UNIVERSITY OF ARKANSAS | College of Education & Health Professions
Education Reform

WORKING PAPER SERIES

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

Jonathan N. Mills and Patrick J. Wolf

April 23, 2019

EDRE Working Paper 2019-10

THE EFFECTS OF THE LOUISIANA SCHOLARSHIP PROGRAM ON STUDENT ACHIEVEMENT AFTER FOUR 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 #10
April 23, 2019

School Choice Demonstration Project, University of Arkansas, Fayetteville, AR

Acknowledgments
We thank Jay P. Greene, Gema Zamarro, Brian Kisida, Joshua Cowen, David Figlio, Brian Gill, Nathan Barrett, and Douglas Harris for their extensive helpful 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. We thank the Smith Richardson Foundation for the grant support that made this research possible. We are grateful to Kathleen Wolf for editorial help. The content of the report is solely the responsibility of the authors and does not necessarily represent the views of the University of Arkansas, the Louisiana Department of Education, or the Smith Richardson Foundation.

* Corresponding author. Address 201 Graduate Education Building, Fayetteville, AR 72701; 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. In this study, we estimate the achievement impacts of ever using an LSP voucher to enroll in one's first-choice private school over the four year period spanning from 2012-13 (Year 1) through 2015-16 (Year 4). In contrast to our previous research, which indicated large initial negative achievement effects of the program that improved after two and three years of participation, the results presented here indicate large negative effects of LSP voucher usage after four years, especially in math. Similar to previous studies, we observe little evidence of differentiation in general effects by gender. However, in contrast to findings presented in our previous studies, African American students are reported to have significantly less negative impacts of voucher usage relative to other students.

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

THE EFFECTS OF THE LOUISIANA SCHOLARSHIP PROGRAM ON STUDENT ACHIEVEMENT AFTER FOUR 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 four 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 previous research indicates large negative impacts of LSP voucher usage on both English Language Arts (ELA) and math achievement after one year of participation (Mills, 2015) that appear to attenuate over the following two years (Mills & Wolf, 2017a; Mills & Wolf,

[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" to "no-lottery winners." Given evidence suggesting over-subscribed schools tend to be better performing (Abdulkadiroglu, Angrist, Dynarski, Kane, & Pathak, 2011) as well as the fact that first-choice schools are likely to be popular schools, it is likely the effects presented here are upper bound estimates of the impact of LSP scholarship usage.

2017b). By Year 3, the achievement of LSP voucher users was 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 (Mills & Wolf 2017b).

In this study, we estimate how achievement of the 2012-13 LSP application cohort was impacted by using an LSP voucher at any point over the four year period spanning from 2012-13 through 2015-16. In general, our study indicates the trend of improvement in LSP scholarship usage effects observed in our prior reports (Mills & Wolf, 2017a; Mills & Wolf, 2017b) did not continue into the fourth year of the program. After four years, students using LSP vouchers to enroll in LSP participating private schools performed noticeably worse on state assessments than their control group counterparts. These estimates are statistically significant in the majority of models and are consistently negative in both math and science. Similar to previous studies, we observe little evidence of differentiation in general effects by gender. However, in contrast to findings presented in our previous studies, African American students are reported to have significantly less negative impacts of voucher usage relative to other students.

There are two important caveats to note regarding our analysis. First, this study deviates from our previous evaluations by estimating the effect of ever using an LSP voucher to attend a private school between 2012-13 and 2015-16. In contrast, our prior studies estimated the achievement impacts of using an LSP voucher to attend a private school in a *specific* year: 2012-13 (Mills, 2015), 2013-14 (Mills & Wolf, 2017a), and 2014-15 (Mills & Wolf, 2017b). There are benefits and drawbacks to estimating the effect of usage in a specific year. A benefit, for example, is that this specification allows us to identify effect for students who have continued to comply with assignment over time, since almost all the students using LSP vouchers in a given

year also used in all previous years. In a sense, this approach captures the impacts of persistent voucher use. On the other hand, this particular Local Average Treatment Effect (LATE) estimator effectively ignores the experience of students who left the program during the four-year time period, as it treats all lottery winners who are no longer enrolled in private schools as treatment group non-compliers. This feature of our prior analytic method is problematic in our current research setting, as the percentage of LSP lottery winners continually using vouchers to attend participating private schools has noticeably declined over time (Sude & Wolf, 2019). Indeed, by Year 4 (2015-16), we observe more LSP lottery winners attending *public* schools than private schools. By estimating the impact of ever using an LSP voucher instead of focusing on enrollment in a specific year, we are able to recapture those initial LSP scholarship users who have switched back to public schools over time. This approach provides a more general – and potentially more policy relevant – picture of the effect of LSP scholarship usage on student achievement as it allows for the identification of impacts even for those scholarship users who find themselves dissatisfied with their private school experience.[3]

Second, while our prior research has generally emphasized findings from models requiring baseline achievement (Mills, 2015, Mills & Wolf, 2017a; Mills & Wolf, 2017b), this requirement dramatically reduces the potential analytical sample by Year 4. Specifically, only LSP applicants with achievement in grades 3 or 4 at baseline meet this sample inclusion standard when examining scholarship usage effects four years after initial assignment. This reduction in sample both limits our analysis's statistical power as well as our ability to distinguish true voucher usage impacts from cohort effects. Therefore, we also present results for an analytical

[3] In Appendix A, we present results from models estimating the achievement impacts of continued LSP voucher usage in Year 4 (2015-16). These estimates are not substantively different than the primary effects presented in this paper: LSP voucher usage is generally associated with negative or statistically insignificant differences after four years.

sample generated by waiving the baseline achievement requirement. Fortunately, if we have correctly identified first-choice school lotteries in our data, these results represent causal estimates of LSP scholarship usage on student achievement. In general, estimated effects for this alternative sample largely correspond with primary analytical sample estimates, with statistically significant negative impacts observed on math and science test scores. In addition, estimated effects are negative and statistically significant in ELA for this sample.

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 four 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. However, we do label specific initiatives appropriately when discussing them.

While economist Milton Friedman (1955) introduced the educational voucher idea in the United States, the theoretical support for its 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 the matching of student academic needs to specific school programs and environments, 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 United States has been experimental. In the sections that follow, we summarize findings from studies of U.S. voucher programs using either gold-standard experimental research designs or highly rigorous quasi-experimental designs.

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 20 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, Indiana, Milwaukee, New York City, Toledo, and Louisiana to determine

their impacts on student achievement. While early evaluations of voucher and voucher-type programs reported effects ranging from neutral to positive, several recent evaluations of statewide programs report negative impacts on test scores.

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. Indeed, 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).

Nevertheless, several recent studies report negative achievement effects of school voucher programs, especially in math. Quasi-experimental evaluations of statewide programs in Ohio (Figlio & Karbownik, 2016) and Indiana (Waddington & Berends, 2018) as well as experimental evaluations of a statewide program in Louisiana (Abdulkadiroglu, Pathak, & Walters, 2018; Mills, 2015; Mills & Wolf, 2017a) and a voucher program in Washington, D.C. (Dynarski et al., 2018) all report voucher usage to have statistically significant negative effects on math achievement.

Program effects often vary over time. An evaluation of the privately-funded Washington Scholarship Fund in D.C. 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 year (Witte et al. 2014).

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).

An experimental evaluation of a small sample of students who applied for a privately funded scholarship program in Toledo, Ohio, concluded that math outcomes were not significantly different between the scholarship and control group students (Bettinger & Slonim, 2006). 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 reflects noticeable effect heterogeneity, with estimated effects varying from negative to positive, varying over time, and varying across student subgroups within programs. Our study adds to the literature on private school voucher programs by examining the effects of a statewide voucher program on intermediate student achievement outcomes using a highly rigorous experimental research design.

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 amount is the lesser of the student's public school state and local allocation or the tuition charged by the participating private school that the student attends. In the 2012-13 school year—the year in which our analysis cohort first applied for LSP vouchers—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.

Private schools must meet certain state government regulations to participate in the program involving admissions, 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 into the program in 2012-13, although school participation in the LSP has increased slightly since.[4]

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 list 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, &

[4] 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.

Roth, 2005). The algorithm prevents gaming, incentivizing families to reveal their true school preference rankings.

Eligible LSP applicants submit up to five private school preferences. The LSP lottery algorithm then 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 space exists 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
- **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

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.

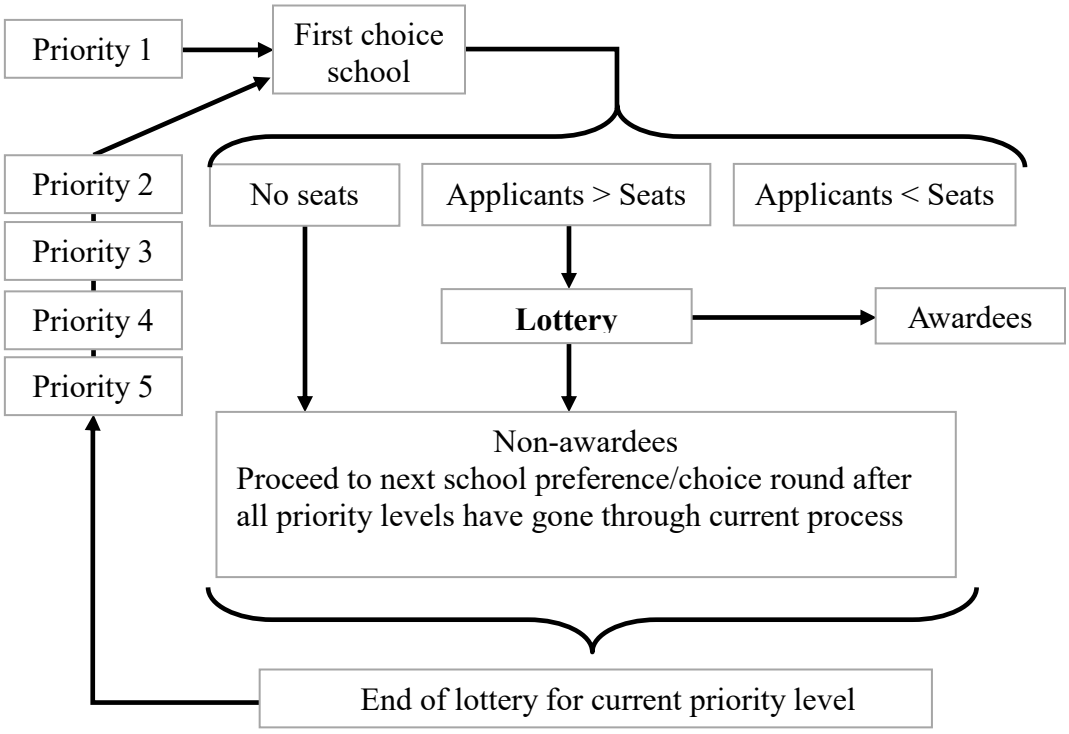

Figure 1. First stage of the Louisiana Scholarship Program award allocation process for the 2012-2013 school year. This figure illustrates the iterative process used to allocate LSP scholarships to students. In addition, this figure highlights the fact that only a subset of students was awarded LSP scholarships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

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 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 two 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 at schools 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 likely are upper bounds of the program's true effect on student achievement.

Data Description

The Louisiana Department of Education (LDOE) provided most of the data for this study in accordance with our data agreement with the state. The LDOE provided:

- Student Information Systems (SIS) files for 2011-12 and 2012-13 which include data on student enrollment and demographic background;
- The 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;[5]

[5] 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.

- State assessment files for the 2011-12 (Baseline), 2012-13 (Year 1 Outcome), 2013-14 (Year 2 Outcome), 2014-15 (Year 3 Outcome), and 2015-16 (Year 4 Outcome) school years, which include data on students' participation in the annual accountability assessments and their scores.[6]

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.[7] 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. In science and social studies, in contrast, students continued to take the LEAP/iLEAP exams aligned with state standards.[8] In 2015-16, Louisiana again switched assessments in response to a standards review.[9] Beginning in 2015-16 (Year 4 Outcome), Louisiana students have taken the LEAP 2025 assessments in grades

[6] 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 one percent of LSP applicants in any given year.
[7] The Louisiana program of assessments offers two alternative assessments for students with disabilities. Performance on these assessments is excluded from our analysis.
[8] 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. 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 in the 2014-15 school year.
[9] The 2015-2016 Louisiana Student Standards professional review involved 100 educators and representatives from universities, the business community, and parent groups across the state. The review resulted in an updated set of state academic standards for students.

3-8 in ELA, math, and science.[10] These assessments, which are aligned with the updated Louisiana student standards, use questions that are academically aligned with the 2014-15 PARCC assessments (Louisiana Department of Education, 2015).

The state-provided assessment data files also include information on student demographics, disability status, and participation in school initiatives such as the free- or 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 applications and who were not multiple birth siblings applying for grades 1 through 5 (totaling 4,499 students). Of these, 2,348 students have outcome data in Year 4 and participated in over-subscription lotteries for their first-choice school, with 49 percent receiving placement. When we focus further on students with baseline achievement data in grades three through four, our analytical sample drops to 780 students. Of these, 328 – or 42 percent – won LSP scholarships to their first choice school. This final sample of students – those with baseline achievement in grades three through four – represent our primary analytical sample of interest.

[10] Students also take a social studies exam; however, this exam is administered as a field test. Performance on this assessment was not made available to the research team for this evaluation.

Analytical Strategy

We begin with a description of our primary analyses, which uses the results of eligible applicants' first-choice school 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 still can 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 students can

receive LSP scholarships 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 four 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 at any point between the 2012-13 and 2015-16 school years [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 his or her first choice school
- A_i is standardized student mathematics or English Language Arts achievement in Year 4 of the program (2015-16)[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," while Rouse (1998) defines enrollment as the number of years enrolled in an attempt to capture potential dosage effects. By defining enrollment conservatively as "ever attending a private school," our study falls in line with the Wolf et al. (2013) evaluation of the DC Opportunity Scholarship Program.

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

We account for nesting of students within lotteries using bootstrapped standard errors (Angrist & Pischke, 2009). The families of students and their post-treatment schools could represent additional nesting factors (Wolf et al., 2013). The results presented here do not account for these sources 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 as lottery includes school of application. Moreover, siblings constitute only 7 percent of our analytical sample and therefore are not a substantial nesting concern.

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 gains for males. Wolf and colleagues also note positive achievement effects for students who were already performing well at baseline. For the most part, we have observed little differentiation in effects associated with student gender and race in our prior evaluations of the LSP (Mills, 2015; Mills & Wolf, 2017a; Mills & Wolf, 2017b). We did observe heterogeneity across baseline achievement terciles in ELA in Year 3 of our outcome evaluation, with the

lowest performers at baseline experiencing test score benefits from the program (Mills & Wolf, 2017b).

Treatment-Control Contrast

While eligible applicants were randomly assigned to receive or to not receive an LSP scholarship to their most-preferred private school, participating families were not required to use the scholarship. Therefore, it is important 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 who received and did not receive LSP scholarships to their first choice schools for the four years following their initial application to the program. The analytical sample presented in Table 1 reflects students who did not list a special education classification on their LSP application, and who were not multiple birth siblings. 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 127 students in 2015-16, are control-group crossovers in our LATE analysis.

While the majority of lottery winners used their scholarships to attend private schools, around 75 percent of students who did not receive scholarships attended public-sector schools in all years of our study. The percentage of first-choice lottery winners attending private schools has declined over time from 77% in Year 1 to only 41% by Year 4. More importantly, Year 4 is the first year in which we observe a larger percentage of first-choice school lottery winners attending public schools (TPS, charter, or magnet) than private schools (44% and 41%, respectively). While our data do not allow us to determine the causes behind this increase in non-

compliance, it is important to note that higher rates of non-compliance will limit the generalizability of our

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)				
Private School	493	77%	58	7%
Public School	113	18%	678	82%
Unknown/Missing School	31	5%	91	11%
Year 2 (2013-14)				
Private School	552	60%	147	14%
Public School	270	30%	777	74%
Unknown/Missing School	93	10%	132	13%
Year 3 (2014-15) - PARCC data				
Private School	684	52%	157	11%
Public School	477	37%	1036	75%
Unknown/Missing School	145	11%	189	14%
Year 4 (2015-16)				
Private School	541	41%	127	9%
Public School	599	46%	1065	77%
Unknown/Missing School	166	13%	190	14%

Notes. All students participated in LSP lotteries. Analysis sample excludes students with disabilities and multiple birth siblings. Year 1 is restricted to students applying for grades 3 through 8 for the 2012-13 school year. Year 2 is restricted to applicants for grades 2 through 7. Year 3 is restricted to applicants for grades 1 through 6. Year 4 is restricted to applicants for grades 1 through 5. *Source*. Authors' calculations.

findings, as our primary analysis provides treatment effect estimates for students who continue to comply in Year 4 with their initial lottery assignment.

Table 1 additionally provides a first look at study sample attrition rates for our treatment and control groups.[14] Attrition represents no more than 14 percent of either group across all four years of data. The difference in attrition rates between treatment and control groups is slightly

[14] Table 1 presents comparisons of raw differences. For a more detailed analysis which accounts for student lotteries, see Appendix Table A1.

larger than the acceptable level (What Works Clearinghouse, 2014) in Year 1, with more attrition in the control than the treatment group (9% versus 4%). We do not observe statistically significant differences in attrition rates between treatment and control groups in Years 3 or 4.[15]

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 groups 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 historical analytical sample which requires baseline achievement for sample inclusion.[16] Columns 2 and 3 present simple averages for each variable for the treatment group and control group, respectively. Column 4 reports the raw difference in these averages. Simple comparisons of raw averages are problematic because they do not account for the fact that students were randomly assigned to treatment or control status within their first-choice school. Instead, the results presented in column 5—"Adjusted Diff."—account for differential probabilities of treatment selection across first-choice school lotteries via fixed effects; and therefore are the focal point of this analysis.

The results are favorable for our analysis. Nearly all of the estimated adjusted differences between lottery winners and losers are statistically insignificant, suggesting that we have

[15] 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 that differential attrition strongly influences our primary LATE estimates.

[16] Other requirements to be in this analytical sample included students not listing a special education classification on their application, not being a multiple birth sibling, having baseline test data in grades three through four, and experiencing a lottery for their first-choice school. A companion analysis for an analytical sample that does not require baseline achievement for sample inclusion is in Appendix Table B1.

adequately identified random lotteries in our analytic sample. Consistent with our prior evaluations of the LSP (Mills, 2015; Mills & Wolf, 2017a; Mills & Wolf, 2017b), we observe that lottery winners provided significantly fewer school preferences on average than lottery losers. This difference is not, however, statistically significant at the .05 level. New to this evaluation, we observe a statistically significant difference between the treatment and control groups on the likelihood of being African American, even with comparisons made within-lottery. Given these differences, our preferred model includes controls for the full set of variables examined in Table 2.

Table 2.

Baseline Equivalence of Treatment and Control Groups on Covariates for BA Sample, Year 4

	N	Treatment Avg.	Control Avg.	Raw Diff.	Adjusted Diff.	*s.e.*
	(1)	(2)	(3)	(4)	(5)	(6)
Female	779	0.54	0.50	0.03	0.00	0.04
Race/Ethnicity						
African American	779	0.88	0.92	-0.04	-0.05**	0.03
Hispanic	779	0.03	0.01	0.01	0.02	0.02
White	779	0.06	0.04	0.02	0.02	0.02
Other	779	0.03	0.02	0.01	0.01	0.02
Limited Eng. Proficiency	780	0.02	0.01	0.01	0.00	0.01
Free- or Reduced-Price Lunch	761	0.78	0.93	-0.15	0.00	0.03
# of Sch. Preferences	780	1.87	2.34	-0.47	-0.16	0.10
Standardized Performance[a]						
ELA Scale Score	780	-0.34	-0.34	0.00	-0.01	0.09
Math Scale Score	779	-0.34	-0.34	0.01	-0.04	0.08
Science Scale Score	779	-0.44	-0.44	0.00	0.07	0.07
Social Studies Scale Score	779	-0.35	-0.34	-0.01	-0.01	0.07

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

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

Notes. BA Sample requires students to have baseline achievement in grades 3 or 4. Analysis sample excludes students with disabilities and multiple birth siblings. The analysis sample represents LSP applicants to grades 1 through 5 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 3 through 4. *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. *Raw Diff.* is the raw difference in means between the treatment and control groups. *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

Results

The following sections present our preliminary estimates of the LSP's impact on student achievement after four years. Throughout, we present results for two analytical samples. The first follows the sample restrictions applied in our previous evaluations of the LSP (Mills, 2015; Mills & Wolf, 2018a; Mills & Wolf, 2018b), which requires baseline achievement for sample inclusion (hereafter referred to as the "BA Sample"). This restriction is motivated by the important power of pre-tests in explaining variation in test outcomes (Peterson & Howell, 2004). Practically speaking, however, this requirement limits our sample to students with test scores in grades three or four in 2011-12, the year immediately prior to their random assignment, which may severely limit the generalizability of the findings from our primary analyses. We therefore additionally present results for an expanded sample of students by dropping the requirement for baseline achievement data (hereafter referred to as the "NBA sample"). This second analytical sample only requires that students applied for placement in 2012-13 (Year 1 of our study) in grades one through five. Assuming we have correctly identified lotteries for first-choice schools, the LATE estimates generated for both the BA and NBA samples will present unbiased estimates of the effects of using an LSP scholarship to attend an LSP school four years after initial assignment.

Our prior work reported large declines in both math and reading achievement after one year of usage (Mills, 2015). These negative effects dropped by half after two years of usage (Mills & Wolf, 2017a) and were not statistically significant after three years of usage (Mills & Wolf, 2017b). In contrast to our previous research, the results presented here indicate large negative effects of LSP voucher usage after four years, especially in math. We additionally observe some statistically significant negative effects of LSP scholarship usage on ELA achievement; however, these results are not consistent across our two analytical samples.

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

Table 3 reports our primary effect estimates for two analytical samples. First, we present results for our historically preferred sample, labeled BA, which requires baseline achievement for a student to be included (columns 1-5). Then we examine results from an expanded sample, labeled NBA, which does not require baseline achievement.

Columns 1 and 6 display coefficient estimates for first stage regressions which use scholarship award to predict the likelihood of continuing to attend an LSP private school in 2015-16 for both analytical samples. These coefficients, which allow us to identify the percentage of individuals who are complying with their initial lottery assignment, indicate moderate rates of compliance during the four-year period.[17] Specifically, 61 percent of individuals in the BA sample and 65 percent of individuals in the NBA sample complied with their initial lottery assignment at any point between 2012-13 and 2015-16.

Our primary estimates of the impact of scholarship usage on the student achievement follow. First we present results from a simple model which only controls for first-choice school lottery fixed effects (columns 2 and 7). Next, we include an indicator for taking the same grade-level assessment in two consecutive years (columns 3 and 8). Fully specified models, which additionally control for student demographics and, for the BA sample, baseline achievement, appear in columns 4 and 9. These are our preferred estimates of the impact of the LSP given the small number of significant differences in baseline characteristics between treatment and control groups observed in Table 2.

[17] Compliance is identified as continuing to observe the result of the lottery. In our context, lottery compliers are represented by treatment group students who enrolled in private schools and control group students not enrolled in private schools. Non-compliance, in contrast, is represented by treatment group individuals who did not use an LSP scholarship to attend a private school and control group students enrolled in private schools.

Table 3.

Estimated Effects of Ever Using an LSP Voucher on Student Achievement after Four Years

	Baseline Achievement (BA) Sample					No-Baseline Achievement (NBA) Sample				
		LATE					LATE			
	First Stage	Simple Model	+ Test Retake	Fully Specified	Omitting New Orleans	First Stage	Simple Model	+ Test Retake	Fully Specified	Omitting New Orleans
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
English Language Arts	0.61***	-0.12	-0.08	-0.11	-0.09	0.65***	-0.13*	-0.11	-0.22***	-0.21**
	(0.04)	(0.18)	(0.17)	(0.15)	(0.15)	(0.02)	(0.08)	(0.08)	(0.08)	(0.09)
Mathematics	0.61***	-0.27**	-0.25**	-0.28**	-0.30***	0.65***	-0.32***	-0.30***	-0.39***	-0.39***
	(0.04)	(0.12)	(0.12)	(0.11)	(0.11)	(0.02)	(0.07)	(0.07)	(0.08)	(0.08)
Science	0.60***	-0.27	-0.24	-0.31**	-0.23	0.65***	-0.16*	-0.14	-0.21**	-0.23***
	(0.04)	(0.19)	(0.19)	(0.15)	(0.14)	(0.02)	(0.09)	(0.09)	(0.10)	(0.09)
Controls										
Test Re-take			X	X	X			X	X	X
Baseline Achieve.			X	X	X					
Demographics				X	X				X	X
# of Sch. Choices				X	X				X	X
New Orleans				X					X	
N	719 - 735		719 - 735	699 - 715	538 - 552	2182 - 2210		2182 - 2210	1606 - 1632	1405 - 1430
Lotteries	103 - 103		103 - 103	103 - 103	73 - 74	255 - 255		255 - 255	214 - 215	178 - 180

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

Notes. *Simple Model* refers estimations that only controls for first-choice school lottery 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 (BA sample only), student demographics, number of school preferences offered, and geography. Columns 5 and 10 omit 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.

While our previous research suggested that the initial large negative test score impacts for LSP voucher users are declining over time, Table 3 presents several cases in which the general effect of voucher usage is negative and statistically significant after four years. First, for the BA sample, we only observe consistent statistically significant negative impacts in math, with LSP voucher users scoring nearly 30 percent of a standard deviation lower than their control-group counterparts in Year 4. We additionally observe a statistically significant negative impact of LSP voucher usage on science test scores on par with those observed in math in our fully specified and researcher preferred model. Point estimates for ELA are not statistically significant; however, this null finding may be due in part to noisy estimation, as the standard errors are large across specifications. The analyses presented in columns 1 through 5 likely suffer from low statistical power, a result driven in part by the requirement of baseline achievement which effectively restricts our analysis to just two grade cohorts of LSP applicants.

We attempt to address this statistical power issue by dropping the baseline achievement requirement for inclusion in our analytical sample. The results presented in columns 6 through 10 of Table 3 are based on the experiences of students applying to use LSP vouchers in 2012-13 in grades one through five and are estimated using a statistical model that omits baseline test scores. Unlike in the BA sample, the estimated impact of LSP voucher usage is negative across all subjects in fully specified models in this expanded sample. As before, the estimated effects are quite large, as LSP voucher users appear to score 22 percent of a standard deviation lower than control group students in ELA, 39 percent of a standard deviation lower in math, and 21 percent of a standard deviation lower in science (Table 3, Column 9). While the estimated impact for science shrinks in magnitude relative to the BA sample, the relatively larger coefficients in ELA and math in the NBA sample correspond with our prior research (Mills & Wolf, 2017b),

which reports evidence of larger negative test score effects of the LSP on applicants to earlier grades.

Table 3 additionally presents models for samples of students applying to LSP voucher schools outside New Orleans (columns 5 and 10). These models attempt to provide a cleaner look at the effects of the program's statewide expansion because New Orleans was already home to both a diverse public charter school market (Harris & Larsen, 2018) as well as the pre-existing piloted version of the LSP. It is possible, for example, that New Orleans private schools are more equipped to work with the less advantaged population of students applying to the LSP due to their experience with the pilot program since 2008. In contrast, estimated effects also depend on the quality of the counterfactual environment (i.e., public schools), and research indicates that student outcomes in New Orleans public schools have improved noticeably in the wake of education reforms implemented during rebuilding efforts following the destruction of Hurricane Katrina (Harris & Larsen, 2018). Both factors suggest LSP effects may differ based on whether students were inside or outside of New Orleans.

In contrast, the results presented in columns 5 and 10 are negligibly impacted by this restriction. Estimates are largely similar in magnitude and mirror the statistical significance of those in our fully specified models (columns 4 and 9). In short, the inclusion of New Orleans students in our preferred models does not appear to affect the substantive conclusions of our evaluation.

Next, we examine how the LSP effects vary over time for these samples of students. Our prior research indicates large negative impacts on ELA and mathematics in the first year of participation (Mills, 2015) that appear to diminish somewhat by Year 2 (Mills & Wolf, 2017a) and are not statistically significant by Year 3 (Mills & Wolf, 2017b). Figures 2 and 3 present

LATE estimates for ELA and mathematics for Years 1 through 4 for consistent samples of students contributing to the analyses presented in Table 3.[18] Figure 2 presents results for students in the BA Sample and Figure 3 focuses on students in the NBA Sample. In both figures, impacts are estimated by conditioning the ever enrollment variable to the possible time period in which a student could enroll in a private school.[19]

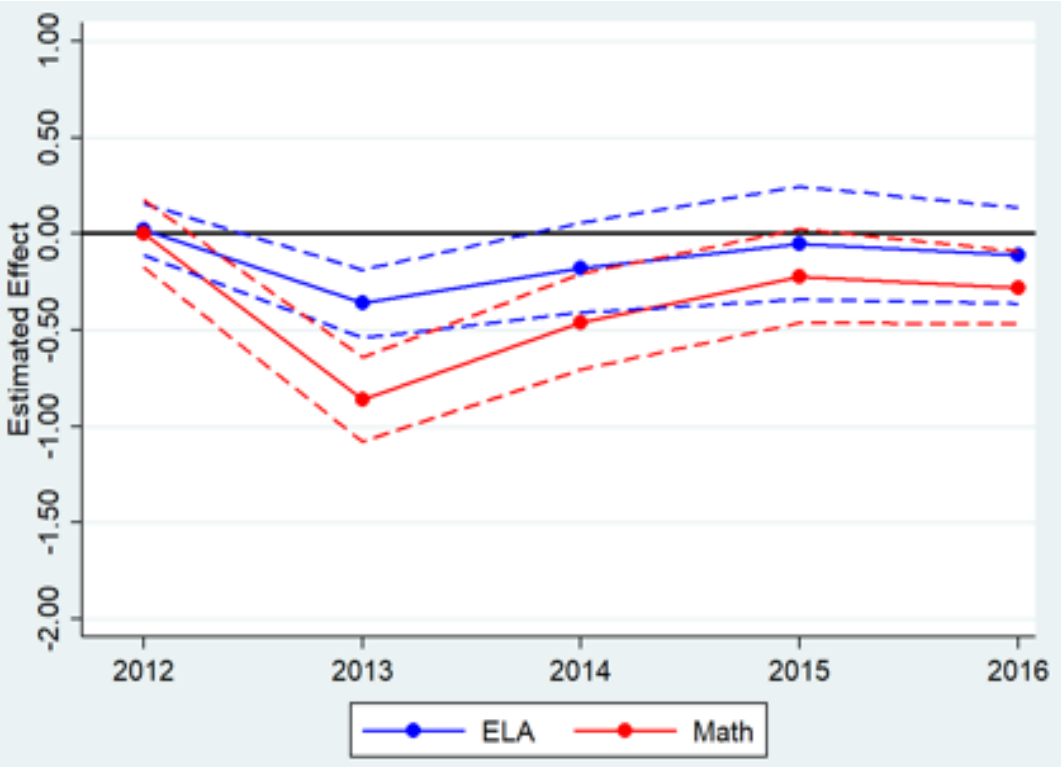

Figure 2. Estimated Local Average Treatment Effects over time for BA sample. Figure presents point estimates from fully specified models for 2011-12 (baseline) through 2015-16 for ELA and math. Results are presented for a consistent sample of students with Spring 2016 outcome data. ELA and math results are based on student achievement on the Louisiana state assessments (LAA) in 2011-12 through 2013-14, PARCC assessments in 2014-15, and LAA in 2015-16. Dashed lines represent 90% confidence intervals for the performance averages.

[18] We present these figures of the impacts of the LSP on a consistent sample of students across time because the annual samples of students with outcome data change over the outcome time horizon. Focusing on a smaller but consistent sample of students means that we can rule out changing student populations as a factor affecting any changes in program effects observed across the years.

[19] For example, Year 2 effects are estimated for students who have complied with lottery assignment at some point during Year 1 and Year 2. Year 3 effects are estimated for students who complied with lottery assignment at some point during Year 1, Year 2, and Year 3, and so on.

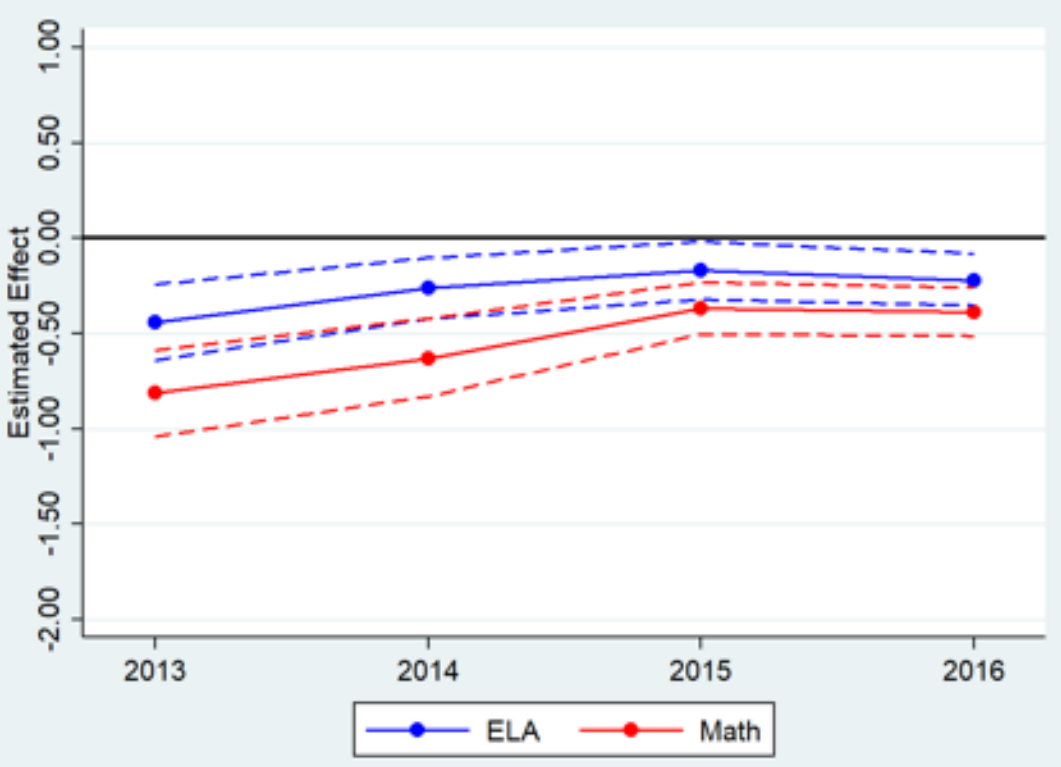

Figure 3. Estimated Local Average Treatment Effects over time for NBA sample. Figure presents point estimates from fully specified models for 2013-14 through 2015-16 for ELA and math. Results are presented for a consistent sample of students with Spring 2016 outcome data. ELA and math results are based on student achievement on the Louisiana state assessments (LAA) in 2012-13 through 2013-14, PARCC assessments in 2014-15, and LAA in 2015-16. Dashed lines represent 90% confidence intervals for the performance averages.

Consistent with our prior work, we observe large declines in ELA and math performance in the first year of voucher usage that become less negative in Years 2 and 3. Effects are generally worse in math than in ELA. By Year 4, however, we observe a direction reversal for the voucher usage effect estimates. Effects are slightly more negative in magnitude in Year 4 relative to Year 3 across all tests and samples and are statistically significant for math in both samples and ELA in the NBA sample only.

Subgroup Effects

The results presented in Table 4 allow us to determine if LSP voucher usage effects are differentiated by gender, race, and baseline achievement. Results are presented for both simple

models which control only for LSP assignment lottery and fully specified models which additionally control for demographics and, in the BA sample, baseline achievement. All models estimate the impact of using an LSP voucher to attend a private school at any point between 2012-13 and 2015-16.

In general, we do not observe consistent evidence that LSP voucher usage effects are moderated by gender. Difference estimates vary across models—sometimes positive and sometimes negative—while only one point estimate is statistically significant. In contrast, we do observe evidence suggesting effects were experienced differently by students of different racial backgrounds. Treatment effects are generally less negative for African American students relative to other qualified LSP applicants (Figure 4), and estimated differences are statistically significant in all but one fully specified model. Previous voucher evaluations have reported similar evidence of effect moderation by race (Mayer et al., 2002; Barnard et al., 2003); however, Year 4 is the first time we have observed such evidence for the LSP. Moreover, unlike studies of the New York Scholarship Program, in which positive effects were observed for African American students (Mayer et al., 2002; Barnard et al., 2003), the point estimates reported in Table 4 generally indicate null to negative effects of LSP voucher usage for this group.

Results are similar for baseline achievement models, with one exception. While we largely do not observe differences by baseline achievement category, students scoring in the middle third of math achievement at baseline have very large and statistically significant

Table 4.

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

	English Language Arts						Mathematics					
	BA Sample			NBA Sample			BA Sample			NBA Sample		
	N	Simple	Full Model	N	Simple	Full Model	N	Simple	Full Model	N	Simple	Full Model
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Gender												
Female students	385	-0.21	-0.05	1182	-0.12	-0.22**	375	-0.45***	-0.35**	1173	-0.35***	-0.37***
		(0.20)	(0.18)		(0.10)	(0.10)		(0.16)	(0.15)		(0.09)	(0.10)
Male students	349	-0.01	-0.17	1028	-0.15*	-0.22*	347	-0.11	-0.21	1029	-0.29***	-0.40***
		(0.19)	(0.18)		(0.09)	(0.12)		(0.16)	(0.15)		(0.09)	(0.11)
Difference		-0.19	0.12		0.03	0.00		-0.34*	-0.13		-0.06	0.02
		(0.19)	(0.19)		(0.11)	(0.14)		(0.21)	(0.18)		(0.11)	(0.14)
Race/Ethnicity												
African American students	658	-0.08	-0.04	1983	-0.09	-0.16*	648	-0.26*	-0.21	1976	-0.29***	-0.34***
		(0.19)	(0.15)		(0.08)	(0.08)		(0.14)	(0.13)		(0.08)	(0.09)
Other students	76	-0.52	-0.58	227	-0.53**	-0.57**	74	-0.60**	-0.75***	226	-0.65***	-0.73***
		(0.42)	(0.41)		(0.23)	(0.24)		(0.28)	(0.18)		(0.22)	(0.20)
Difference		0.44	0.54		0.44*	0.41*		0.34	0.54**		0.36	0.39*
		(0.43)	(0.40)		(0.23)	(0.24)		(0.33)	(0.21)		(0.25)	(0.23)
Baseline achievement												
Lower Third	249	0.08	-0.10				245	0.25	0.20			
		(0.33)	(0.28)					(0.22)	(0.22)			
Middle Third	254	0.05	-0.06				253	-0.78***	-0.82***			
		(0.25)	(0.22)					(0.17)	(0.17)			
Upper Third		-0.16	-0.01					-0.40	-0.24			
		(0.28)	(0.28)					(0.26)	(0.31)			

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

Notes. BA Sample requires students to have baseline achievement in grades 3 or 4. *NBA Sample* does not require baseline achievement. Performance measures standardized within grade based on control group score distributions. *Simple Model* refers estimations that only controls for first-choice school lottery fixed effects. *Full Model* refers to models controlling for test retaking, baseline achievement (BA sample only), student demographics, number of school preferences offered, and geography. 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.

negative effects by year 4. We cannot, however, make comparisons across point estimates because these estimates are drawn from separate regressions. Interestingly, students initially testing in the bottom third of the test distribution at baseline are shown, at times, to be outperforming their counterparts by Year 4. However, none of the estimated effects are significant.

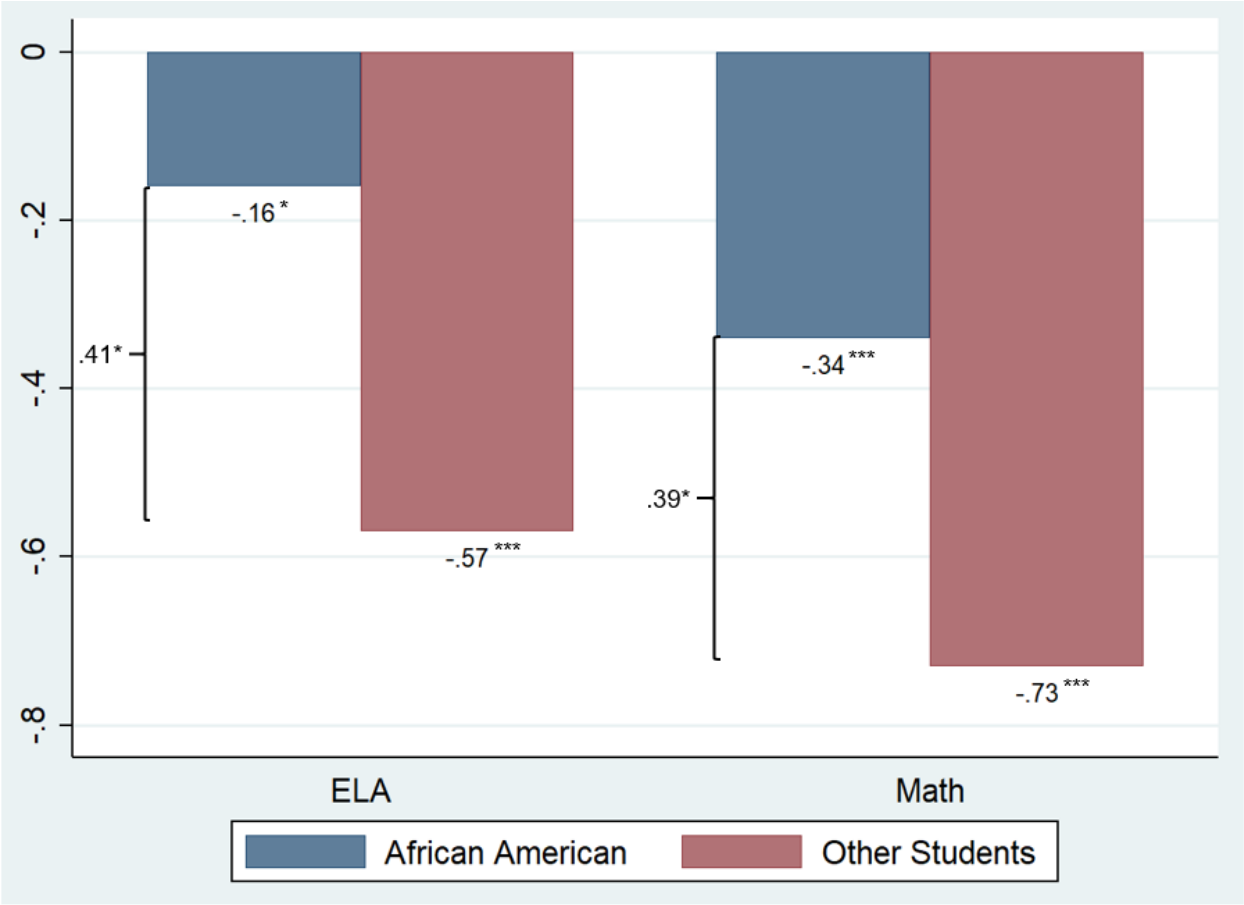

Figure 4. Differential effects of LSP usage by race for NBA sample.
Figure presents separate estimates of the impact of ever using an LSP voucher to attend a private school between 2012-13 and 2015-16 for African American students and students of other races and ethnicities for ELA and math. Models are fully specified. *** - p<.01, ** - p<.05, * - p<0.10.

In short, while we do not observe strong evidence that effects were differentiated by gender or baseline achievement, our results indicate African American students generally experienced less negative achievement impacts from LSP voucher usage than non-African American students.

Does Differential Attrition Impact the Estimates?

Finally, we examine if differential attrition between treatment and control group members is impacting our results. Table 1 indicates noticeable sample attrition which has generally increased over time. Moreover, attrition rates appear to differ between treatment and control group members, with generally higher observed attrition rates among those initially assigned to the LSP control group. These differences, if driven by non-random factors, raise concerns of biased effect estimates (Gerber & Green, 2012; Lee, 2009; What Works Clearinghouse, 2014). If, for example, those in the control group with lower expected outcomes both in public and private schools leave the sample with higher probability, our LATE estimates will be negatively biased.

In Table 5, we examine the extent to which differential attrition impacts 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.20 In our case, we are concerned about disproportionately high levels of attrition among low performers in the control 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 4 performers in the treatment group. Similarly, we create a lower bound by excluding the highest performers.

[20] 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 LSP applicants who were assigned an LSP scholarship but decided to forgo their scholarship and instead enroll in a private school at their own expense. While we cannot validate this assumption empirically, it seems unlikely such "defiers" exist in our data – especially given the program's income threshold.

Table 5.

Examining Effects of Differential Attrition

	BA Sample			NBA Sample		
	N	Simple	Full Model	N	Simple	Full Model
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: English Language Arts						
Primary LATE	735	-0.12	-0.12	2210	-0.13*	-0.21**
		(0.18)	(0.16)		(0.08)	(0.08)
Lee Lower Bound	724	-0.25	-0.22	2190	-0.19**	-0.27***
		(0.17)	(0.15)		(0.08)	(0.08)
Lee Upper Bound	725	0.01	-0.03	2190	-0.04	-0.13
		(0.16)	(0.15)		(0.07)	(0.08)
Panel B: Mathematics						
Primary LATE	723	-0.27**	-0.28**	2202	-0.32***	-0.38***
		(0.12)	(0.11)		(0.07)	(0.08)
Lee Lower Bound	713	-0.36***	-0.35***	2181	-0.37***	-0.43***
		(0.11)	(0.10)		(0.07)	(0.08)
Lee Upper Bound	708	-0.13	-0.17	2181	-0.25***	-0.32***
		(0.12)	(0.12)		(0.07)	(0.07)
Controls						
Test Re-take		X	X		X	X
Baseline Achieve.			X			X
Demographics			X			X
# of Sch. Choices			X			X
New Orleans			X			X

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

Notes. BA Sample requires students to have baseline achievement in grades 3 or 4. *NBA Sample* does not require baseline achievement. *Simple Model* refers estimations that only controls for first-choice school lottery fixed effects. *Full Model* refers to models controlling for test retaking, baseline achievement (BA sample only), student demographics, number of school preferences offered, and geography. 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 5 presents little evidence that support the claim that our effect estimates are biased by differential attrition. Nearly all point estimates indicate that LSP voucher usage negatively impacted student achievement, especially in math. Even when they are not statistically significant, we observe negative effect estimates with one exception: a small positive estimate for the upper bound effect on ELA in a simple specification for the BA sample that is not

statistically significant. In short, there is no strong evidence that our findings are driven by differential attrition.

Exploring LSP's Accountability Provisions

School accountability forms an important, and contentious, topic in school choice theory. Some commentators argue, for example, that schools are naturally held accountable in school choice systems by families who can always demonstrate their dissatisfaction by simply moving to another school (Greene, 2011). Other observers, however, argue that transaction costs associated with changing schools could hinder the ability of families to hold schools accountable by voting with their feet, especially if the families are relatively low in social capital (Goldhaber et al., 2005). Given these concerns, other commentators have argued for a strong centralized regulator to ensure schools are properly serving their students', as well as society's, educational needs (Smarick, 2012).

Since 2010, Louisiana has placed a strong emphasis on the importance of centralized accountability mechanisms in a choice-heavy education system. Public schools, for example, are subject to potential sanctions and closure due to poor performance on the state's School Performance Score (SPS), an accountability index that takes into account student achievement and graduation rates. This regulatory emphasis is also present in the LSP, with participating private schools subject to similar oversight by the LDOE. Private schools in the LSP are required to comply with state health and safety codes, are subject to financial audits, and can potentially face sanctions for poor performance of their LSP voucher users. Private schools in the LSP that fail to meet these requirements or improve their students' test score outcomes can be prohibited from enrolling any new students via the program.

In this section we conduct an exploratory analysis designed to provide an initial look at the relationship between the LSP's accountability mechanisms and student achievement. We examine how local average treatment effects differed for students whose first-choice school was sanctioned at any point between the 2012-13 and 2015-16 school years (hereafter referred to as "future sanctioned schools") as well as for students listing other, never sanctioned, LSP private schools as their first choice school. In effect, this analysis provides a thought experiment as to the relationship between later sanctions and the effects of the LSP on student achievement. We stress that this analysis does not provide causal estimates of the effects of LSP sanctions on student achievement. Sanctions occurred after the point of random assignment used in our analysis and will reflect, in part, the general effect of the program. We caution the reader that the findings presented here merely describe the landscape of accountability practice in the LSP and the extent to which the sanctioning element of the system was implemented according to its design.

Between 2012-13 and 2015-16, 35 participating LSP private schools were sanctioned (see Table 6). The overwhelming majority of LSP sanctions are for academic reasons. All but 4 schools were sanctioned either for having an SCI equivalent to an F according to the state's school letter grading system (13 total) or having fewer than 25 percent of LSP students failing to score proficient or above (17 total). Given that the majority of sanctions occurred in response to poor academic performance ratings, it is natural to assume that, in general, treatment effects will be more negative among students in future sanctioned schools.

Table 6

List of School Sanctions, by Year

	Sanction cause	Number of schools
2012-13	SCI equivalent to an F letter grade	---
	<25% proficient	7
	Financial Audit Concerns	---
2013-14	SCI equivalent to an F letter grade	---
	<25% proficient	---
	Financial Audit Concerns	1
2014-15	SCI equivalent to an F letter grade	12
	<25% proficient	8
	Financial Audit Concerns	---
2015-16	SCI equivalent to an F letter grade	1
	<25% proficient	2
	Financial Audit Concerns	4
Total		49

Source. Authors' calculations.

Table 7 describes students who listed future sanctioned LSP schools as their most-preferred LSP school compared with all other eligible applicants for the 2012-13 school year. Column 4 presents results from independent *t*-tests of difference in means between the two groups. Students listing future sanctioned schools as their first choice differ from other applicants along a number of demographic dimensions. They are more likely to be African American and less likely to be White; they listed more school preferences on their applications; and, they performed noticeably worse on the Louisiana state assessments at baseline (2011-12) than other applicants. In sum, while LSP applicants are generally disadvantaged relative to the average Louisianan public school student, the results presented in Table 7 suggest that students listing future sanctioned schools as their top choice on their LSP application are further disadvantaged relative to other applicants.

Table 7

Baseline Equivalence of Treatment and Control Groups on Covariates, Year 4

	N	Sanctioned school listed as first choice avg.	Other applicants avg.	Diff.	*s.e.*
	(1)	(2)	(3)	(4)	(5)
Female	9,736	0.51	0.50	0.01	0.01
Race/Ethnicity					
African American	9,736	0.92	0.80	0.12***	0.01
Hispanic	9,736	0.02	0.03	-0.01***	0.00
White	9,736	0.04	0.13	-0.09***	0.01
Other	9,736	0.02	0.04	-0.02***	0.00
Limited Eng. Proficiency	3,957	0.02	0.01	0.01**	0.00
Free-or Reduced-Price Lunch	5,470	0.94	0.94	0.00	0.01
# of Sch. Preferences	9,736	2.15	1.90	0.25***	0.03
Grades					
Kindergarten	9,736	0.19	0.23	-0.05***	0.01
First grade	9,736	0.13	0.12	0.01	0.01
Second grade	9,736	0.10	0.11	-0.01	0.01
Third grade	9,736	0.11	0.10	0.01	0.01
Fourth grade	9,736	0.08	0.11	-0.03***	0.01
Fifth grade	9,736	0.08	0.09	-0.01	0.01
Sixth grade	9,736	0.09	0.08	0.01	0.01
Seventh grade	9,736	0.09	0.05	0.04***	0.01
Eighth grade	9,736	0.06	0.03	0.02***	0.00
Ninth grade	9,736	0.05	0.04	0.00	0.00
Tenth grade	9,736	0.02	0.01	0.01**	0.00
Eleventh grade	9,736	0.01	0.01	0.00	0.00
Twelfth grade	9,736	0.00	0.00	0.00	0.00
Standardized Performance					
ELA Scale Score	3,953	-0.47	-0.38	-0.10***	0.03
Math Scale Score	3,953	-0.57	-0.44	-0.14***	0.03
Science Scale Score	3,945	-0.59	-0.49	-0.10***	0.03
Social Studies Scale Score	3,945	-0.53	-0.37	-0.16***	0.03

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

Notes. Sample represents all eligible applicants for LSP scholarships in the 2012-13 school year. Treatment refers to students receiving LSP scholarships to their first choice private school. Diff is the difference between students who received scholarships to their first choice school and other eligible applicants. "s.e." indicates standard error of the difference. Test scores are standardized within grade based on the observed distributions of scale scores across Louisiana. Source. Authors' calculations.

Table 8 presents Local Average Treatment Effect (LATE) estimates for students listing future sanctioned schools as their first choice school and other applicants. Results are presented for both ELA and math, as well as for analytical samples requiring baseline achievement (BA sample) and samples that relax this requirement (NBA sample). Simple models include controls only for first-choice lotteries while fully specified models additionally control for student demographics and, in BA samples, baseline achievement.

Table 8.

Local Average Treatment Effects for Students Attending Sanctioned and Non-sanctioned Schools, across All Years

	BA Sample			NBA Sample		
	N	Simple	Full Model	N	Simple	Full Model
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. English Language Arts						
Sanctioned schools	388	-0.23	-0.15	1352	-0.22**	-0.22**
		(0.30)	(0.24)		(0.10)	(0.10)
Non-sanctioned schools	347	0.01	-0.07	858	0.03	-0.21
		(0.24)	(0.21)		(0.13)	(0.14)
Difference		-0.24	-0.07		-0.24	0.00
		(0.38)	(0.32)		(0.16)	(0.17)
Panel B. Mathematics						
Sanctioned schools	379	-0.40**	-0.27	1346	-0.36***	-0.40***
		(0.19)	(0.18)		(0.10)	(0.11)
Non-sanctioned schools	344	-0.13	-0.28*	856	-0.24**	-0.36***
		(0.15)	(0.15)		(0.10)	(0.11)
Difference		-0.27	0.01		-0.12	-0.04
		(0.22)	(0.23)		(0.14)	(0.15)
Controls						
Test Re-take		X	X		X	X
Baseline Achieve.			X			X
Demographics			X			X
# of Sch. Choices			X			X
New Orleans			X			X

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

Notes. BA Sample requires students to have baseline achievement in grades 3 or 4. *NBA Sample* does not require baseline achievement. 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.

The row titled "Difference" is the focus of this analysis as it represents the estimated difference in LATE between students applying to future sanctioned schools and other students. In general, none of the estimated differences reported in Table 8 are statistically significant at conventional levels. Moreover, estimated differences are often not very large in magnitude in models controlling for student demographics. In summation, we do not find strong evidence of differential effects experienced between students attending future sanctioned schools and students attending other types of private schools after four years.

We caution, however, that this analysis is not causal in nature. While our analysis leverages oversubscription lotteries to estimate treatment effects, the comparisons between students applying to future sanctioned schools and other schools are merely exploratory. A major problem facing this analysis is that school sanctions are endogenously related to treatment effects: almost by definition we should expect students applying to schools that were later sanctioned for poor performance to have experienced more negative treatment effects. The relationship is nearly axiomatic.

At the same time, this analysis demonstrates that, among eligible LSP applicants, students listing future sanctioned schools as their most-preferred private school are disadvantaged across several demographic measures relative to other LSP applicants. It is possible that many of the private schools opting to participate in the LSP were simply not prepared to meet the educational needs of large numbers of these students transferring into their schools all at once. Given that the majority of sanctions were issued due to poor academics, this scenario seems likely.

Conclusion

This study examines how the statewide expansion of the Louisiana Scholarship Program (LSP)—one of the largest school voucher programs in the U.S.—affected student achievement after four years. This research contributes to the existing literature on the participant effects of publicly funded voucher programs for three 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. Finally, this study examines

achievement effects after four years, which allows for a more though understanding of how the effects of this particular voucher program evolved over time.

In contrast to our earlier research, which reported large negative impacts of LSP voucher usage after one year (Mills, 2015) that improved over time (Mills & Wolf, 2017a; Mills & Wolf, 2017b), the results presented here indicate large negative effects of LSP voucher usage after four years, especially in math. These negative effects are robust to alternative sample selection requirements. Some negative effects of LSP usage are observed on English Language Arts test scores after four years, but those findings are sensitive to different samples and definitions of enrollment. In addition, similar to our prior work, we observe little evidence of differentiation in general effects by gender. However, in contrast to findings presented in our previous studies, African American students experienced less negative impacts of voucher usage relative to non-African American students.

Our analysis suffers from several limitations. First, we analyze achievement using student performance on Louisiana state assessments which are criterion-referenced exams aligned with Louisiana's state public school academic standards. These results could, for example, reflect a lack of alignment of private school curricula with the Louisiana state standards rather than true negative impacts of LSP schools on student achievement. In the few years and grades where a less-aligned test was administered, the negative achievement impacts of the LSP were smaller and, in some cases, not statistically significant (Mills & Wolf, 2017a).

Moreover, the assessment regime switched twice in the observed four-year period. While we can address this problem of changing tests in part through standardizing test scores within each year, it is important to recognize the potential disruptive impact test changes can have on both students and schools. Finally, we can only observe test scores on the Louisiana assessments

for students in grades three through eight, which potentially limits the generalizability of results to all LSP applicants. Most noticeably, we are unable to examine effects for applicants applying for vouchers in 9th grade and beyond. Fortunately, in a companion report to this analysis, Holmes, Mills, and Wolf (2019) are able to shed some light on the experiences of older applicants by examining the impact of LSP voucher use on their rates of college enrollment.

Second, our historical requirement that students have baseline achievement data for inclusion in the primary analytical sample has left us with a particularly small sample by Year 4. Taken jointly, the Louisiana test-based accountability system's focus on student test scores in grades three through eight and our baseline achievement requirement limits our primary analytical sample to students who tested in grades three or four in the baseline year of 2011-12. Not only does this requirement leave us with a small sample that limits the statistical power of our analysis, it also restricts our ability to generalize findings beyond these grade cohorts. Nevertheless, we observe similar negative and statistically significant effect estimates when we drop this baseline achievement requirement and expand the sample to all applicants applying for placements in grades one through five in 2012-13. The similarity of these findings suggests that voucher usage has negatively impacted student achievement among those students applying to earlier grades.

Finally, we cannot offer a definitive conclusion as to what might have caused both the observed negative effects as well as the apparent reversal in trend towards improvement reported in our prior work. In earlier reports, we speculated that negative effects were driven in part by the challenges associated with the dramatic expansion of the program statewide. Private schools, for example, had little time to prepare for the incoming students and had no previous experience in administering the Louisiana state assessments. We would expect, however, that negative

effects associated with a lack of preparation among private schools would diminish over time as the program gained maturity. Similarly, while student transfers into a new school are generally negatively correlated with student achievement, this disruption, too, should dissipate over a four year period. In contrast, even when we focus on students who are still attending private schools after four years, we continue to observe negative achievement effects in math and science.

Moreover, it is possible that the promising effects observed after three years reflected a decrease in accountability pressures facing public schools rather than true improvement on the part of private schools. When Louisiana temporarily switched to PARCC assessments aligned with Common Core standards in 2014-15 (Year 3 of our analysis), the Louisiana Department of Education treated the school year as transitionary, as public schools were held blameless for their students' performance on the new state assessments. While the transitionary period also applied to LSP participating private schools, it is possible the Year 3 results reflect, in part, a differential response by public schools to this temporary relaxation of accountability pressure. In 2015-16, or Year 4 of our analysis, the Louisiana assessments were again high stakes tests for both public and LSP private schools. Therefore, it is possible that the null results in Year 3 were more reflective of the decreased accountability pressures experienced by public schools during the transitionary period than a true improvement in performance among LSP voucher users.

Unfortunately, our study is only designed to identify if LSP voucher use impacted student achievement over time. It cannot identify the causes behind the observed effects. All we can say is that, after four years, students receiving LSP vouchers in the first cohort of applicants to the statewide program via lottery who used them to enroll in private schools were performing behind their control-group counterparts on Louisiana state assessments of math, science, and in some specifications also ELA.

The purpose of this work is to provide the most rigorous assessment of the effect of the Louisiana Scholarship Program on the student achievement of participants. 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. While the early negative effects dissipated somewhat over time, we continue to observe negative test score effects of LSP voucher usage that are especially large in math, four years after initial assignment.

Sources Cited

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.

Bettinger, E., & Slonim, R. (2006). The effect of educational vouchers on academic and non-academic outcomes: Using experimental economic methods to study a randomized natural experiment. *Journal of Public Economics*, *90*, 1625–48.

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.

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., Guttman, B. (2018). Evaluation of the DC Opportunity Scholarship Program: Impacts two years after students applied. 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 2018-4010.

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.

Figlio, D. & Karbownik, K. (2016). *Evaluation of Ohio's EdChoice Scholarship Program: Selection, competition, and performance effects.* Washington D.C.: Thomas B. Fordham Foundation. Available at http://edex.s3-us-west-2.amazonaws.com/publication/pdfs/FORDHAM%20Ed%20Choice%20Evaluation%20Report_online%20edition.pdf.

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.

Goldhaber, D., Guin, K., Henig, J. R., Hess, F. M., & Weiss, J. A. (2005). How school choice affects students who do not choose. In Eds. J. R. Betts & T. Loveless (Eds.), *Getting choice right* (101-129). Washington, D.C.: Brookings Institution Press.

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

Greene, J. P. (2011). *Why American needs school choice*. New York, NY: Encounter Books.

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

Harris, D. N. & Larsen, M. F. (2018, July). *The effects of the New Orleans post-Katrina market-based school reforms on student achievement, high school graduation, and college outcomes* (Technical Report). New Orleans, LA: Education Research Alliance for New Orleans. Retrieved December 21, 2018 from https://educationresearchalliancenola.org/files/publications/Harris-Larsen-Effects-of-New-Orleans-Post-Katrina-Market-Based-School-Reforms.pdf

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 (2013). *Louisiana Scholarship Program: Annual Report 2012-2013*. Retrieved January 24, 2016, from http://www.louisianabelieves.com/docs/school-choice/scholarship-annual-report-%282013%29.pdf?sfvrsn=4

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

Louisiana Department of Education (2015). LEAP 2025 Vision: The Need to Transition. PowerPoint presentation retrieved from https://www.louisianabelieves.com/docs/default-source/links-for-newsletters/2015-2016-fall-assessment-guide.pdf

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. (2017a). Vouchers in the bayou: The effects of the Louisiana Scholarship Program on student achievement after two years. *Educational Evaluation and Policy Analysis*.

Mills, J. N. & Wolf, P. J. (2017b). *The effects of the Louisiana Scholarship Program on student achievement after three years* (Louisiana Scholarship Program Evaluation Report #7). New Orleans, Louisiana: Education Research Alliance for New Orleans. Retrieved from: http://educationresearchalliancenola.org/publications/the-effects-of-the-louisiana-scholarship-program-on-student-achievement-after-three-years

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 (EDRE Working Paper 2016-17). Social Science Research Network. Available at 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

Smarick, A. (2012). *The urban school system of the future: Applying the principles and lessons of chartering*. Lanham, MD: Rowman & Littlefield Education.

Sude, Y., & Wolf, P. J. (2019, April). Do you get cream with your choice? Characteristics of students who moved into or out of the Louisiana Scholarship Program (EDRE working paper no. 2019-13). Social Science Research Network.

Waddington, R. J. & Berends, M. (2018). Impact of the Indiana Choice Scholarship Program: Achievement effects for students in upper elementary and middle school. *Journal of Policy Analysis and Management, 37*(4), 1-26. Available at https://onlinelibrary.wiley.com/doi/abs/10.1002/pam.22086.

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.

Appendix A:

Defining Usage as Enrollment in 2015-16

In previous papers, we have estimated the impact of LSP scholarship usage in a specific year (Mills, 2015; Mills & Wolf, 2017a). For our Year 4 impacts analysis, we have instead opted to estimate the impact of ever using an LSP voucher to attend a Louisiana private school. This decision is primarily motivated by the declining rate of voucher usage by treatment group students between 2012-13 and 2015-16. By estimating the impact of ever using an LSP voucher instead of focusing on enrollment in a specific year, we are able to account for the experiences of initial LSP voucher users who have switched back to public schools over time. In a sense, this approach provides a more general—and potentially more policy relevant—picture of the effect of LSP scholarship usage on student achievement because it allows for the identification of impacts even for those scholarship users who find themselves dissatisfied with their private school experience.

In this section, we examine result sensitivity to this choice of enrollment specification by re-estimating all models to produce estimated effects of LSP voucher usage in the 2015-16 school year. In general, findings largely mirror those observed in the main analysis: LSP voucher usage is associated with large negative impacts on achievement in Year 4, especially in math.

Analytical Strategy

The analyses presented below are based on a variant of the 2SLS model used to generate the primary effect estimates presented in the main body of the report. Specifically, we use the following 2SLS model to estimate the effects of LSP scholarship usage on student achievement after four 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 2015-16 school year
- R_i is a fixed effect for a student's first choice school lottery[21]
- T_i indicates if a student received an LSP scholarship to their first choice school
- A_i is standardized student mathematics or English Language Arts achievement in Year 4 of the program (2015-16)[22]
- 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 key parameter of interest in this analysis is $\hat{\tau}$, which now represents the estimated impact of using an LSP voucher to attend a private school four years after initial assignment. As in our main analysis, our estimates represent local average treatment effects (LATEs). A LATE is the estimated impact of usage for students who continued to comply with their initial lottery assignment in the 2015-16 school year. As before, models account for nesting of students within first-choice private school lotteries via bootstrapped standard errors.

Estimated Impacts of LSP Voucher Usage in 2015-16

The following sections present our preliminary estimates of the LSP's impact on student achievement after four years. Our prior work reported large declines in both math and reading

[21] 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.

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

achievement after 1 year of usage (Mills, 2015). These negative effects dropped by half after two years of usage (Mills & Wolf, 2017a) and were not statistically significant after three years of usage (Mills & Wolf, 2017b). In Year 4, we see a reversal in the trending reduction in negative effects observed in years 2 and 3. Instead, across a majority of specifications and analytical samples, LSP scholarship users appear to be performing significantly worse on statewide assessments in math. We additionally observe some statistically significant negative effects of LSP scholarship usage on ELA achievement; however, these results are not consistent across analytical samples.

Table A1 presents our primary estimates of the achievement impacts of LSP voucher usage in 2015-16. We present results for two analytical samples: one requiring baseline achievement in grades 3 and 4 ("BA Sample") and another which does not require baseline achievement for sample inclusion ("NBA Sample").[23] The latter sample is effectively restricted to eligible LSP students applying for placement in grades 1 through 5 in the 2012-13 school year (Year 1 of our study).

Columns 1 and 6 display coefficient estimates for first stage regressions which use scholarship award to predict the likelihood of continuing to attend an LSP private school in 2015-16 for both analytical samples. These coefficients, which allow us to identify the percentage of individuals who are continuing to comply with their initial lottery assignment, indicate high rates of non-compliance by Year 4.[24] Specifically, less than 20 percent of

[23] Our empirical strategy is to identify causal impacts of voucher usage by leveraging oversubscription lotteries for first-choice schools. Assuming we have correctly identified lotteries for first-choice schools, the LATE estimates generated for both the BA and NBA samples will present unbiased estimates of the effects of using an LSP scholarship to attend an LSP school four years after initial assignment.

[24] Compliance is identified as continuing to observe the result of the lottery. In our context, lottery compliers is represented by treatment group students enrolled in private schools and control group students who are not enrolled in private schools. Non-compliance, in contrast, is represented by treatment group individuals who are not using an LSP scholarship to attend a private school and control group students who are enrolled in private schools.

individuals in the BA sample and 30 percent of individuals in the NBA sample continued to comply with their initial lottery assignment four years after randomization. While all estimates are statistically significant, the low compliance rates raise the concern that first-stage lottery results may be weakly instrumenting for scholarship usage in a specific year.

Our primary estimates of the impact of scholarship usage on the student achievement follow. First, we present results from a simple model that only controls for first-choice school lottery fixed effects (columns 2 and 7). Next, we include an indicator for taking the same grade-level assessment in two consecutive years (columns 3 and 8). Fully specified models, which additionally control for student demographics and baseline achievement, when indicated, appear in columns 4 and 9. These are our preferred estimates of the impact of the LSP given the small number of significant differences in baseline characteristics between treatment and control groups observed in Table 2.

While our previous research suggested the initial large negative test score impacts for LSP scholarship users are declining over time, the results presented in Table A1 indicate some large negative impacts of LSP scholarship usage reappearing in 2015-16. First, for the BA sample, we only observe statistically significant negative impacts in math, with LSP scholarship users scoring nearly 90 percent of a standard deviation lower than their control-group counterparts in Year 4. In addition, the standard errors of the estimated effects in ELA and science are large, suggesting the negative impact estimates of the LSP in those domains may be statistically insignificant mainly because they are statistically noisy. The analyses presented in columns 1 through 5 likely suffer from low statistical power, a result driven in part by the requirement of baseline achievement that effectively restricts our analysis to just two grade cohorts of LSP applicants.

Table A1.

Estimated Effects of Ever Using an LSP Voucher on Student Achievement after Four Years

	Baseline Achievement (BA) Sample					No-Baseline Achievement (NBA) Sample				
		LATE					LATE			
	First Stage	Simple Model	+ Test Retake	Fully Specified	Omitting New Orleans	First Stage	Simple Model	+ Test Retake	Fully Specified	Omitting New Orleans
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
English Language Arts	0.19***	-0.36	-0.24	-0.33	-0.27	0.28***	-0.29	-0.24	-0.53**	-0.50**
	(0.05)	(0.66)	(0.61)	(0.50)	(0.50)	(0.03)	(0.18)	(0.18)	(0.21)	(0.21)
Mathematics	0.19***	-0.90*	-0.81	-0.87*	-0.89*	0.28***	-0.74***	-0.68***	-0.94***	-0.90***
	(0.05)	(0.52)	(0.50)	(0.45)	(0.47)	(0.03)	(0.17)	(0.17)	(0.19)	(0.19)
Science	0.18***	-0.88	-0.76	-0.96	-0.65	0.28***	-0.38*	-0.32	-0.52**	-0.53**
	(0.05)	(0.87)	(0.81)	(0.64)	(0.52)	(0.03)	(0.22)	(0.21)	(0.25)	(0.21)
Controls										
Test Re-take			X	X	X			X	X	X
Baseline Achieve.			X	X	X					
Demographics				X	X				X	X
# of Sch. Choices				X	X				X	
New Orleans				X						
N	719 - 735		719 - 735	699 - 715	538 - 552	2182 - 2210		2182 - 2210	1606 - 1632	1405 - 1430
Lotteries	103 - 103		103 - 103	103 - 103	73 - 74	255 - 255		255 - 255	214 - 215	178 - 180

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

Notes. *Simple Model* refers to estimations that only control for first-choice school lottery 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 (BA sample only), student demographics, number of school preferences offered, and geography. Columns 5 and 10 omit 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.

We attempt to address this statistical power issue by dropping the baseline achievement requirement for inclusion in our analytical sample. The results presented in columns 6 through 10 of Table 3 are based on the experiences of students applying to use LSP scholarships in 2012-13 in grades one through five and are estimated using a statistical model that omits baseline test scores. Unlike in the BA sample, the estimated impact of LSP scholarship usage is negative across all subjects in fully specified models in this expanded sample. As before, the estimated effects are quite large: LSP scholarship users appear to score 53 percent of a standard deviation lower than control group students in ELA, 94 percent of a standard deviation lower in math, and 52 percent of a standard deviation lower in science. While the estimated impact for science shrinks in magnitude relative to the BA sample, the relatively larger coefficients in ELA and math in the NBA sample correspond with our prior research (Mills & Wolf, 2017b), which reports evidence of larger negative test score effects of the LSP for applicants to earlier grades.

Table A1 additionally presents models for samples of students applying to LSP scholarship schools outside New Orleans (columns 5 and 10). These models attempt to provide a cleaner look at the effects of the program's statewide expansion, as New Orleans was already home to both a diverse public charter school market (Harris & Larsen, 2018) as well as the pre-existing piloted version of the LSP. It is possible, for example, that New Orleans private schools are more equipped to work with the less advantaged population of students applying to the LSP due to their experience with the pilot program since 2008. In contrast, estimated effects also depend on the quality of the counterfactual environment (i.e., public schools), and research indicates that student outcomes in New Orleans public schools have improved noticeably in the wake of education reforms implemented during rebuilding efforts following the destruction of

Hurricane Katrina (Harris & Larsen, 2018). Both factors suggest LSP effects may differ based on whether students were inside or outside of New Orleans.

In contrast, the results presented in columns 5 and 10 are negligibly impacted by this restriction. Estimates are largely similar in magnitude and mirror the statistical significance of those in our fully specified models (columns 4 and 9). In short, the inclusion of New Orleans students in our preferred models does not appear to affect the substantive conclusions of our evaluation.

Next, we examine how the LSP effects vary over time for these samples of students. Our prior research indicates large negative impacts on ELA and mathematics in the first year of participation (Mills, 2015) that appear to diminish somewhat by Year 2 (Mills & Wolf, 2017a) and are not statistically significant by Year 3 (Mills & Wolf, 2017b). Figures 2 and 3 present LATE estimates for ELA and mathematics for Years 1 through 4 for consistent samples of students contributing to the analyses presented in Table A1.[25] Figure A1 presents results for students in the BA Sample and Figure A2 focuses on students in the NBA Sample.

Consistent with our prior work, we observe large declines in ELA and math performance in the first year of voucher usage that become less negative in Years 2 and 3. Effects are generally worse in math than in ELA. By Year 4, however, we observe a direction reversal for the scholarship usage effect estimates. Effects are more negative in magnitude in Year 4 relative to Year 3 across all tests and samples, and are statistically significant in the majority of specifications.

[25] We present these figures of the impacts of the LSP on a consistent sample of students across time for two reasons. First, the annual samples of students with outcome data change over the outcome time horizon. Focusing on a smaller but consistent sample of students means that we can rule out changing student populations as a factor affecting any changes in program effects observed across the years. Second, compliance rates change over time. Both of these factors imply that we may observe different trends in effects over time for the samples of students contributing to the Year 4 analysis. In a sense, this robustness test provides a first look at the generalizability of these findings or, in contrast, the uniqueness of the samples.

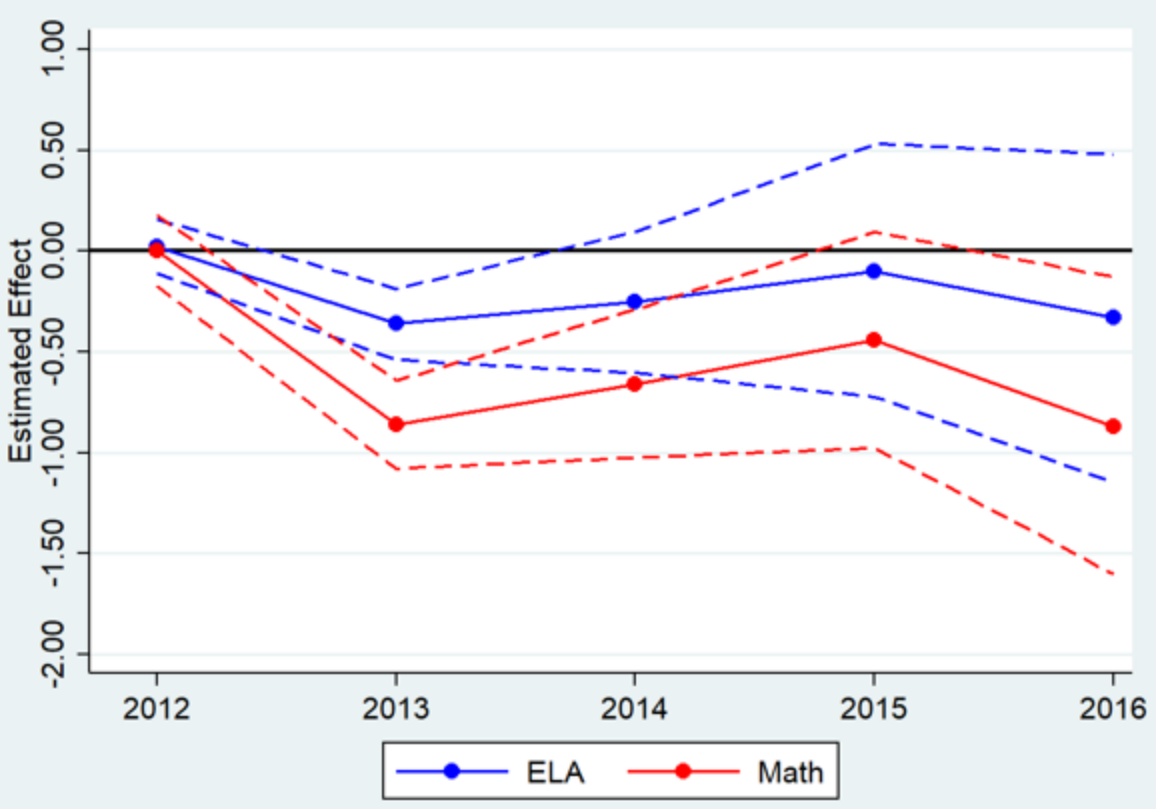

Figure A1. Estimated Local Average Treatment Effects over time for BA sample. Figure presents point estimates from fully specified models for 2011-12 (baseline) through 2015-16 for ELA and math. Results are presented for a consistent sample of students with Spring 2016 outcome data. ELA and math results are based on student achievement on the Louisiana state assessments (LAA) in 2011-12 through 2013-14, PARCC assessments in 2014-15, and LAA in 2015-16. Dashed lines represent 90% confidence intervals for the performance averages.

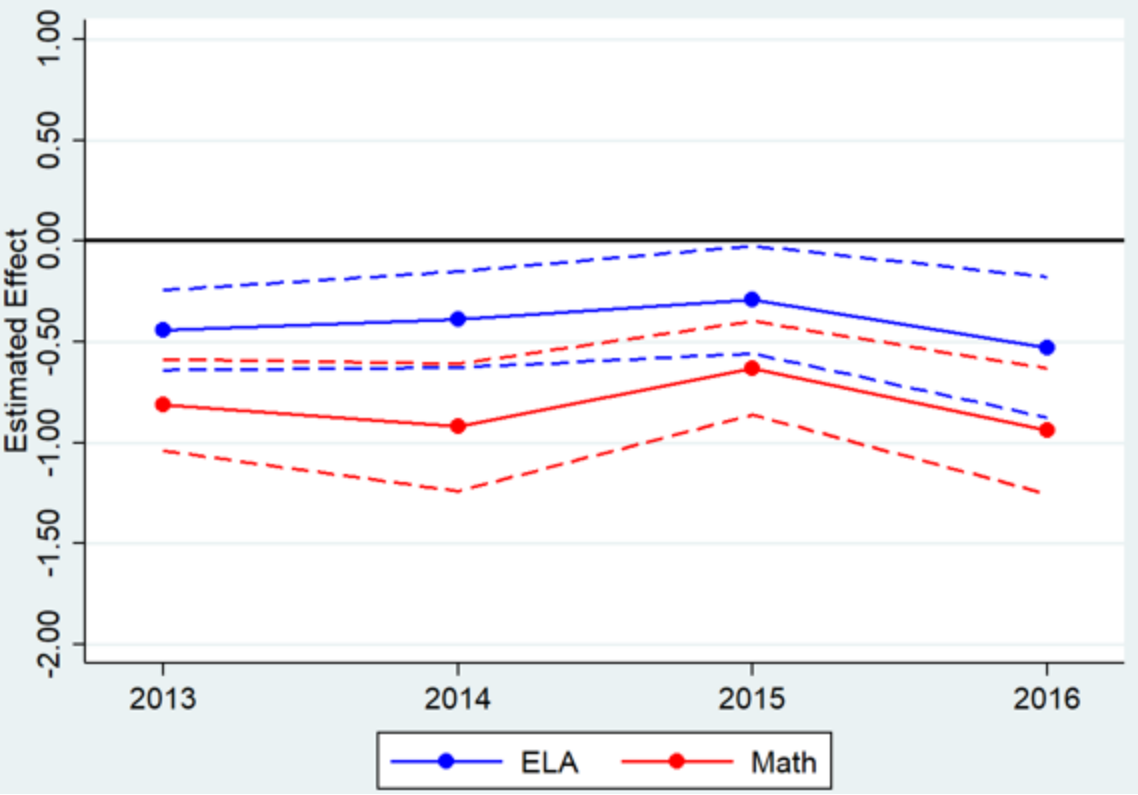

Figure A2. Estimated Local Average Treatment Effects over time for NBA sample. Figure presents point estimates from fully specified models for 2013-14 through 2015-16 for ELA and math. Results are presented for a consistent sample of students with Spring 2016 outcome data. ELA and math results are based on student achievement on the Louisiana state assessments (LAA) in 2012-13 through 2013-14, PARCC assessments in 2014-15, and LAA in 2015-16. Dashed lines represent 90% confidence intervals for the performance averages.

Interestingly, the Year 4 estimates are roughly the same size as the effect estimates for Year 2. Year 3 was considered a transitional year for the state's accountability system as the state experimented with PARCC (Mills & Wolf, 2017b). Neither public nor private schools in the LSP faced sanctions for poor student test score results in that one year within the time horizon of our study. Just as switching from low-stakes to high-stakes altered the estimates of the test-score effects of the Milwaukee school voucher program in a prior study (Witte et al., 2014), the change from high-stakes to low-stakes in Year 3 of our LSP evaluation also may render those effect estimates anomalous. Our Year 3 estimates of the test score impacts of the LSP might be more accurate than the estimates for the other years. In all other years of our study, the performance of the LSP and control group students was assessed using some variant of the LEAP exam that was both aligned to the public school curriculum and more familiar to the personnel in public schools compared to their counterparts in private schools. The PARCC, like the LEAP, was designed to be closely aligned with the mandatory public school curriculum but was equally unfamiliar to both public and private school personnel for the single year that it was used as the state accountability test. Arguably, that common level of unfamiliarity removed a home-test advantage from the control group in Year 3 that may have contributed to its higher average test scores relative to the treatment group in Years 1, 2, & 4.

Subgroup Effects

The results presented in Table A2 allow us to determine if LSP scholarship usage effects are differentiated by gender, race, and baseline achievement. Results are presented for both simple models which control only for LSP assignment lottery and fully specified models which additionally control for demographics and, in the *BA Sample*, baseline achievement. All models specify enrollment as enrollment in an LSP school four years after initial assignment.

Table A2.

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

	English Language Arts						Mathematics					
	BA Sample			NBA Sample			BA Sample			NBA Sample		
	N	Simple	Full Model	N	Simple	Full Model	N	Simple	Full Model	N	Simple	Full Model
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Gender												
Female students	385	-0.53	-0.20	1182	-0.27	-0.47**	375	-1.27*	-0.97*	1173	-0.75***	-0.83***
		(0.71)	(0.52)		(0.20)	(0.23)		(0.68)	(0.51)		(0.18)	(0.20)
Male students	349	-0.12	-0.51	1028	-0.37*	-0.61*	347	-0.59	-0.76	1029	-0.75***	-1.13***
		(0.70)	(0.62)		(0.22)	(0.34)		(0.58)	(0.55)		(0.23)	(0.35)
Difference		-0.38	0.31		0.09	0.14		-0.66	-0.21		0.00	0.30
		(0.45)	(0.51)		(0.20)	(0.36)		(0.55)	(0.49)		(0.22)	(0.38)
Race/Ethnicity												
African American students	658	-0.38	-0.21	1983	-0.25	-0.41*	648	-0.99	-0.79	1976	-0.73***	-0.86***
		(0.76)	(0.51)		(0.19)	(0.22)		(0.63)	(0.68)		(0.19)	(0.22)
Other students	76	-1.08	-1.22	227	-0.88**	-1.12**	74	-1.38	-1.71	226	-1.16***	-1.49***
		(1.12)	(1.17)		(0.40)	(0.50)		(1.51)	(3.81)		(0.3)	(0.42)
Difference		0.72	1.01		0.63	0.71		0.40	0.92		0.43	0.63
		(1.02)	(1.05)		(0.40)	(0.52)		(1.29)	(3.33)		(0.43)	(0.49)
Baseline achievement												
Lower Third	249	0.28	-0.43				245	1.43	1.63			
		(2.20)	(18.17)					(17.12)	(110.61)			
Middle Third	254	0.12	-0.13				253	-2.03	-1.94**			
		(0.93)	(0.57)					(2.61)	(0.97)			
Upper Third	242	-0.41	-0.03				235	-0.91	-0.50			
		(0.94)	(0.80)					(0.61)	(0.68)			

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

Notes. BA Sample requires students to have baseline achievement in grades 3 or 4. *NBA Sample* does not require baseline achievement. Performance measures standardized within grade based on control group score distributions. *Simple Model* refers estimations that only controls for first-choice school lottery fixed effects. *Full Model* refers to models controlling for test retaking, baseline achievement (BA sample only), student demographics, number of school preferences offered, and geography. 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.

While large in magnitude, we do not observe statistically significant evidence that LSP scholarship usage effects are moderated by either gender or race. While estimated effects are negative and statistically significant for some groups, none of the estimated difference coefficients are statistically significant. For example, while both male and female scholarship users in the NBA Sample perform poorly in ELA, we cannot determine with acceptable levels of statistical precision that treatment effect differed between these two groups.

Results are similar for baseline achievement models, with one exception. While we largely do not observe differences by baseline achievement category, students scoring in the middle third of math achievement at baseline have very large and statistically significant negative effects by Year 4. We cannot, however, make comparisons across point estimates because these estimates are drawn from separate regressions. Interestingly, students initially testing in the bottom third of the test distribution at baseline are shown, at times, to be outperforming their counterparts by Year 4. None of the estimated effects are significant, however.

In short, we do not observe strong evidence that effects were differentiated by gender, race, or baseline achievement. This pattern of findings differs somewhat from the analysis presented in the main body of this paper (Table 4), in which African American students are observed to have less negative test score impacts in math than non-African American students. We believe this difference is due in large part to the diminished relevance of first-choice lottery outcomes as instruments when compliance is specified as enrollment in an LSP school in a specific year. In the analysis featured in this Appendix, we are interested in estimating the impact of LSP voucher usage in the specific school year of 2015-16. Compliers, in this context, are treatment group students who are still enrolled in a private school four years after initial

assignment and control group compliers are those students who are still enrolled in a public school in Year 4. As noted, the percentage of students continuing to use LSP vouchers to enroll in private schools has declined dramatically since 2012-13, which has ultimately left us with relatively small estimates in our first stage regressions. While statistically significant, standard diagnostics indicate these analyses sometimes suffer from a weak instruments problem. In contrast, the estimates presented in the main body of this paper—in which usage is defined as ever using an LSP voucher to attend a private school between 2012-13 and 2015-16—do not exhibit such issues. As such, we emphasize in this report the results focused on estimating effects of using an LSP voucher at any point during the four years following initial random assignment rather than the year specific analysis presented in this appendix.

Appendix B:

Additional Tables

Table B1.

Baseline Equivalence of Treatment and Control Groups on Covariates for NBA Sample, Year 4

	N	Treatment Avg.	Control Avg.	Raw Diff.	Adjusted Diff.	*s.e.*
	(1)	(2)	(3)	(4)	(5)	(6)
Female	2,348	0.54	0.51	0.03	0.01	0.03
Race/Ethnicity						
African American	2,348	0.90	0.90	-0.01	-0.03**	0.01
Hispanic	2,348	0.02	0.02	0.00	0.01	0.01
White	2,348	0.05	0.05	0.00	0.01	0.01
Other	2,348	0.03	0.02	0.01	0.01	0.01
Limited Eng. Proficiency	n/a	n/a	n/a	n/a	n/a	n/a
Free-or Reduced-Price Lunch	1,740	0.95	0.95	0.00	0.00	0.01
# of Sch. Preferences	2,348	1.98	2.51	-0.53	-0.13**	0.06
Standardized Performance						
ELA Scale Score	n/a	n/a	n/a	n/a	n/a	n/a
Math Scale Score	n/a	n/a	n/a	n/a	n/a	n/a
Science Scale Score	n/a	n/a	n/a	n/a	n/a	n/a
Social Studies Scale Score	n/a	n/a	n/a	n/a	n/a	n/a

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

Notes. NBA sample does not require baseline achievement for sample inclusion. Analysis sample excludes students with disabilities and multiple birth siblings. The analysis sample represents LSP applicants to grades 1 through 5 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 3 through 4. *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. *Raw Diff* is the raw difference in means between the treatment and control groups. *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 B2.

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.*
BA Sample: Baseline Achievement Required					
Spring 2013	1,932	0.04	0.09	-0.05***	(0.01)
Spring 2014	1,721	0.07	0.10	-0.03*	(0.02)
Spring 2015	1,333	0.08	0.11	-0.03	(0.02)
Spring 2016	899	0.12	0.14	-0.02	(0.03)
NBA Sample: No Baseline Achievement Requirement					
Spring 2013	2,519	0.05	0.11	-0.05***	(0.01)
Spring 2014	2,819	0.10	0.12	-0.01	(0.01)
Spring 2015	3,145	0.11	0.14	-0.03**	(0.01)
Spring 2016	2,688	0.12	0.14	-0.02	(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