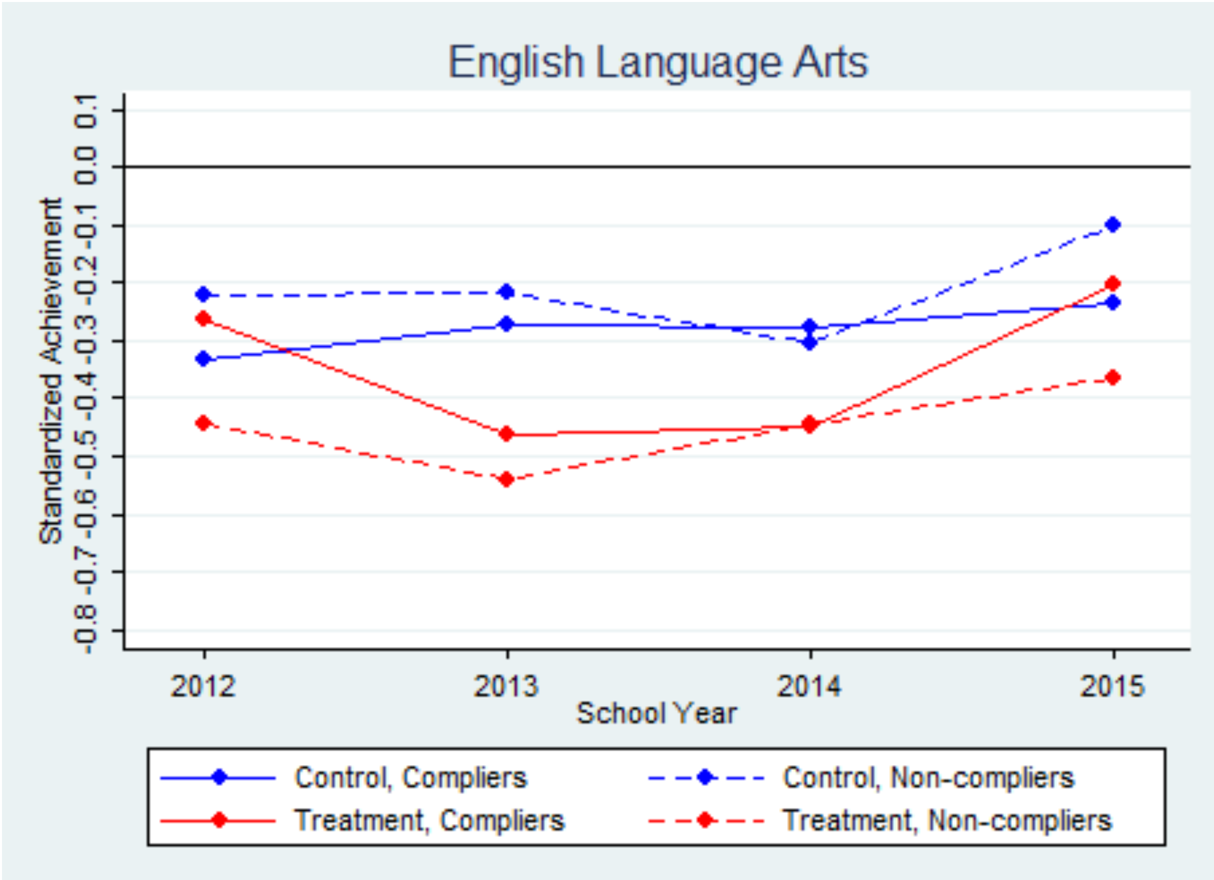

Figure 4. Average ELA standardized achievement for compliers and non-compliers contributing to the Year 3 analysis. Figure presents average ELA achievement that has been standardized to the state's test distribution in each year for four groups of students: (1) control group students complying with their 2012-13 LSP lottery outcome in 2014-15 (Year 3), (2) control group students who are not complying with their lottery outcome in 2014-15, (3) treatment group students complying with their lottery outcome in 2014-15, and (4) treatment group students not complying with their lottery outcome in 2014-15. The black line indicates average performance in a given year. This figure shows that all groups generally performed behind the state average in all years and that non-compliers among the treatment group were generally among the lowest performers. It also shows that both treatment compliers and non-compliers experienced decreases in achievement in 2012-13, that has generally trended upward since, with a stronger upward trend experienced among treatment group compliers between 2013-14 and 2014-15.

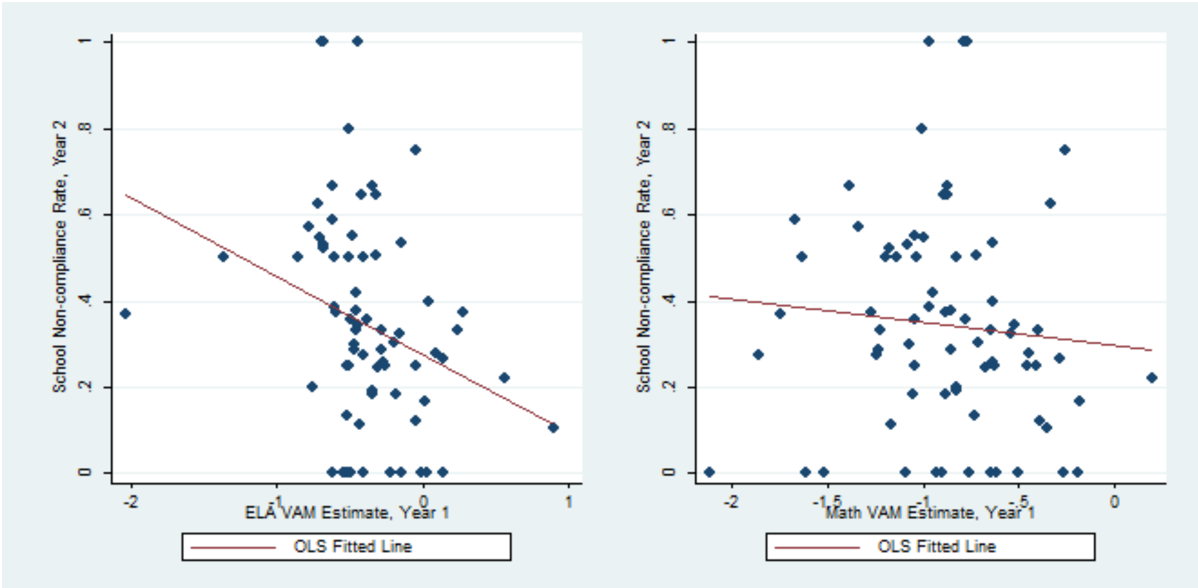

Figure 5. Association between private school non-compliance rates in Year 2 with prior year school value-added model estimates. The figure plots the relationship between school-level non-compliance and prior year VA estimates, for both ELA and math. The results indicate a negative relationship between school non-compliance and prior year VA, suggesting students in particularly poor performing private schools were more likely to transfer to public schools. This relationship is weaker in mathematics relative to ELA, and is indeed not statistically significant. We have omitted a single outlier school with a particularly low VA estimate.

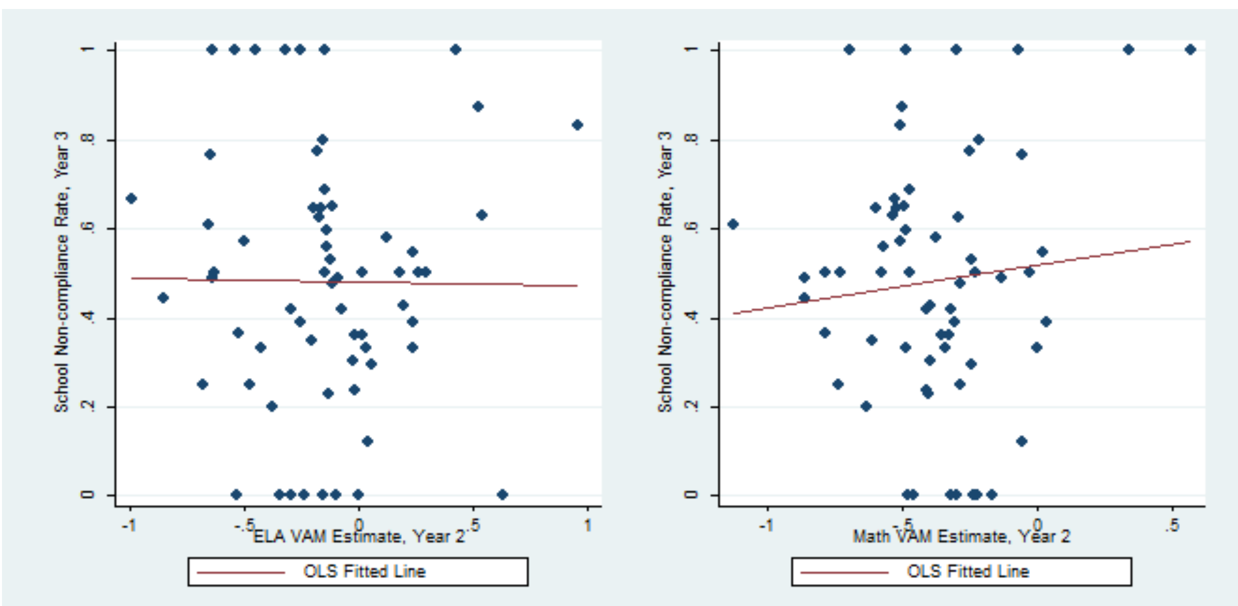

Figure 6. Association between private school non-compliance rates in Year 3 with prior year school value-added model estimates. The figure plots the relationship between school-level non-compliance and prior year VA estimates, for both ELA and math. The results are not statistically significant for either ELA or math.

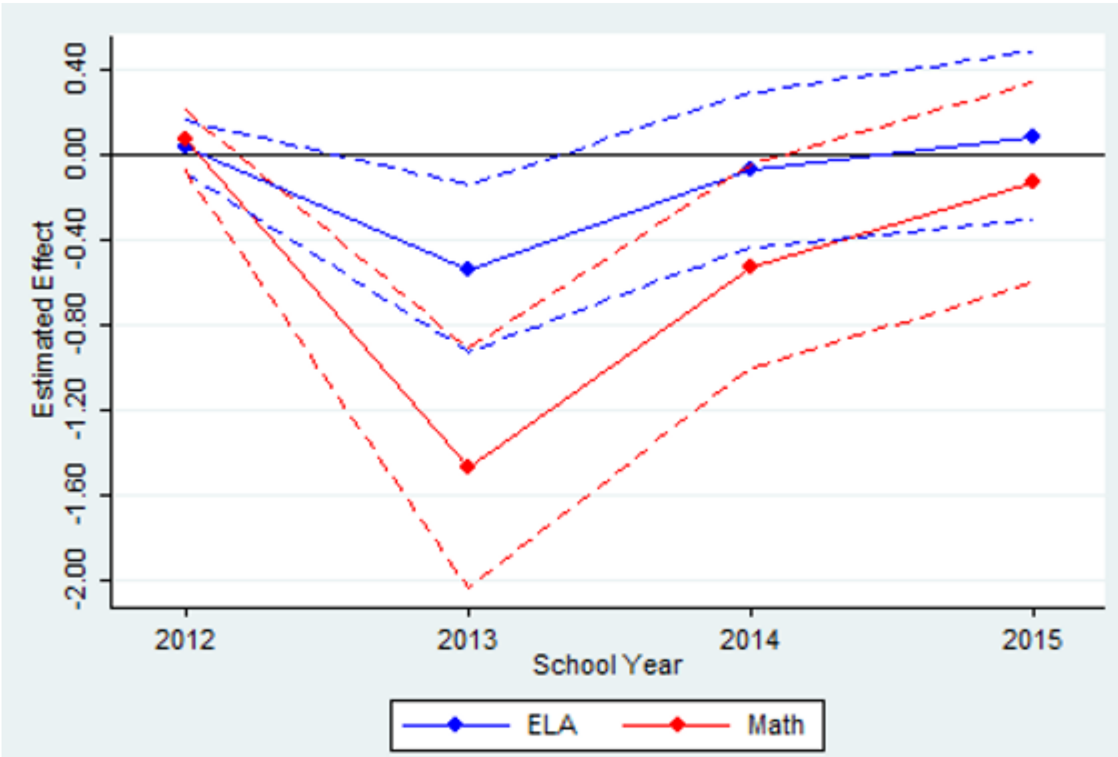

Figure 7. Comparing Year 3 treatment group compliers and control group compliers over time. Figure presents estimated differences between treatment and control group compliers contributing to the Year 3 analysis over time. The vertical axis is achievement that has been standardized to the control group's test distribution.

Appendix

Table A1.

Differential attrition rates between treatment and control across time

	N	Awarded LSP to 1st Choice School	Not Awarded LSP to 1st Choice School	Diff.	*s.e.*
LA Assessments, Spring 2013	1359	0.03	0.08	-0.05***	(0.01)
LA Assessments, Spring 2014	1358	0.07	0.10	-0.03*	(0.02)
PARCC Assessments, Spring 2015	1333	0.08	0.11	-0.03	(0.02)
LA Assessments, Spring 2015	1333	0.07	0.12	-0.04**	(0.02)

*** - p<.01, ** - p<.05, * - p<0.10

Notes. All models include lottery fixed effects. Standard errors (parentheses) account for clustering within risk sets. *Source.* Authors' calculations

Table A2.

Baseline equivalence of treatment and control groups on covariates in expanded sample, Year 3

	N	Treatment Avg.	Control Avg.	Adjusted Diff.	*s.e.*
	(1)	(2)	(3)	(4)	(5)
Female	2,746	0.53	0.51	0.01	0.02
Race/Ethnicity					
African American	2,746	0.90	0.89	-0.02	0.01
Hispanic	2,746	0.03	0.03	0.01	0.01
White	2,746	0.05	0.06	0.00	0.01
Other	2,746	0.03	0.02	0.01	0.01
Limited English Proficiency	n/a	n/a	n/a	n/a	n/a
Free-or-Reduced Price Lunch	2,087	0.94	0.95	-0.01	0.01
Number of School Preferences Listed	2,746	2.01	2.54	-0.22***	0.05
Standardized Performance					
ELA Scale Score	n/a	n/a	n/a	n/a	n/a
Math Scale Score	n/a	n/a	n/a	n/a	n/a
Science Scale Score	n/a	n/a	n/a	n/a	n/a
Social Studies Scale Score	n/a	n/a	n/a	n/a	n/a

*** - p<.01, ** - p<.05, * - p<0.10

Notes. Analysis sample excludes students with disabilities and multiple birth siblings. FRL information for Sample 1 members drawn from 2011-12 Student Information Systems data. “n/a” indicates the variable is unavailable for the given comparison. *Treatment* refers to students receiving LSP scholarships to their first choice private school. All other students comprise the control group. Sample 1 represents LSP applicants to grades three through seven in 2012-13 who did not list a special education exclusion on their LSP application and were not multiple birth siblings. Sample 2 is additionally restricted to students with baseline in grades three through six. Sample 1’s demographic set are drawn from LSP application data. Sample 2’s demographics are drawn from the 2011-12 testing data, which contain richer student data compared to the application data. Adjusted Diff is the difference between Treatment and Control group students, controlling for first-choice school lottery fixed effects. “s.e.” indicates standard error of the difference, which accounts for clustering within lotteries. *Source*. Authors’ calculations

Table A3.

Examining magnitude and composition of non-compliance over time for consistent sample of students contributing to Year 3 analysis, Mathematics Achievement

	Treatment (Received LSP to First-Choice School)						Control (Did Not Receive LSP to First-Choice School)					
			State Standardized Math Achievement						State Standardized Math Achievement			
	N	%	Baseline 2011-12	Year 1 Diff.	Year 2 Diff.	Year 3 Diff.	N	%	Baseline 2011-12	Year 1 Diff.	Year 2 Diff.	Year 3 Diff.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Year 1												
Compliers	315	78%	-0.418	-0.480	-0.290	-0.019	643	95%	-0.387	0.097	-0.003	0.084
Non-compliers	90	22%	-0.363	-0.086	-0.123	0.044	31	5%	-0.359	-0.584	-0.621	-0.189
Year 2												
Compliers	246	61%	-0.316	-0.510	-0.402	-0.106	577	86%	-0.403	0.094	0.044	0.108
Non-compliers	159	39%	-0.544	-0.214	-0.012	0.150	97	14%	-0.283	-0.102	-0.476	-0.149
Old Non-compliers	78	19%	-0.403	-0.069	-0.065	0.115	29	4%	-0.332	-0.565	-0.675	-0.149
New Non-compliers	81	20%	-0.680	-0.348	0.038	0.184	68	10%	-0.262	0.102	-0.391	-0.149
Year 3												
Compliers	202	50%	-0.319	-0.508	-0.415	-0.103	583	87%	-0.409	0.081	0.032	0.104
Non-compliers	203	50%	-0.491	-0.280	-0.085	0.092	91	14%	-0.239	-0.034	-0.434	-0.140
Old Non-compliers	158	39%	-0.536	-0.212	-0.012	0.134	82	12%	-0.223	-0.082	-0.487	-0.139
New Non-compliers	45	11%	-0.331	-0.521	-0.339	-0.058	9	1%	-0.387	0.394	0.100	-0.152

Notes. Sample includes students contributing to fully specified LATE estimates presented in Table 3, column 4. Sample inclusion requires baseline achievement in grades three through five, Spring 2015 outcome data, and non-missing values for all demographic variables. Achievement estimates are standardized to the state's testing distribution in a given year. *Year 1 Diff* refers to the difference in standardized achievement in 2012-13 from 2011-12. *Year 2 Diff* refers to the difference in standardized achievement in 2013-14 from 2011-12. *Year 3 Diff* refers to the difference in standardized achievement in 2014-15 from 2011-12. New Non-compliers among the treatment group are highlighted in Year 2 and Year 3 for clarity. *Source*. Authors' calculations.

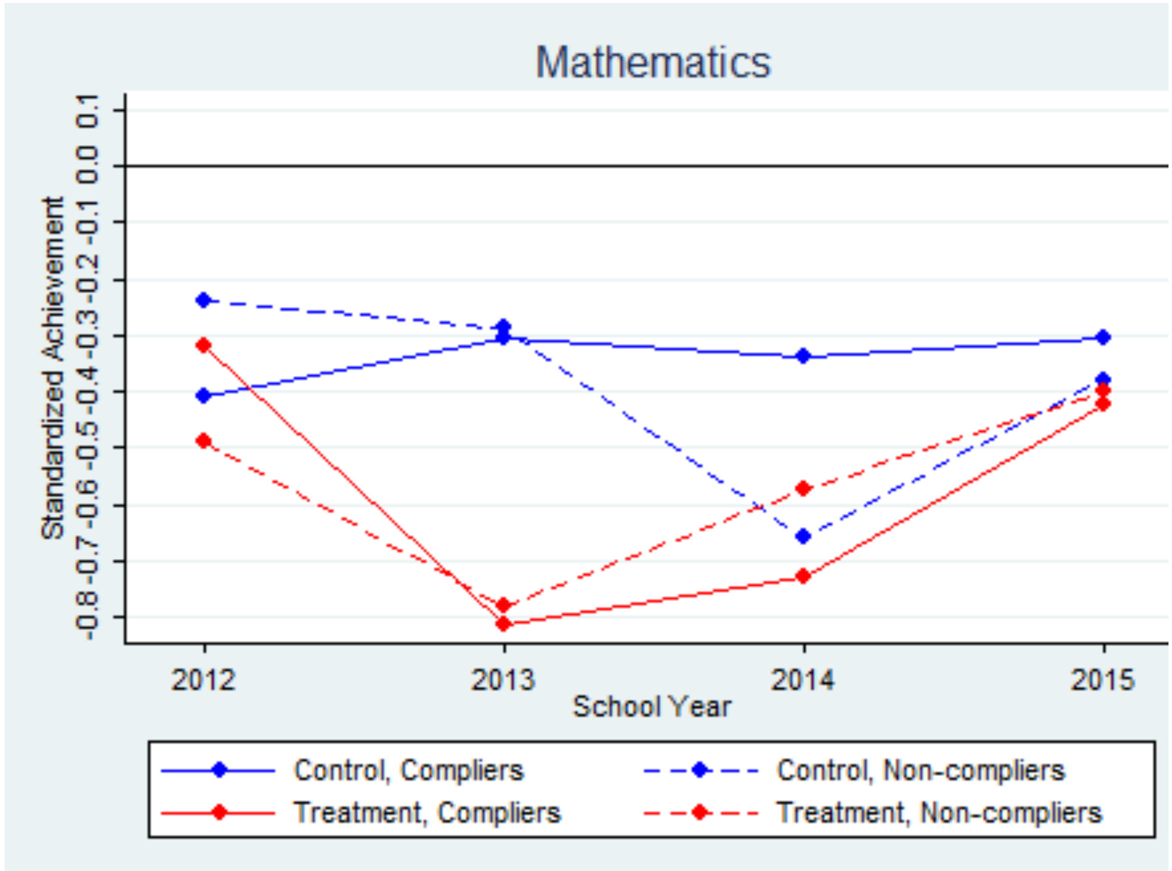

Figure A1. Average mathematics standardized achievement for compliers and non-complie contributing to the Year 3 analysis. Figure presents average ELA achievement that has bee standardized to the state's test distribution in each year for four groups of students: (1) con group students complying with their 2012-13 LSP lottery outcome in 2014-15 (Year 3), (2) control group students who are not complying with their lottery outcome in 2014-15, (3) treatment group students complying with their lottery outcome in 2014-15, and (4) treatme group students not complying with their lottery outcome in 2014-15. The black line indicat average performance in a given year. This figure shows that all groups generally performec behind the state average in all years and that compliers among the treatment group were generally among the lowest performers. It also shows that both treatment compliers and no compliers experienced decreases in achievement in 2012-13 that has generally trended upw since, with a stronger upward trend experienced among treatment group compliers betweer 2013-14 and 2014-15.