DISCUSSION PAPER SERIES

IZA DP No. 14342

Effects of Scaling up Private School Choice Programs on Public School Students

David N. Figlio
Cassandra M.D. Hart
Krzysztof Karbownik

APRIL 2021

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 14342

Effects of Scaling up Private School Choice Programs on Public School Students

David N. Figlio
Northwestern University and IZA

Cassandra M.D. Hart
University of California, Davis

Krzysztof Karbownik
Emory University and IZA

APRIL 2021

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Effects of Scaling up Private School Choice Programs on Public School Students*

Using a rich dataset that merges student-level school records with birth records, and leveraging a student fixed effects design, we explore how the massive scale-up of a Florida private school choice program affected public school students' outcomes. Program expansion modestly benefited students (through higher standardized test scores and lower absenteeism and suspension rates) attending public schools closer to more pre-program private school options. Effects are particularly pronounced for lower-income students, but results are positive for more affluent students as well. Local and district-wide private school competition are both independently related to student outcomes.

JEL Classification: H75, I21, I22, I28

Keywords: school choice, school competition, student achievement, behavioral outcomes

Corresponding author:
Krzysztof Karbownik
Department of Economics
Emory University
Rich Memorial Building
1602 Fishburne Drive
30307 Atlanta, Georgia
USA
E-mail: krzysztof.karbownik@emory.edu

* We are grateful to the Florida Departments of Education and Health for providing de-identified, matched data used in this analysis. Figlio acknowledges support from National Science Foundation, National Institute of Child Health and Human Development, the Bill and Melinda Gates Foundation, the Institute for Education Sciences (CALDER grant), and the Shelter Hill Foundation for assistance in building the dataset and/or conducting this research. We thank Michael Dinerstein, Stefanie Fischer and seminar participants at the CESifo Economics of Education Meeting and University of New South Wales for helpful comments and suggestions. The conclusions expressed in this paper are those of the authors and do not represent the positions of the Florida Departments of Education and Health or those of our funders. All errors are our own.

I. Introduction

Programs using public funds for children to attend private schools of their choice are on the rise in the United States: As of 2019, 25 states, plus the District of Columbia and Puerto Rico, had voucher or scholarship programs in place, many of them targeted to specific populations like students with disabilities or low-income students (EdChoice, 2019), and as of the time of writing numerous other states are considering enacting similar programs. Among the most controversial issues associated with private school choice programs involves what happens to the students remaining in public schools. On the one hand, private school choice programs could encourage public schools that might otherwise have been complacent to vigorously improve the education they offer in order to avoid losing "clients" to a more-affordable private schooling sector (Epple, Romano, & Urquiola, 2017; Urquiola, 2016). On the other hand, public school students could be harmed by private school choice programs if the programs drain resources from the public schools or if choice-induced sorting of students disadvantages those remaining in public schools (Epple, Romano, & Urquiola, 2017).

These theoretical predictions assume an established program, so it is important to know what happens to traditional public schools as school choice programs expand and mature. For example, we might expect to see the most pronounced effects when a new school choice program is initially announced, as the programs may be most salient when they are new and receiving publicity. But we might also expect the effects of school choice programs to become more pronounced as the programs grow; as schools see a growing share of students opting into choice programs, they may feel more compelled to respond.[1] Furthermore, it is plausible that local competition has a compounding effect at the district level, inducing further gains or losses to public student's outcomes. Other important reasons why the scale-up of a voucher program may have different effects than the introduction of a program is that longer-run scale-up implies much more of a sense of program permanence. It is not uncommon for policies and programs to change dramatically over time, or to be abolished; for instance, Montana's tax credit scholarship program was in legal limbo for years due to court challenges (Totenberg & Naylor, 2020). Some

[1] Of course, the direction or strength of effects could also depend on other factors, like the quality of nearby private schools. Studies have reached mixed conclusions on the causal effects of attending private school on voucher users' outcomes, with some studies finding benefits to students in terms of either test scores or longer-term attainment outcomes and others finding null or negative effects on test scores (see Epple, Romano, & Urquiola, 2017 for a review of this literature). In this study, we lack data on private school quality to test whether competitive effects differ based on their quality.

programs grow in magnitude over time, while others remain relatively small pilot programs, and still others fluctuate wildly in size due to changing legislative appropriations. Only by studying a program's growth and development over a long stretch of time can we begin to fully understand how voucher programs alter the public school landscape.

Nonetheless, nearly every paper written in the U.S. context – including those written to date by this paper's authors – investigates the introduction of a school voucher program, rather than studying how the programs affect public schools as they approach some longer-run steady-state. The weight of the U.S. evidence shows small but positive effects of the introduction of private school voucher programs on public school students' test scores (Chakrabarti, 2008; Egalite, 2016; Figlio & Hart, 2014; Figlio & Karbownik, 2016; Greene & Winters, 2007; Hoxby, 2003; Rouse et al., 2013; see Urquiola, 2016; Epple, Romano, & Urquiola, 2017; and Egalite & Wolf, 2016 for recent overviews of this literature). These studies generally focus on the very immediate short-run effects, evaluating the first one to four years after the initial introduction of school voucher programs, when both the pros and cons of the choice program may be constrained due to the small number of initial participants. They do not, however, consider what happens when the private school choice program scales up and the number of students using private school vouchers grows to encompass a sizeable fraction of the overall student body. To date, with the exception of a pair of informative but single-market school-level analyses from Milwaukee (Hoxby, 2003; Chakrabarti, 2008), we do not know much about whether scaling up private school choice programs helps or harms public schools. Our paper complements this prior work and examines scale-up of a major statewide voucher program – the largest of its kind – over a timeframe of over 15 years in the third largest state in the U.S. The statewide nature of the voucher program is also helpful because there are many distinct competitive markets, permitting us to employ a quasi-experimental shift-share research design (Bartik, 1991) to identify the effects of voucher competition on public school performance.

One paper provides some evidence on the impacts of scale-up of public school choice (charter) programs, though even that highly-informative paper still focuses on a relatively small scale-up, much smaller than the degree that we consider here, and an arguably short time frame. Gilraine, Petronijevic, & Singleton (2021) show that North Carolina students who experience increased exposure to charter competition due to new school approvals resulting from the removal of a charter cap saw improvements in math (but null effects in reading). In their setting, competition increases in a relatively narrow window of two years, resulting in roughly a 25% increase in the potential degree of charter school competition. Our longer time scale allows us to consider a program that has grown almost seven-fold from its original size and currently serves a participant population that is on average nearly 4% the size of the K-12 student population in Florida while in some districts the participation rate is over 10%.[2] Moreover, our paper considers a different sector providing competition (private vs. public charter), extends the competition measures beyond local level, explores both cognitive and behavioral outcomes, provides extensive heterogeneity analyses, and is set in a different state.[3]

In this paper, we leverage extraordinary child-level data that matches birth records to school records and employ student fixed effects to evaluate a statewide school voucher program, the Florida Tax Credit Scholarship Program, that grew over the course of about a decade from much less than one percent to roughly four percent of the state's student body participating. We exploit differences in the initial competitive landscape faced by different schools – using five separate measures of voucher competition introduced by Figlio & Hart (2014) – as well as differences over time in the expansion of the voucher program, to determine whether students attending public schools that face increased exposure to private school choice as a result of a statewide voucher program experience educational (test scores) and behavioral (absenteeism and suspensions) benefits or losses. These two sources of variation provide the shift – the state-level expansion in the availability of vouchers – and the share – pre-policy fractions of private schools that we consider the most likely competitors to public schools – in the shift-share quasi-experimental research design (Bartik, 1991). We find evidence – described below – that our quasi-experimental setting satisfies the exogeneity-of-shares identifying assumptions recently detailed by Goldsmith-Pinkham, Sorkin, & Swift (2020).

We find evidence that as public schools are more exposed to private school choice, their students experience increasing benefits as the program scales up. In particular, higher levels of private school choice exposure are associated with lower rates of suspensions and absences, and

[2] Specifically, 108,098 students participated in the program as of 2017-18, compared to a K-12 public school student population of nearly 2.78 million students (Florida Department of Education, 2018; Florida Department of Education, n.d.a)

[3] Two other recent papers consider scale up of charter programs (Ridley and Terrier, 2018; Cohodes et al., 2019), however, these papers focus on even smaller expansions in Massachusetts and only the former is concerned with effects on students remaining in traditional public schools. They find small positive effects on test scores but do not explore behavioral outcomes.

with higher standardized test scores in reading and in math. These results are not uniform: We carry out an extensive heterogeneity analysis facilitated by matched birth and school records from Florida, and find that the public school students most positively affected by increased exposure to private school choice are comparatively low-SES students (those with lower family incomes and lower maternal education levels). Nonetheless, we also observe statistically significant but smaller gains for higher-SES students who are unlikely themselves to be targeted by the means-tested vouchers. Furthermore, competitive landscapes faced by individual schools and the district as a whole are both independently important, with the latter having larger effect sizes on student outcomes. In terms of potential mechanisms, we are able to rule out alternative explanations related to changing composition of students remaining in the public schools, changes in district-level competition from public school choice options such as charters or magnets, or effects on the resources that public schools have. Thus, in our view, the increase in competitive pressure resulting from increased voucher utilization is the most plausible channel for the estimated gains in test scores and behavior.

II. Florida Context

We focus on the competitive effects of the Florida Tax Credit (FTC) Scholarship Program as it scaled up over a 15-year period. Announced in spring 2001, the FTC program provides dollar-for-dollar tax credits to corporations that donate to non-profit Scholarship Funding Organizations (SFOs); the SFOs then use these contributions to offer scholarships to low-income students for use at private schools (Florida Department of Education, 2018). Until recently, and during the study period covered by this paper, in order to receive a scholarship the first time, students must either have spent the previous full year in a Florida public school or be entering kindergarten or first grade. In 2002-03, the first year of operation, the program spent $50 million to fund scholarships for 15,585 students, with a maximum value of $3,500 for each scholarship. Scholarships need not cover the full amount of private school tuition; families are allowed to supplement the scholarship as necessary to meet tuition bills. Initially, eligibility was restricted to students with a family income no greater than 185 percent of the federal poverty line (see Florida Statute 220.187 (2001) and subsequent amendments), or $47,637.50 for a family of four in 2019 dollars.

The program has expanded along several dimensions since 2002-03, its first year of operation. Table 1 charts the expansion of the program in terms of the designated funds for the program (Column 1), realized spending (Column 2), the number of students enrolled in FTC (Column 3), the ratio of FTC participants to the total K-12 public school population (Column 4), the number of participating private schools (Column 5), the maximum income level eligible for participation (Column 6), and the maximum scholarship level available (Column 7). By 2017-18, the program cost roughly $640 million and awarded scholarships to 108,098 students (Florida Department of Education, 2018), or about 3.6 percent of all K-12 students in Florida. Since private high schools tend to be more costly and thus attract fewer students, participation as a percentage of K-8 public school enrollment – of primary interest for our analysis in this paper – is even higher at 5.0% and 3.5% in elementary and middle school grades, respectively.

The growth in participation may flow from a number of factors. First, the number of private schools participating in the FTC program nearly doubled from 924 schools participating in 2003-04 to 1818 schools participating in 2017-18. The growth in availability of schools means that some areas may have effectively lacked a convenient voucher-accepting school in the early years, but had options available in later years. Increases in the number of participants may also reflect a loosening of the income-based restrictions. In 2006 school year, the program introduced a rule allowing students using the scholarships to continue in the program if their family income did not exceed 200% of the federal poverty line. In 2010, this continuing-student eligibility threshold was raised to 230% of the federal poverty line. In 2016, the eligibility restrictions were changed to allow partial scholarships for entering students with incomes between 185%-260% of the federal poverty line. These rule changes expanded the pool of students eligible to participate.

At the same time, the program may have become more attractive to students because of increases in the maximum scholarship available. The maximum scholarship grew from $3500 in 2002-03 to $7,208 in 2017-18. While state-level data on private school tuition is scarce, based on our calculations using national statistics from NCES, the $7,208 scholarship would cover 100% (67%) of the average tuition of a Catholic elementary (high) school and about 83% (40%) of the average tuition of a non-Catholic, religious elementary (high) school.[4]

[4] Specifically, figures from the 2011-2012 Schools and Staffing Survey suggest that the national average tuition at Catholic, other religious, and non-sectarian elementary schools were $5,330, $8,676, and $18,170, respectively (National Center for Education Statistics, 2013). The corresponding figures for high schools for each school type were $9,790, $16,520, and $25,180. Translated into 2017 dollars at a 9% inflation rate, this suggests that a $7,208 scholarship for an elementary (high school) student would have covered 124% (67.5%), or 83.1% (40.0%), or 36.4% (26.3%) of the average tuition at Catholic, other religious, or non-sectarian schools respectively. Based on data from Florida Active Public School Directory over 60% of FTC-participating private schools are religious (16% are Catholic, with 44% representing other religions).

Given the massive, in an educational context, scale up over time, the participation rates vary both across years and across geographic areas. To give a better sense of the spread of participation across space and over time, Figure A1 shows the district-level ratio of FTC participation to K-12 public school enrollment over four school years: 2005-06, 2009-10, 2013-14, and 2017-18. Here, darker grey colors represent a heavier concentration of FTC participation. Two patterns stand out. First, the map darkens considerably over time, reflecting an overall growth in participation rates between 2006 and 2018. No district has more than 3% participation in 2005, while roughly 45% of districts (representing nearly 60% of students) have participation rates in excess of 3% in 2017. Second, there is substantial spatial variation in FTC participation, particularly in later years. While eight of Florida's 67 districts have less than 1% participation in 2017, another eight have participation rates of over 7% while a district with highest participation rate reaches almost 13%. While we lack district-by-grade level data on participation, the state-level statistics on participation by grade suggest that the rates are likely higher for elementary grade students in these districts, a population on which we focus on in our empirical analyses. There is not a clear pattern in the relationship between district size and FTC participation rates. For instance, while the districts with the highest rates of participation (7% or more) in 2017 include highly populous districts like Miami-Dade County, they also include sparsely populated districts like Jefferson County. This suggests meaningful differences across the state in the extent to which public schools should perceive competition from private schools.

II. Methods

A. Data and Sample

We draw upon data provided by the Florida Department of Education and Florida Department of Health. The Department of Education collects data on all students, including test scores, absences, and suspension data for students in grades PK-12. The Florida agencies merged these data to birth records for children born in Florida between 1992 and 2002, providing measures of families' socioeconomic status at birth, as well as neonatal outcomes such as birth weight. Because we also received the same data on the set of children born in Florida but never attending Florida public schools, we can characterize selection into our sample. We measure a public school's competitive landscape of nearby private schools based on files maintained by the Florida Department of Education, which provide locational data (latitude and longitude, as well as addresses) for public and private schools as well as the grades that each school serves (Figlio & Hart, 2014). One of our competition measures also relies on a list of houses of worship (churches, mosques, and temples) in Florida, including location data; these were obtained from a database maintained by ReferenceUSA.[5]

Our sample is limited in two key ways. First, we focus on outcomes for students in grades 3-8, because test scores serve as one of our main outcomes and they are most consistently available for this set of grades. We also require students to be present in grade one so that we can assign our competitive pressures measures as explained in Section II C. While our empirical sample includes students observed in any grade between 3-8, in robustness tests we confirm that our results are similar if we limit it to students present in all grades. The second is that due to data availability and in order to have complete coverage of the rich set of measures provided by the birth records data, we restrict our main sample only to those students with Florida birth certificates (i.e., those students born in Florida). Roughly 81 percent of children represented in Florida birth records are ultimately observed in the Florida public school data, a match rate that tracks closely with the share of Florida-born students who appear in Florida public schools according to the American Community Surveys (Figlio, Guryan, Karbownik, & Roth, 2014).[6] Records of children who started in a public Florida kindergarten but left the public school system prior to the start of testing in the third grade or had missing test score information in all years accounted for 14.8 percent and 0.8 percent of the remaining matched sample, respectively.[7] Additionally, 0.8 percent of the matched sample was excluded from testing because of severe disabilities. This suggests that our data provide good coverage of the overall universe of students affected by the competitive pressures from the school voucher program. Since this attrition is not systematically correlated with our treatment of interest, the estimates should not be biased.

---

[5] Note that these data reflect the landscape in 2010, due to data availability limitations. However, we argue that it is unlikely that the voucher law prompted the establishment of any new houses of worship.

[6] It is noteworthy that the voucher program's scale-up affected somewhat who shows up in public schools to begin with. As we discuss in Online Appendix A, in areas with greater competition, we see a diminishing share of students coming from lower-income families enrolling in public schools, consistent with the means-testing criteria for program eligibility. To the extent that student fixed effects account for these time-invariant characteristics, and there are no time-varying covariates differentially correlated with scale up in more vs. less competitive areas, our estimates should not be biased.

[7] Leaving the public school system between kindergarten and the commencement of testing in grade 3 is not consistently correlated – in terms of sign and statistical significance – with competitive pressures faced at entry into the school system, suggesting that attrition from the sample is not endogenous to scale up.

Overall, our main analytic sample includes student data for roughly 1.2 million unique students in the 2002-03 to 2016-17 academic years, although we use several additional prior years of data to characterize the initial schools for students in earlier cohorts as well. When we refer to academic years in data for the remainder of the paper, we will refer to spring of the academic year when the testing takes place.

Since our matched data are limited to students born between 1992 and 2002, they do not include test score information for years prior to the program's initiation. Thus, in order to provide evidence that pre-program competitive landscape was not correlated with trends in student outcomes prior to the establishment of the FTC program, we supplement our main analysis with data from earlier years. This data set includes information on all public school students in Florida (and not only those who were also Florida-born as in our main analysis) who were tested between 1998/99 and 2006/07 school years. This analysis may further alleviate concerns regarding our preferred sample that is limited to matched birth-school records. It further allows us to execute an event study design that illuminates the lack of pre-policy trends. On the other hand, this supplementary dataset has three major limitations: (1) since students are not tested in all grades, we cannot apply our individual fixed effects identification strategy and thus we have to rely on school-level fixed effects analysis, which provides for weaker internal validity;[8] (2) this dataset has only limited demographic information on students since we lack birth record data; and (3) we do not have data on disciplinary outcomes pre-policy. For these reasons we use it only as a supplemental data source supporting our identifying assumption, and we come back to these issues when we discuss our event-study analysis in Section III.

B. Models

We estimate the effects of expansions of school choice program by estimating within-student models of the following form:

$$Y_{isglt} = \beta Expansion_t \times Competition_{sl} + \theta_{il} + \delta_{gt} + \varepsilon_{isglt}$$

[8] This yields another limitation of the data, tied to the way we construct our competition measures. As described below, we assign the competition measures to students based on the school they attend in first grade. However, the choice of first grade school may be endogenous for cohorts who enter first grade after the program is announced. While this concern is minimized when we use student fixed effects, it is a substantial threat to validity in analyses relying on school fixed effects. Therefore, our analyses using school fixed effects is restricted to cohorts entering grade 1 prior to the program announcement, only including years through 2007. It limits the utility of this analysis for our primary research question, since most of the expansion happens in subsequent years (Table 1).

where Y_{isglt} captures an outcome measure for student *i* who entered the FLDOE data in grade one (G1) school *s*, observed in grade *g* corresponding to academic stage (elementary or middle school) *l* in year *t*, θ_{il} is a student-by-stage fixed effect that allows separate within-student effects for the elementary school (grades 3-5) and middle school (grades 6-8) stages, but constrains school effects to be determined by schools that we anticipate students attending given their first grade placement.[9] Note that the inclusion of the student-by-stage fixed effects also hold constant time-invariant factors affecting students throughout their careers at a given academic stage, such as prior parental investments into children's human capital.[10] Implicitly, this serves a similar function to controlling for lagged test scores as a means of capturing prior endowments and investments. The term δ_{gt} is a grade-by-year fixed effect and we include robust standard errors (ε_{isglt}) clustered by student's G1 school.

The coefficient of interest is β, which estimates the interaction between a measure, $Expansion_t$, that captures the degree of program use statewide in year *t*, and a measure, $Competition_{sl}$, that captures whether each student's school is expected to face an above-median or below-median degree of competitive pressure, based on the pre-program competitive landscape. Importantly, for reasons we describe below in the Competition Measures section, the competitive pressure measures that we expect each student's school to face are projected based on the school that each child *initially attends* in first grade, rather than the actual school attended in any given grade, and we project these different measures based on whether the child is in elementary or middle school. This helps us avoid identifying off of changes in competitive pressure generated by endogenous moves by students during schooling. The competitive pressure measures are further based on pre-program competitive landscapes, as explored in Figlio & Hart (2014), rather than the actual competitive landscapes after the voucher program is

[9] We focus on first grade rather than kindergarten because first grade is the first mandated grade of attendance in Florida. In practice, there is extremely high correspondence between kindergarten school attendance and first grade school attendance observed in the Florida data. As explained below we anchor each student to their grade one (G1) school and then rely on empirically observed flows of students between elementary and middle schools. Therefore, a school effect itself is not identified in this equation given the student-by-stage fixed effects.

[10] Since our model includes individual fixed effects, it is problematic to further include lagged test scores in this estimation, which could presumably account for dynamic responses of cognitive skills to competitive pressures, because coefficient on β will be inconsistently estimated (Nickell, 1981). Nonetheless, since our sample sizes are very large, we have also estimated models with once-lagged test scores as control variable. This analysis produces, if anything, more positive estimates in the range of 0.2 to 1.0 percent of a standard deviation as compared to our baseline results of 0.3 to 0.7 percent of a standard deviation per 10 percent increase in the program size. Thus, we conclude that our results are robust to this specification check, and if anything, on the conservative side.

introduced, because the latter may be endogenous to public school quality. Given this design, the student-by-stage fixed effect implicitly holds the initial level of pre-program competition constant within each student-by-stage cell. This means that the effect for the interaction term is identified off of program expansion rather than by any movement of the students between schools, or off the introduction of new private schools in response to the incentives introduced by the voucher program. Thus, the coefficient of interest describes whether expansion matters more for schools with higher initial degrees of competitive pressure than for schools with relatively little initial competitive pressure. We multiply our estimates by 100 to ease interpretation of effect sizes in our figures and tables.

We rely on exogeneity of shares in the Bartik (1991) quasi-experimental shift-share research design for identification, a la Goldsmith-Pinkham, Sorkin, & Swift (2020). As we study multiple time periods but a single sector, we are operating in a panel setting, and identification relies on assumptions regarding relevance and exogeneity of shares.[11] We demonstrate below that that voucher program growth increases competitive pressure, implying that the relevance identifying assumption likely holds. Furthermore, as recommended by Goldsmith-Pinkham, Sorkin, & Swift (2020) in an event-study setting, we also show that the parallel trends assumption holds. This bolsters our confidence that the shares are exogenous to changes in outcomes and that the identifying assumptions are likely to be satisfied.[12]

C. Measures

Outcomes. Our main cognitive outcomes rely on standardized measures of math and reading scores for 3rd-8th grade students on Florida's state tests. We standardize each test within year and grade using our empirical sample of Florida-born students to maintain consistency across years, but the results are robust to using measures available for a subset of our sample years that are standardized on the whole-state population.[13] We use school years 2002-03 to

[11] To be precise Goldsmith-Pinkham, Sorkin, & Swift (2020) consider K industries over T time periods in L locations. In this paper we consider a special case of this where we have one industry (private schools), fifteen periods (school years), and either approximately 1700 (schools) or 67 (school districts) locations.
[12] An alternative approach is proposed by Borusyak, Hull, & Jaravel (2020) with the focus on the exogeneity of shifts but it requires many uncorrelated shocks which we clearly do not have as we are limited to a single policy change. Nonetheless, this is not concerning since exogeneity of shares is likely to hold in our application as detailed below.
[13] This is important due to several changes in the structure of the tests over the period covered by our analysis. The Florida Comprehensive Assessment Test (FCAT) was replaced by an updated version (FCAT 2.0) in 2010-11, and then by the Florida Standards Assessments (FSA) in 2014-15 (Florida Department of Education, n.d.b).

2013-14 and 2002-03 to 2016-17 for math and reading, respectively. This discordance comes from the fact that, in math, more advanced students were able to exercise more choice about which assessments to take starting in school year 2014-15; for instance, students taking Algebra I in eighth grade could take an algebra-specific examination rather than an examination on 8th grade math generally. We therefore exclude the years with less consistency in tests from our analysis. We also construct the measure of averaged math and reading test scores for each student for school years 2002-03 to 2013-14.[14]

Another major contribution of our paper is to explore effects of competitive pressure on a novel set of behavioral outcomes: likelihood of being suspended and absence rates. Our suspension measure is an indicator variable for whether a student has ever been suspended in a given school year, while our absence measure captures the share of days that a student is reported absent net of days they are suspended. Thus, the former can be thought of as an indicator for more serious disciplinary problems while the latter is a measure of truancy. We observe suspension and absenteeism outcomes through the 2011-12 school year.

Competition. Following Figlio & Hart (2014), we use five measures to capture the degree of competitive pressure that each school is likely to face. The "Density" measure captures the number of private schools serving the same grade range of students (i.e., elementary or middle school grades) within a five-mile radius of each public school. The "Distance" measure captures the distance between each public school and the nearest private competitor serving the same grade range; this measure is multiplied by -1 so that a positive sign on the measure will indicate greater competitive pressure. The "Diversity" measure captures the number of different religious denominational categories represented among the private schools within a five-mile radius of each public school; we group each school into one of ten denominational categories (including non-religious) for this measure.[15] The "Slots" measure captures the number of private school students served in the same grade range within a five-mile radius, standardized by the number of grades served. The "Houses of Worship" measure captures the number of houses of worship in a five-mile radius. This measure captures the underlying religiosity of the

[14] In the main regression for years 2002-03 to 2013-14, 98.8 percent of observations have both math and reading scores, 0.3 percent have only math, and 0.9 percent have only reading. By comparison in years 2014-15 to 2016-17, 74.7 percent have both scores, 1.7 percent have only math, and 23.6 percent have only reading.

[15] The ten categories are Non-religious, Catholic, Protestant, Baptist, Evangelical, Non-denominational, Jewish, Islamic, Christian Other, and Religious Other.

community, which may be associated with demand for private religious education, as well as the possibility that private schools may co-locate in the buildings that serve as houses of worship (Figlio & Hart, 2014). This measure is related to others commonly used to capture demand for religious education in the literature on private school competition, such as the share of a population that is Catholic (Hoxby, 1994; Dee, 1998; Jepsen, 2002) or the density of Catholic churches in a locality (Jepsen, 2002), but captures religiosity across a greater number of faith traditions.

Because presenting all five measures is unwieldy for the purposes of robustness checks and heterogeneity analyses, we also construct a single composite "Competitive Pressure Index" measure based on a principal components analysis of the five aforementioned measures. The principal component analysis produced a single component with an eigenvalue greater than 1; the loadings for this component were used to generate the Competitive Pressure Index score for each school. The component loadings generated by the principal components analysis for the first two components are documented in Appendix Table A1.

We make two important decisions in assigning competitive pressure measures to schools. The first addresses the concern that the competitive pressure faced by a school in any given year during the program's scale up may be endogenous to perceived school quality; for instance, private schools may be tempted to enter into markets where the public alternatives are of poorer quality (Arsen & Ni, 2008). In that case, competitive pressure would be conflated with other unobserved factors plausibly correlated with student outcomes. To avoid this problem, we measure the competitive pressure that each elementary and middle school in Florida faced using the competitive landscape in place in 2000, the last year before the voucher program was announced. Because these measures reflect the competitive landscape prior to the announcement of the scholarship program, the level of competition captured in these measures cannot logically be a result of strategic responses to the program. In supplemental analyses (presented in Appendix Figure A2), we demonstrate that this empirical design meets the instrument relevance assumption described by Goldsmith-Pinkham, Sorkin, & Swift (2020): We confirm that our year-2000 measures are strongly correlated with more current (but potentially endogenous) measures of competition, namely the number of private schools accepting FTC scholarships within a five mile radius of the public school (Panel A) and the distance to the nearest school accepting an FTC scholarship (Panel B).[16]

The second decision addresses the concern that students may move between public schools based on their perception of school quality. While endogenous public school selection would have to be correlated with differential expansion in order to be a threat to identification in our quasi-experimental shift-share research design, we take another step in order to eliminate the possibility that potentially endogenous school switches influence our estimates of the competitive effects of school voucher program expansion. Specifically, we calculate students' *predicted* elementary- and middle-school competition levels based on the school that they attend in first grade. This treats students as if they remain in the same elementary school they entered in first grade, and thus abstracts from any potentially endogenous moves. For students' elementary-school stage, the competitive pressure measures therefore capture the pre-policy competitive landscape of students' grade 1 (G1) schools. For the middle-school stage, we create a weighted average of the competitive landscapes that we would anticipate students to face based on 1) the flow of students empirically observed in our data from each elementary school (grade 1) to each middle school (grade 6) in Florida, and 2) the pre-program competitive landscape of Florida middle schools. Specifically, for each student attending a given G1 school, we observe the middle school that they actually attend, and we capture the pre-policy degree of competition faced by that middle school. We then weight these measures with empirical flows between elementary and middle schools to obtain the expected middle-school stage competition for each student based on the G1 school they attend. Importantly, in our estimating equation the fixed effect θ contains the interaction of an individual fixed effect with an indicator for whether the child is in a middle-school grade, so that the competitive pressure that we expect children to face as they progress from elementary to middle school is allowed to vary with expansion.

Because our main interest is in whether effects from program expansion are more marked in schools that face greater competitive pressure, we dichotomize each competition measure to ease interpretation of the interaction terms. Thus, for each variable, the competitive pressure

[16] Specifically, we use the Florida Department of Education's Active Private School Directory (representing the available private schools until the 2019-2020 school year). The directory indicates whether each private school accepts the FTC scholarship, so that for each public school, we are able to determine how many grade-overlapping private schools accept the FTC scholarship in a five mile radius. We are unable to calculate our Diversity, Slots and Houses of Worship measures for more recent years.

indicator captures whether the student's projected school is above or below the median on each competitive pressure measure. In the main analysis, the median split point is calculated with schools (rather than students) as the level of analysis, and is calculated separately for grades 1 to 5 and 6 to 8. In robustness tests and some extensions, we present results with the median split point calculated at the student level.

Program expansion. Our main measure of program expansion captures the logged number of students participating in the FTC Scholarship program in a given year. The interaction of the logged expansion measure and the median-split measure of competitive pressure can therefore be interpreted as the relative effect of a one percent expansion in the number of students served in schools initially facing an above-median degree of competitive pressure, compared to the effects of expansion in locations with lower competitive pressure.

Student characteristics. We have a variety of student characteristics from birth records. In particular, we capture student sex, mother's race, mother's ethnicity, whether the child's mother was born in the US, mother's marital status at the time of birth, mother's years of education at the time of birth, or whether the birth was paid for by Medicaid. These characteristics are time-invariant and are therefore captured by student fixed effects in our main estimating equation; however, we use some of them to provide extensive heterogeneity analysis to further our understanding of mechanisms at play.

In school records we also observe information on students' free and reduced price lunch status (FRPL), which we use as another stratifying characteristic in our heterogeneity tests. This measure varies within student across years but we focus our analysis on two groups of students: those that were never on free and reduced price lunch through their public school career; and those that were ever designated eligible for free or reduced price lunch.

Table A2 provides descriptive statistics for the full population of Florida births (column 1) and our main empirical sample (column 2). The comparison between these two columns makes it clear that the set of children remaining in Florida to attend public school is negatively selected in terms of maternal education compared with all children born in Florida. We are more likely to observe children whose mothers are high school dropouts (24.9 percent vs. 20.9 percent), and less likely to observe children whose mothers are college graduates (14.7 percent vs. 20.2 percent). We are also more likely to observe Black children (23.3 percent vs. 19.4

percent). At the same time, ethnicity, immigrant origin and maternal age at birth are comparable in these two samples.

In subsequent columns of Table A2 we investigate whether characteristics of students differ based on the degree of the pre-policy competitive pressure faced. Columns 3 and 4 provide statistics for children whose schools are above- or below-median in competitive pressure based on our Competitive Pressure Index. It appears that locations facing less competitive pressure prior to program's introduction have more white students (68.0 percent versus 37.3 percent) and have markers suggestive of higher socioeconomic status (66.6 percent ever on free or reduced price lunch vs. 75.8 percent). Interestingly, however, composition of parental education is relatively similar across these locations.

Panel B of this table presents mean values of our five measures of competitive pressure and the combined Competitive Pressure Index based on principal components analysis while panel C shows the five outcome variables. Descriptively, it appears that children attending schools with above median competitive pressure have, on average, poorer outcomes than children attending schools with below median competitive pressure. For example, math and reading test scores are 9.1 and 10.8 percent of a standard deviation lower in the former as compared to the latter sample, respectively. These patterns may be because lower-SES families (who tend to have lower test scores on average) are more likely than higher-SES households to live in more densely-populated urban areas (which tend to have higher degrees of competition). Regardless of the explanation, these cross-sectional differences underscore the importance of our empirical strategy that identifies competitive pressure effects based on the roll-out of the voucher program, controlling for student fixed effects.

III. Results

A. Main Results

We find consistent evidence that as the voucher program scales up, students in areas with more pre-program competitive pressure see a significantly greater improvement in outcomes than do students in areas with lighter pre-program competitive pressure (Table 2). While the magnitude of the coefficients varies across competition pressure measures, this pattern is consistent across all five underlying measures of competitive pressure (Panels A-E) for four of the five outcomes that we measure: averaged math and reading (Column 1), reading (Column 3), suspension rates (Column 4), and absence rates (Column 5). When we combine all the measures into our single Competitive Pressure Index measure (Panel F), we likewise find statistically significant increases in test scores and reductions in behavioral problems. Recall that the outcomes are multiplied by 100, and thus, our results for the Diversity measure in Panel A suggest that a 10 percent increase in the number of students participating in the voucher program is associated with a 0.4 percent of a standard deviation greater improvement in combined math/reading scores for students in schools with above-median density of private competitors, compared to students schools facing lower degrees of competitive pressure. This effect is larger for reading (about 0.7 percent of a standard deviation) as compared to math (about 0.2 percent of standard deviation). We also find reductions in both suspensions and absences, indicating that behavioral outcomes in the schools facing higher competition likewise improve. The former declines by 0.4 percent and the latter by 0.5 percent compared to their means in specifications using the Diversity measure. While the Density, Diversity, Houses of Worship, and Slots measures suggest that expanded competitive pressure is significantly and positively related to math achievement, results are non-significant when the Distance measure is used (Column 2).

The Competitive Pressure Index estimates presented in panel F – the estimates using our preferred measure – imply that a 10 percent increase in the number of students participating in the voucher program, in schools with above-median as compared to those with below-median baseline competitive pressure, increases math and reading test scores by 0.3 and 0.7 percent of a standard deviation, respectively. At the same time, suspensions decline by 0.9 percent while absence rates decline by 0.6 percent, relative to their base rates. The fact that both suspensions and absences decline in response to scale up also suggests that our test score effects are not driven by public schools using more rigid disciplinary policies to achieve academic gains. Since our identification forces us to compare scale up effects in more- vs. less-competitive locations, these estimates should be treated as a lower bound on the total effect of voucher program scale up. Our preferred measure combines all the competitive inputs into a single index, thus for transparency and clarity we present subsequent heterogeneity and robustness analyses using the Competitive Pressure Index measure, but we point readers to the working paper version of this paper (Figlio, Hart, & Karbownik, 2020) which shows these additional results for all other measures available.

Our results are similar in pattern if we simply look at the changes in effects of the initial level of competitive pressure over time (as the program was expanding) rather than as a function of a specific measure of program usage. In other words, the key terms here are interactions of initial pre-program competitive pressure level-by-year rather than interactions of initial pre-program competitive pressure level-by-logged number of participants. Panel A of Figure 1 shows the year-by-year marginal effects for being located in a market with more baseline competitive pressure for each outcome using our preferred measure of competition pressure– the Competitive Pressure Index. In this analysis, we standardize all five outcome variables to have mean zero and standard deviation of 100 in their respective empirical samples, thus making the direct comparison of magnitudes across multiple outcomes feasible.

The graph suggests that schools located in markets with more competitive pressure saw a roughly 14 percent of a standard deviation greater increase in combined math and reading scores by 2014 relative to schools in markets with less-competitive pressure. Given that by 2014, the program had expanded by nearly 300 percent compared to its original size, this coefficient is strongly consistent with our main table.[17] With respect to behavioral outcomes, improvements emerge later in the time period. Students attending schools in markets with more competitive pressure saw consistently greater reductions in suspensions (absences) relative to peers in schools facing less competitive pressure starting in 2006 (2009). We present corresponding graphs for each competitive pressure measure and each outcome (unstandardized) separately in the Appendix Figure A3.

B. Benchmarking Effect Sizes

One benchmark to contextualize the size of our causal estimates is the extent to which expansion of the voucher program is associated with closing the gap between schools located in high-competitive pressure and low-competitive pressure areas. Our descriptive statistics in Table A2 suggest that students in schools facing higher competitive pressures tend to have poorer outcomes across all measures except for absences, which are similar in both groups. Those poorer outcomes include lower math scores (gap of 9.1 percent of a standard deviation), reading

[17] More specifically, given our estimate of 5.111 (on a logged competitive pressure measure where the combined math and reading test score dependent variable was multiplied by 100 to show significant digits), the 284 percent increase in the size of the program would be associated with an 14.52 percent of a standard deviation increase in the combined math and reading test scores; this is very close to the estimate of 14.46 percent of a standard deviation presented for 2014 in Figure 1.

scores (gap of 10.8 percent of a standard deviation), combined scores (gap of 10.0 percent of a standard deviation), and higher likelihood of suspensions (gap of 1.4 percentage points or 4.1 percent of a standard deviation). Given these figures and the effect sizes presented above, a ten percent increase in the size of a voucher program would be expected to close between 2.9 and 6.8 percent of the test score gaps, and 9.2 percent of the gap in suspensions. The closure of these gaps is especially meaningful because students with poorer average academic outcomes (including Black students, Hispanic students, and students using free and reduced price lunch) tend to be over-represented in schools facing higher degrees of competitive pressure.

However, it is worth highlighting that, as illustrated in Table 1, the program has expanded by much more than ten percent; the number of scholarship users was nearly seven times higher in 2017-18 than in 2002-03—and thus the realized gains should actually be much larger. If we use our preferred Competitive Pressure Index estimates from Table 2 (Panel F), we would expect a more conservative doubling of the program size to result in a greater differential improvement of 5.1 percent of a standard deviation in combined math/reading scores in areas with high competitive pressure vs. in areas with low competitive pressure, and a 1.3 percentage point greater decline in suspension rates (9.4 percent relative to the sample mean); these effect sizes represent a meaningful change in the gaps in outcomes between schools facing higher and lower degrees of competitive pressure.

A potentially more intuitive set of benchmarks may also help contextualize the size of our reduced-form results. For example, when considering combined math and reading, the mean test score difference between students ever and never observed on free or reduced price lunch is 73 percent of a standard deviation. Likewise, the test score gap between children with high school-graduate and college-graduate mothers is over 60 percent of a standard deviation. The same gaps in likelihood of being suspended are 12.2 and 8.5 percentage points, respectively. Thus, it is worth noting that the program effects, even when considering maximum observed scale-up, are still relatively modest – although certainly economically meaningful – compared to these larger, long-standing gaps.

To further asses the magnitude of our estimates, it is also helpful to compare them to other estimates in the education and human capital formation literature, and in particular to those obtained using data from Florida. For instance, Figlio & Hart (2014), who looked at the introduction of the program, found that a one standard deviation increase in the pre-program competition pressure predicted a differential improvement in test scores of 1.5 to 2.7 percent of a standard deviation in combined math and reading standardized scores. We can most directly compare the present results to those if we draw on specifications that use continuous (rather than median-split) measures of competitive pressures, which we show in robustness checks presented later in the paper (Table 3, panel C). Those figures suggest that our conservative quantification, assuming doubling (tripling) of the program, would result in a roughly 1.5 (2.9) percent of a standard deviation increase in combined math and reading standardized scores. Thus, a doubling or tripling of the program size is required to increase the salience of competition pressure—with attendant benefits to public school student test performance – to the same degree as the initial introduction of the competitive pressures. At the same time, over the course of our sample, the program increased almost seven-fold, and thus the effects due to the growth of the program now outweigh the initial introduction effects.

A doubling of the program, yielding a 5.1 percent of a standard deviation effect, is also comparable to or larger than, depending on the exact outcome and specification, effect sizes from charter expansion studied by Ridley & Terrier (2018) and Gilraine, Petronijevic, & Singleton (2021). It is further about the same as the effect of a 10-percent increase in birth weight (Figlio et al., 2014), a quarter of the size of the effect of school entry cutoff on cognitive development (Dhuey et al., 2019), and 80% of the birth order gap in reading scores (Breining et al., 2020). More generally, it is equivalent to roughly 10-percent of the effect of child care subsidies on children's GPA (Black et al., 2014). Note, however, that all these papers estimate total effects of the treatment while our estimates represent gains in high-competition relative to low-competition areas, thus providing lower bound on the total effect. Nonetheless, we feel that these comparisons should help readers in understanding the magnitude of our coefficients which appear meaningful from the economic and education policy perspectives.

We view benchmarking the effect sizes off of a doubling of the program (i.e., a relative measure of increase in size) as useful because no other program has expanded to the same extent in absolute terms; however, as documented in Figure 1 the effective gains of Florida students between first year of program operation (2002-2003 school year) and end of our data span are much larger at 12 and 17 percent of a standard deviation in math (school year 2013-2014) and reading (school year 2016-2017), respectively. Importantly, it appears that even after almost seven-fold expansion and coverage of about 4 percent of K-12 population, and in some counties more than 7% of K-12 students and an even higher fraction of the K-8 students who are of particular interest in this paper, the test score gains do not decline.

C. Heterogeneity

We next address questions of whether different types of students differentially benefit from increased competitive pressure, running our analyses separately for each subsample of students in turn. We present the results using the composite Competitive Pressure Index measure for all five outcome variables (Tables A3 and A4 and panels B and C of Figure 1).[18]

Results are generally consistent in pattern across all subgroups; however, the exact magnitudes and statistical significance vary somewhat. Lower socioeconomic status students– whether measured by use of free or reduced-price lunch or by mother's education level– see larger effects across all outcomes. This is evident both in Table A3 and in Figure 1. Within test score outcomes, these differences are more pronounced for reading than for math. Given that more affluent children should never have been eligible for the program, the fact that the expansion of the program was nonetheless associated with improvements for this group of children in more competitive landscapes suggests that the benefits of competitive pressure are diffuse and extend, albeit to a lesser degree, to children that the public schools face no risk of losing to private schools due to the voucher program.

Similarly, we also can divide families into deciles of socio-economic status, using a measure introduced for these data by Autor et al. (2019). The socioeconomic status composite index is created through a principal components analysis, similar to our competitive pressure index. Specifically, the principal components analysis generates factor loadings based on mother's marital status, age, and years of education at birth as well as an indicator for whether the birth was Medicaid-funded and median zip code-level neighborhood income at the time of birth; and uses these to construct a composite socioeconomic status gradient index. We then separate the sample by SES deciles to observe scale up effects across the SES gradient.

Appendix Figure A4 presents coefficients associated with increasing competitive pressure on students by SES decile. Effects are strongest for families in the bottom six deciles, but expanded competitive pressure is associated with benefits for all families except for the very top SES decile. Taken together, these patterns of results suggest that voucher expansion may

[18] See Figlio, Hart, & Karbownik (2020) for results for each underlying measure separately.

work partly through stimulating competition near lower-SES schools, competition-induced additional focus on lower-SES students (regardless of school SES), or a combination of both.

We also observe differences in magnitudes across racial and ethnic groups as well as by immigration status (Table A4). First, test score gains are very similar for Black and White children. The former group, however, does not experience significant behavioral benefits in terms of absences or suspensions. Second, Hispanics experience larger gains in reading compared to the other two racial/ethnic groups but smaller and statistically insignificant increases in math. It also appears that increased competitive pressure is particularly beneficial for Hispanic students in terms of reductions in suspensions. Third, students with foreign-born mothers see a pattern of results comparable to that of Hispanic students. This is not surprising given that Hispanic children are disproportionately likely to have non-native born mothers in our sample (68 percent of Hispanic children in our sample has mother born outside of the U.S., compared to only 9 percent for non-Hispanic children).

D. Robustness

While our results vary somewhat across outcomes and different sociodemographic groups, they are largely robust to different modeling decisions. We present a set of robustness checks, using our preferred competitive pressure measure (the Competitive Pressure Index) in Tables 3 and 4. Our main results from Table 2 are reproduced in Panel A of each table for ease of comparisons.

A first set of tests in Table 3 looks at whether our results are sensitive to different constructions of the competitive pressure measure that underlies the interaction term. Our main measure uses a school-level sample to determine whether each school had a higher- or lower-than-median level of competition pre-policy; i.e., each school entered the sample to determine the median split once regardless of how many students attended. Panel B uses a measure that defines the median based on the degree of competition faced by the median student's school, rather than the median school. These results are very similar to the main results. Panel C uses the underlying, continuous Competitive Pressure Index measure of pre-program competitive pressures in our interaction terms rather than the median-split term. The pattern of results remains the same, although consistent with the fact that this measure has a different underlying distribution, the magnitudes of the coefficients are predictably different compared to results in panel A.

Another way to look at the influence of the pre-voucher private school landscape is to split the private school competition variable at a more granular level than the above/below median division used in our main specifications. Figure 2 shows point estimates for versions of the models where the competitive pressure variable is stratified into quintiles of competition and interacted with the logged expansion measure. Panel A shows results for test scores while Panel B shows results for our behavioral outcomes. The effects of expansion on test scores are more pronounced for schools with higher levels of competition pre-policy, and this pattern is particularly evident for reading. Absences show similar patterns to the cognitive effects, with more competitive landscapes associated with consistently greater improvements in the outcome across the span of competition. On the other hand, for suspensions, there is relatively little difference in the second and third quintiles of pre-policy competition compared to the schools with the least-competitive landscapes, but pronounced differences emerge for schools in the fourth- and fifth-quintiles. Across the board, however, we see consistent patterns of more pronounced advantages for students from schools in the most competitive areas as the program expands, roughly consistent with a dose-response relationship.

Table 3, Panel D uses an alternate measure of program expansion, substituting a logged measure of program funding in place of the logged measure of student participation. In panel E, we assign the middle-school pre-policy competition measures based on the actual grade 6 (middle) schools initially attended by each student, thus potentially allowing for endogenous selection into middle school based on its quality. The results in both cases remain unchanged.

Our main results use all available data to generate flows of students between elementary and middle schools that we then use to generate the expected competitive pressure measures faced by children in middle school. However, this approach may be subject to endogeneity if these flows are affected by program expansion differentially in areas with higher vs. lower baseline competition pressures. Thus, in panel F, we investigate if our results persist when we generate our expected competitive pressure measures for middle school-stage students using only pre-policy announcement flows between elementary and middle schools. Since our data do not go back far enough to track children from grade 1 to grade 6 using only pre-announcement cohorts, we cannot execute this analysis based on first attended grade but rather utilize transitions between grades 5 and 6 for school years 1999 to 2001. As shown in Panel F, this refinement is inconsequential. Finally, in Panel G, we eliminate the distinction between elementary and middle schools altogether and consistently assign students to the level of pre-program competition experienced by their grade-one school. Again, our conclusions remain unchanged.

The next set of tests in Table 4 checks for sensitivity of results to different modeling assumptions and changes in the exact samples used. Panel B examines the possibility that our results may be driven partly by regional factors associated both with competitive landscapes, changes over time, and with broader student success. To address this concern we include a set of region-by-year fixed effects.[19] This specification produces results that are qualitatively similar, albeit smaller in magnitude than our main results, particularly for reading and suspensions. This suggests that part of the effects of competition may be between regions rather than solely at a hyper-local level, a point to which we return in Section V.

Our results are also largely robust to different sample choices. Our earlier work (Figlio & Hart, 2014) used a sample restricted to students attending schools with at least one competitor within five miles. Table 4, Panel C imposes a similar restriction, and we reconstruct the median split to reflect the exclusion of the schools that do not have a competitor with 5 miles from the analysis. Panel D tests whether our results are sensitive to restricting our sample to the set of years for which we have all outcomes available (ending our sample in 2011-12, after which we no longer observe suspension and absenteeism outcomes). Panel E tests sensitivity to restricting our sample to the panel of students whom we observe for six consecutive years, when a normal progression would have taken them from grade 3 to grade 8. In all cases, our results remain similar.

The final two panels (F and G) address the concern that our test measures in the main specifications are standardized using the Florida-born sample rather than a statewide sample. This decision allowed us to use a greater set of years, because we only have data to standardize scores using the entire state population through the 2012/13 school year. Panel F replicates our main specification using our sample-standardized outcome measure and dropping years after 2012/13 school year while Panel G substitutes in the set of test score measures standardized using the whole state population. Again, these results are very similar to our primary findings.

[19] We identify six regions in Florida, using Florida Association of Counties designations. The six regions are: Northwest, Northeast, West Central, East Central, Southwest, and Southeast.

In a final check of the robustness of our results to the use of different samples, we estimate a series of "leave-one-out" models where we drop each district in turn to ensure that no single district is driving our results (Appendix Figure A5). In almost all specifications, the results remain very similar. However, they are slightly more sensitive for reading and suspensions than other outcomes when we leave out Miami-Dade County, the largest school district in the state; the suspensions coefficient becomes non-significant in this case.

A final and critical concern for our identification is that there may be secular improvements over time that happen to be more pronounced in areas with high competition, but that are occurring regardless of the voucher policies. In Figure 3, we test this using data from a different sample based on all public school students but a limited set of years as described in Section II A. Here we do not have student outcome data in all grades in all years (specifically in the pre-policy years), and thus looking at within-student changes is not feasible to examine pre-policy trends. For this reason, we must control for school fixed effects rather than student fixed effects in this specific test. One worry inherent in this approach is that students might select into grade-one schools as a consequence of the policy. To ameliorate the concern that student sorting to grade-one schools may be affected by the policy, we limit the sample to students who started schooling prior to the policy's introduction (i.e., those born before September 1, 1994) and therefore whose grade-one schools prima facie could not have been affected by the policy. The relationship between competition and student outcomes in each year is compared to the relationship in the omitted year of 2001 (Figure 3, triangle markers). We provide additional information about these analyses in Online Appendix B.

Because in this analysis we change both the identification strategy (school fixed effects rather than individual fixed effects as in our main results) and the years used (1999-2007 rather than 2003-2017 as in our main results), we also include a set of point estimates that shows the estimates that would be generated, to the extent feasible, if each of these changes was implemented in our preferred sample. We include point estimates that replicate our main estimation strategy (individual fixed effects) using 2003-2007 data (circle markers), as well as a set of point estimates that uses our 2003-2007 sample, but includes school-by-stage rather than individual fixed effects (square markers). Both of these sets of estimates use an omitted year of 2003 as in Figure 1. The pattern of results in those analyses looks very similar to results in Figure 1, Panel A, with slightly more positive estimates using the school fixed-effects models.

Figure 3 shows that pre-policy competition is unrelated to changes in student scores before the voucher policy was announced in spring 2001. The coefficients for school year 1999 and 2000 are not significantly different from zero in five out of six cases, and the only statistically significant result (for reading in 1999) suggests trends that, if anything, ran in the opposite direction of our post-policy findings. Thus, our reading results can be considered conservative estimates.[20] Thereafter, as for the main results, we see that schools in more competitive areas improve more quickly than schools in less competitive areas, with positive and significant coefficients in each year from 2002 onward and for each outcome. These figures strongly suggest that there were no pre-existing trends boosting outcomes for students in higher-competition areas prior to the voucher policy, providing further support that the exogeneity-of-shares assumption for our shift-share research design meets the standards described by Goldsmith-Pinkham, Sorkin, & Swift (2020). Thus, we are confident that our results reflect effects of the scale up of the program itself differentially affecting schools in higher-baseline competition areas, rather than reflecting any prior differential trends.

IV. Alternative Explanations

So far, we have suggested that our results are due to increased competitive pressure on public schools associated with voucher program expansion. However, there may be alternative explanations to these findings. For instance, voucher programs may change the composition of students remaining in the public schools over time, and these changes might be related to the degree of voucher competition individual schools face. Voucher programs also influence the resources that public schools have, and these resource effects might work in opposite directions. On the one hand, the voucher program reduces funding to school districts that lose state funding allocations for students attending private school. On the other hand, if the vouchers lead to fewer students per grade, class sizes might plausibly go down in the public schools. To the extent possible, we investigate these alternative explanations in this section.

[20] Our results are likely conservative for another reason. There was putatively another voucher policy operating prior to the introduction of the FTC program. The Opportunity Scholarship Program, announced in 1999 in conjunction with the state's new accountability system, provided students in schools with two consecutive "F" grades with vouchers to use at either public or private schools. However, the program was under legal challenge almost as soon as it was announced, and had very limited uptake. As of summer 2002, students in only 8 schools statewide were eligible for the program (Rouse et al., 2013), and according to program officials, it never served more than a few hundred students before the private school voucher aspect was struck down in 2006 by the Florida Supreme Court (Figlio & Hart, 2014). To the extent that the OSP contaminates our results, it should make the contrast between the pre- and post-FTC periods look more similar, rendering our results conservative.

Beyond what we can examine empirically, another alternative explanation could be changes in other policies that might likewise be driving gains in test scores and improvements in disciplinary outcomes. Most obviously, the national No Child Left Behind Act (NCLB) was under discussion at the same time that the FTC program was passed. Since schools with greater pre-program competition were lower performing (Table A2), one might be concerned that this legislation, which was intended to put pressure on low-performing schools, may be driving our findings. We think this is unlikely for several reasons. First and foremost, it would have to be the case the accountability pressure has been growing in the same way as the voucher program and there is no policy evidence on that. Second, prior to NCLB, Florida already had comparably stringent accountability law which, however, did not put substantial pressure on public schools (Rouse et al., 2013). Finally, our results extend to higher-SES students that are unlikely to be targeted by either the voucher program or the NCLB and their test scores likewise grow, albeit at smaller rate than for lower-SES students, as the program scales up, implicating increasing competition as a more likely driver of our results.

A. Peer Composition

First, consider the possibility that our results are due to changes in school composition brought about by differing degrees of voucher competition. These composition changes could result in observed impacts through peer effects associated with who remains in the public schools as the program scales up. For instance, if students who leave public schools to use the voucher program tend to be lower-achieving or more subject to disciplinary problems on average, then the loss of those peers to the private sector could leave behind an easier-to-educate core of students, and result in positive impacts on student learning. These compositional changes could produce benefits even if schools exert no more effort in response to the competitive pressure caused by vouchers.

To investigate this, we carry out analyses to see whether schools facing increased competitive pressure have students remaining in the school who would have had higher predicted test scores and lower predicted rates of suspensions and absences, all else equal, based solely on their background characteristics. Columns 1-5 of Table 5 present the results of an analysis that is parallel to our main specification, with two key changes. First, this analysis is conducted at the school level rather than the student level and accordingly uses school-by-stage fixed effects rather than student-by-stage fixed effects. Second, the dependent variable is the average of predicted peer outcomes in each school, with the predicted values based solely on background characteristics of the students enrolled.[21] If we see that schools with more competition also have student cohorts with higher predicted scores (or lower suspension and absence rates) enrolled over time as the program scales us, this would provide evidence that changes in student composition, rather than any effort by the school, may explain the effects we documented above.

This exercise produces no strong evidence that positive peer effects drive our results. For only one peer outcome (suspensions) is there even a marginally significant relationship, but the coefficient is in the opposite direction as our main results, suggesting that schools in more competitive landscapes were more likely to enroll students with higher predicted suspension risk as the voucher program expanded. That would be akin to reverse cream skimming, work against our findings, and thus lead to lower bound estimates. Even for the coefficients where the peer effects operate in the same direction as for our main results in Table 2 (reading and absences), the coefficients in Table 5 are negligible in magnitude compared to our main results e.g., the point estimate for predicted reading (Table 5) is one-fiftieth of the point estimate for actual reading scores (Table 2). This is on top of the fact that peer effects in test scores literature in general suggests relatively small, if any, effects on students (see e.g., Sacerdote (2014) for a recent review). Overall, this suggests that our results are unlikely to be driven by changes in student composition associated with increased voucher utilization.

B. Resources

The voucher program could also have induced changes in resources received by affected schools. We lack reliable data on most measures of school resources, and candidate measures such as school-level measures of expenditures, when they exist, largely reflect either teacher

[21] Specifically, we regress individual-level test scores and behavioral outcomes onto student background characteristics measured at birth (child's month and year of birth, sex, birth weight, gestational age, birth order, prenatal care start, abnormal conditions at birth and congenital anomalies; as well as mother's education, race, ethnicity, place of birth outside of U.S., state of birth other than Florida, health problems, age, marital status and Medicaid-paid birth), and use the resulting coefficients to predict outcomes for each student. The R^2 from these regressions for combined math and reading, math, reading, suspensions, and absences are 0.24, 0.21, 0.21, 0.10, and 0.07, respectively. Thus, our explanatory variables have non-trivial predictive power for these outcomes, or at least for test scores. We then aggregate these predicted outcomes at the G1 school-by-stage-by-year level and link it to competition-by-expansion interaction measured at G1 school-by-stage-by-year level since this is the level of variation in out pre-program competition measures as explained in Section II. The results are very similar if instead we aggregate both outcomes and competition at G1 school-by-year level. They are likewise similar when weighted by number of students in each aggregated cell.

experience levels or student participation in special education (Lankford & Wyckoff, 1995).[22] But arguably the most salient resource indicator – average class size in a school – is measurable in the Florida data. We therefore explore whether our estimates of the competitive effects of voucher scale-up are potentially due to changes in class size associated with increases in competitive pressure. These may occur mechanically, to the extent that voucher programs draw students away from the public schools they would have attended, or may be the result of strategic decisions by principals to make the school more attractive to students and parents. We draw on class size archives from 2006-07 through 2016-17 posted by the Florida Department of Education, which report the average class size for each school separately for students in grades PK-3, grades 4-8, and grades 9-12.[23] For each school and year, we calculate the weighted average class size at the school level based on the grade range served by each school. This class size variable, varying at school-by-year level, is then used as an outcome in regressions akin to those used to produce our peer effect estimates.

Table 5, column 6, shows the results of this exercise. Schools facing landscapes with more initial competitive pressure did have somewhat smaller class sizes as the program expanded. However, based on past literature on the relationship between class size and student outcomes, the magnitudes of the coefficients are too small to realistically explain away much of our main cognitive and behavioral effects. Our estimate of -0.221 implies a 0.022 student reduction in class size per 10 percent increase in the program. To contextualize the expected effects on test scores of a reduction of this magnitude, we make use of the 22 percent of a standard deviation increase in test scores associated with a roughly seven-student reduction in class size effect implied by the Tennessee STAR experiment (Krueger, 1999); this estimated effect is similar in magnitude to those found by Angrist & Lavy (1999), Lindahl (2005), Chetty et al. (2011), and Fredriksson et al. (2013) in related studies. If we assume the same proportionate effect would apply to competition-induced class size decrease, a 0.022 reduction in class size would imply expected test score improvements scarcely different from 0 (0.022*22/7=0.07 percent of a standard deviation). This would account for only about 14% of

[22] Many sources of data on education finance—like the NCES Local Education Agency Finance Surveys and the NCES Longitudinal School District Fiscal-Nonfiscal Files—reflect finances at the district level. In Florida, districts are very large, and schools within the same district could plausibly be exposed to very different levels of competition, limiting the usefulness of these files for our purposes.

[23] Ideally, we would observe class size information starting in 2002/2003 school year; however, data for these earlier cohorts are not available.

the effect of competition on test scores that we estimated in Table 2, where our coefficient on the PCA measure implied a 0.511 percent of a standard deviation increase predicted for a 10 percent program expansion.

Given that the range of results in the extant literature (e.g., Jepsen & Rivkin, 2009; see Chingos, 2013 for a thorough review of the class size literature) estimate that class size reductions of about 10 students produce improvements in test scores of between 0.05 and 0.22 standard deviations, and other papers (e.g., Leuven & Løkken, 2020) sometimes find very small class size effects on other meaningful outcomes, the share of our estimated effects that can be explained by class size reductions may be even smaller still. While class size may be contributing to the observed positive effects of competitive pressure on cognitive and behavioral outcomes, and clearly suggests that public schools change some inputs in response to increased competitive pressures, it seems unlikely to be a quantitatively meaningful driver of the findings.

V. District-Level vs. Local Competition

A key remaining question is the level at which competition matters most. On the one hand, perhaps schools are primarily affected by competition only in their immediate vicinity, and competition throughout the rest of the district is irrelevant. This may be the case if school-level efforts were the primary channel through which competitive effects worked and evidence from charter school expansions supports this notion (Gilraine, Petronijevic, & Singleton 2021). On the other hand, if a sizable portion of the response to competition comes from district-level policies (or from formal or informal conversation between school leaders, who meet at the district level), then we might observe district-level competition being more (or equally) important as the immediate vicinity of each school. Of course, school-level competition – measured in terms of geographic proximity – is also likely measured with more error than district-level competition, as counties have defined boundaries, so that could also help to explain a finding of district-level competition mattering more than school-level competition. Even beyond the measurement error issue, it is not clear *ex ante* how – in terms of geography – we should define a market when it comes to the school competition, and plainly there are policies at both school- (e.g., class size) and district-level (e.g., teacher compensation schemes) that might be changed in response to the increasing competitive pressures.

To examine this, we create a district-level version of our competition measure, capturing whether the district-level competition each child is expected to face, based on their first-grade school, was above or below median. The district-level competition measure is created by aggregating the school-level measure, weighted by student population, to generate the average degree of competition faced by each student in each district. In order to use a median split number that has roughly equal numbers of students in above- and below-median districts, we construct the splits using the level of competition faced by the median *student* as the cutpoint.[24] To make sure that the school and district-level competition variables are consistent with each other, we also use the student-level median to stratify the school-level competition variable, paralleling the version of the measure used in the robustness check in Panel B of Table 3. As for the rest of the analyses, the school and district-level competition measures here are both based on pre-policy private school landscapes, and the above-median indicators are interacted with the year-specific measure of voucher program expansion. Pearson correlation coefficient between the two measures is moderate, at 0.62, which means that there is a degree of independence between school- and district-level competition.[25]

Results are given in Table 6. Column 1 replicates the robustness check from Panel B of Table 3 for comparison, while Column 2 supplements the school-level competition-by-expansion interaction term with the district-level competition-by-expansion interaction term. For every outcome except for absences (Panels A-E), the effect of district-level competition dominates the school-level measures. In all cases the school-level interaction terms fall in magnitude compared to Column 1, although with the exception of math scores, they retain both their signs and statistical significance. This pattern of results suggests that while the local, neighborhood-level competition that schools face matters, there is an independent effect of being in a higher-competition district, suggesting the potential importance of district-level responses to the salience of private school competition.

In supplemental analyses (Appendix Table A5), we create separate categories capturing both the school-level and district-level competition simultaneously. That is, we categorize whether each student is in a high-competition school/high-competition district (41% of sample); low-competition school/high-competition district (10% of sample); high-competition school/low-

[24] There are two small districts that are narrowly classified as above-median in terms of the competition faced by their elementary schools but below-median in terms of the competition faced by their middle schools. In order to treat these districts similarly at both stages, we define them as above-median competition at both stages, but our results are not sensitive to this decision.

[25] For comparison, this is lower than Pearson correlation between math and reading scores in our data, which is 0.72.

competition district (9% of sample), or low-competition school/low-competition district (40% of sample, omitted category). We find that while each of the other configurations has increasingly positive outcomes associated with the expansion of the voucher program relative to the low-competition school/low-competition district reference group, the exact pattern differs depending on the outcome. Effects are largest when both levels of competition are above median for reading, suspension, and absences outcomes. For math and the math/reading composite outcome, effects are largest when school experiences lower levels of local competition but is located in a district with above median competition. Furthermore, in all cases, we can reject the null hypothesis that these competition categorizations have identical estimated effects; with F-statistics ranging from 5.7 to 80.2

One concern that these results may raise is that different district-level efforts actually may drive our results rather than the expansion of school competition per se. In columns 3 to 7 of Table 6 (and in even-numbered columns of Table A5), we explore whether other district-level variables (along with school-level variables pertaining to our mechanisms explored in Section IV) might explain away the apparent effects of competition. For instance, perhaps public school choice options—magnets and charter schools—have been simultaneously expanding more rapidly in districts with greater pre-program private school competition. If so, these schooling options, rather than private school competition per se, may explain our results. In Column 3, we add a control for the share of magnet and charter schools per 1000 students in the district. Our results are little changed for most outcomes, suggesting that other district-level forms of competition do not explain away our results. An exception is that both the school- and district-level competition interaction terms fall in magnitude for the suspensions outcome, and the school-competition-by-expansion interaction becomes non-significant.

In Column 4, we test whether the results in Column 2 are robust to the inclusion of district-level average salaries. This addresses the possibility that districts with more competition could also have been undertaking other policy changes that could improve educational outcomes—such as offering higher salaries to recruit and retain a more stable teaching workforce. Adding teacher salary measures has minimal effects on the coefficients for either school-level or district-level competition.

Similarly, in Columns 5 and 6, we explore whether, respectively, the inclusion of school-level class-size information or the inclusion of predicted peer ability levels introduced in Table 5 alters the results. The inclusion of these variables does little to meaningfully move the coefficients on the expansion-by-competition measures at either the school- or district-level. Finally, in Column 7, we include all district- and school-level variables simultaneously. For most outcomes, Column 7 produces results very similar to Column 2, again with the exception that the estimates on suspension become non-significant with respect to school-level competition which is driven by changes introduced in Column 3.

Taken together, these results suggest that cross-district differences in exposure to private school competition matter beyond local levels of competition investigated in most prior research. In our application, they are particularly important drivers of the improvement in math. For the remaining outcomes, even within districts, being more exposed to local competition continues to have meaningful independent effects. This suggests that there may be district-level mechanisms at work, such as district policies adopted in response to competitive threats, but we are unable to disentangle these mechanisms further given the data at hand.

V. Conclusions

School choice programs have been growing in the United States and worldwide over the past two decades, and thus there is considerable interest in how these policies affect students remaining in public schools. Although we now have relatively comprehensive knowledge on the immediate short-run effects stemming from the introduction of such programs, the evidence on the effects of these programs as they scale up is virtually non-existent. Here, we investigate this question using data from the state of Florida where, over the course of our sample period, the voucher program participation increased nearly seven-fold. We build on past research in that, to date, this is the largest voucher expansion in U.S.; it represents the largest school voucher program in the country; and we can study it over 15 years, whereas previous research focused on much smaller scale expansions and was mostly limited to studying effects one to four years after the program's introduction.

We find consistent evidence that as the program grows in size, students in public schools that faced higher initial competitive pressure levels see greater gains from the program expansion than do those in locations with less initial competitive pressure. Importantly, we find that these positive externalities extend to behavioral outcomes— absenteeism and suspensions—that have not been well-explored in prior literature on school choice from either voucher or charter programs. Our preferred competition measure, the Competitive Pressure Index, produces estimates implying that a doubling in the number of students participating in the voucher program increases test scores by 3 to 7 percent of a standard deviation and reduces behavioral problems by 6 to 9 percent relative to their sample means. This benchmark is greater than most other expansions studied in the extant literature but much smaller than the seven-fold scale up of the program that happened in Florida. We show that these results are very robust to alternative plausible ways of measuring competition and expansion, as well as to different modeling choices. They also cannot be explained by changes in student composition or school resources, to the extent that these are measurable in our data. Our results are also consistent with past work showing modest benefits to the initial introduction of voucher programs (e.g., Hoxby, 2003; Figlio & Hart, 2014; Egalite, 2016; Egalite & Wolf, 2016; Figlio & Karbownik, 2016), while extending upon these findings to show the persistence and growth of these positive effects as the program scaled up. Importantly, our findings generate additional nuance in demonstrating that the level of competition faced by the district could be even more important than the marginal degree of competition faced by the individual school in driving the effects of the expansion of the voucher program on student outcomes. This suggests that future work that looks at district-level responses (or peer networks within a school district) as potential mechanisms may be a fruitful area for research.

Finally, we find that public school students who are most positively affected come from comparatively lower socioeconomic background, which is the set of students that schools and districts should be most concerned about losing under the Florida Tax Credit Scholarship program. However, smaller effects remain statistically significant – in most cases – even for students who are very unlikely to be targeted by vouchers themselves, suggesting that benefits may come partially through generalized school improvements rather than through improvements targeted solely at voucher-eligible students.

References

Angrist, J.D., & Lavy, V. (1999). Using Maimonides' Rule to estimate the effect of class size on scholastic achievement. *Quarterly Journal of Economics,* 114(2), 533-575.

Arsen, D., & Ni, Y. (2008). *The competitive effect of school choice policies on traditional public schools.* Arizona State University/University of Colorado at Boulder. Tempe, AZ: Education Public Interest Center.

Autor, D., Figlio, D., Karbownik, K., Roth, J., & Wasserman (2019). Family disadvantage and the gender gap in behavioral and educational outcomes. *American Economic Journal: Applied Economics,* 11(3), 338-381.

Bartik, T. (1991). Who benefits from state and local economic development policies? Kalamazoo, MI: W.E. Upjohn Institute.

Black, S.E., Devereux, P.J., Løken, K.V., & Salvanes, K.G. (2014). Care or cash? The effect of child care subsidies on student performance. *Review of Economics and Statistics,* 96(5), 824-837.

Borusyak, K., Hull, P., & Jaravel, X. (2021). Quasi-experimental shift-share research designs. *Review of Economic Studies*, forthcoming

Breining, S., Doyle, J., Figlio, D., Karbownik, K., & Roth, J. (2020). Birth order and delinquency: Evidence from Denmark and Florida. *Journal of Labor Economics,* 38(1): 95-142.

Chakrabarti, R. (2008). Can increasing private school participation and monetary loss in a voucher program affect public school performance? Evidence from Milwaukee. *Journal of Public Economics,* 92, 1371-1393.

Chetty, R., Friedman, J.N., Hilger, N., Saez, E., Schanzenbach, D.W., & Yagan, D. (2011). How does your kindergarten classroom affect your earnings? Evidence from Project STAR. *Quarterly Journal of Economics,* 126(4), 1593-1660.

Chingos, M.M. (2013). Class size and student outcomes: Research and policy implications. *Journal of Policy Analysis and Management,* 32(2), 411-438.

Cohodes, S., Setren, E., & Walters, C. (2019). Can successful schools replicate? Scaling up Boston's charter school sector. NBER working paper 25796.

Dee, T.S. (1998). Competition and the quality of public schools. *Economics of Education Review,* 17(4), 419-427.

Dhuey, E., Figlio, D., Karbownik, K., & Roth, J. (2019). School starting age and cognitive development. *Journal of Policy Analysis and Management,* 38(3), 538-578.

EdChoice. (2019, February 12). *School choice in American*. Retrieved from EdChoice: https://www.edchoice.org/school-choice/school-choice-in-america/#

Egalite, A. J. (2016). *Competitive impacts of means-tested vouchers on public school performance: Evidence from Louisiana and Indiana.* Cambridge, MA: Program on Education and Governance Working Paper Series, Harvard Kennedy School. Retrieved February 23, 2019, from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.696.8049&rep=rep1&type=pdf

Egalite, A. J., & Wolf, P. J. (2016). A review of the empirical research on private school choice. *Peabody Journal of Education,* 91(4), 441-454.

Epple, D., & Romano, R.E. (2008). Educational vouchers and cream-skimming. *International Economic Review,* 49(4), 1395-1435.

Epple, D., Romano, R. E., & Urquiola, M. (2017). School vouchers: A survey of the economics literature. *Journal of Economic Literature,* 55(2), 441-492.

Figlio, D. N., & Hart, C. M.D. (2014). Competitive effects of means-tested school vouchers. *American Economic Journal: Applied Economics,* 6(1), 133-156.

Figlio, D.N., Hart, C.M.D., & Karbownik, K. (2020). Effects of scaling up private school choice programs on public school students. NBER Working Paper 26758. Accessed from https://www.nber.org/papers/w26758

Figlio, D., & Karbownik, K. (2016). *Evaluation of Ohio's EdChoice Scholarship Program: Selection, competition, and performance effects.* Washington, DC: Thomas B. Fordham Institute.

Figlio, D., Guryan, J., & Karbownik, K., Roth, J. (2014). The effects of poor neonatal health on children's cognitive development. *American Economic Review,* 104(12), 3921-3955.

Florida Department of Education. (2018, September). *Fact Sheet: Florida Tax Credit Scholarship Program*. Retrieved February 13, 2019, from http://www.fldoe.org/core/fileparse.php/5606/urlt/FTC-Sept-2018.pdf

Florida Department of Education. (n.d.a). *Student Enrollment: State Level.* Retrieved January 31, 2019 from https://edstats.fldoe.org/SASWebReportStudio/gotoReportSection.do?sectionNumber=1

Florida Department of Education. (n.d.b). *FCAT Historical*. Retrieved February 3, 2019, from http://www.fldoe.org/accountability/assessments/k-12-student-assessment/archive/fcat/

Florida Statute 220.187 (2001).

Fredriksson, P., Öckert, B., & Oosterbeek, H. (2013). Long-term effects of class size. *Quarterly Journal of Economics,* 128(1), 249-285.

Gilraine, M., Petronijevic, U., & Singleton, J.D. (2021). Horizontal differentiation and the policy effect of charter schools. *American Economic Journal: Economic Policy*, forthcoming

Goldsmith-Pinkham, P., Sorkin, I., & Swift, H. (2020). Bartik instruments: What, when, why, and how? *American Economic Review*, 110(8), 2586-2624.

Greene, J. P., & Winters, M. A. (2007). An evaluation of the effect of DC's voucher program on public school achievement and racial integration after one year. *Catholic Education: A Journal of Inquiry and Practice,* 11(1), 83-101.

Hoxby, C.M. (1994). Do private schools provide competition for public schools? National Bureau of Economic Research (Cambridge, MA) Working Paper No. 4978,

Hoxby, C. M. (2003). School choice and school productivity: Could school choice be a tide that lifts all boats? In C. M. Hoxby, *The Economics of School Choice* (pp. 287-241). Chicago, IL: University of Chicago Press.

Jepsen, C. (2002). The role of aggregation in estimating the effects of private school competition on student achievement. *Journal of Urban Economics,* 52: 477-500.

Jepsen, C., & Rivkin, S. (2009). Class size reduction and student achievement: The potential tradeoff between teacher quality and class size. *Journal of Human Resources,* 44(1), 223-250.

Krueger, A.B. (1999). Experimental estimates of education production functions. *Quarterly Journal of Economics,* 114(2), 497-532.

Lankford, H., & Wyckoff, J. (1995). Where has the money gone? An analysis of school district spending in New York. *Educational Evaluation and Policy Analysis,* 17(2), 195-218.

Leuven, E., & Løkken, S.A. (2020). Long-term effects of class size in compulsory school. *Journal of Human Resources,* 55(1), 309-348.

Lindahl, M. (2005). Home versus school learning: A new approach to estimating the effect of class size on achievement. *Scandanavian Journal of Economics,* 107(2), 375-394.

National Center for Education Statistics (2013). Schools and Staffing Survey (SASS), "Private School Data File," 1999-2000, 2003-04, 2007-08, and 2011-12. Accessed 4/7/2020 at https://nces.ed.gov/programs/digest/d13/tables/dt13_205.50.asp.

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica,* 49(6), 1417-1426.

Ridley, M., & Terrier, C. (2018). Fiscal and education spillovers from charter school expansion. NBER working paper 25070.

Rouse, C. E., Hannaway, J., Goldhaber, D., & Figlio, D. (2013). Feeling the Florida heat? How low-performing schools respond to voucher and accountability pressure. *American Economic Journal: Economic Policy, 5*(2), 251-281.

Sacerdote, B. (2014). Experimental and quasi-experimental analysis of peer effects: two steps forward? *Annual Review of Economics*, 6, 253-272.

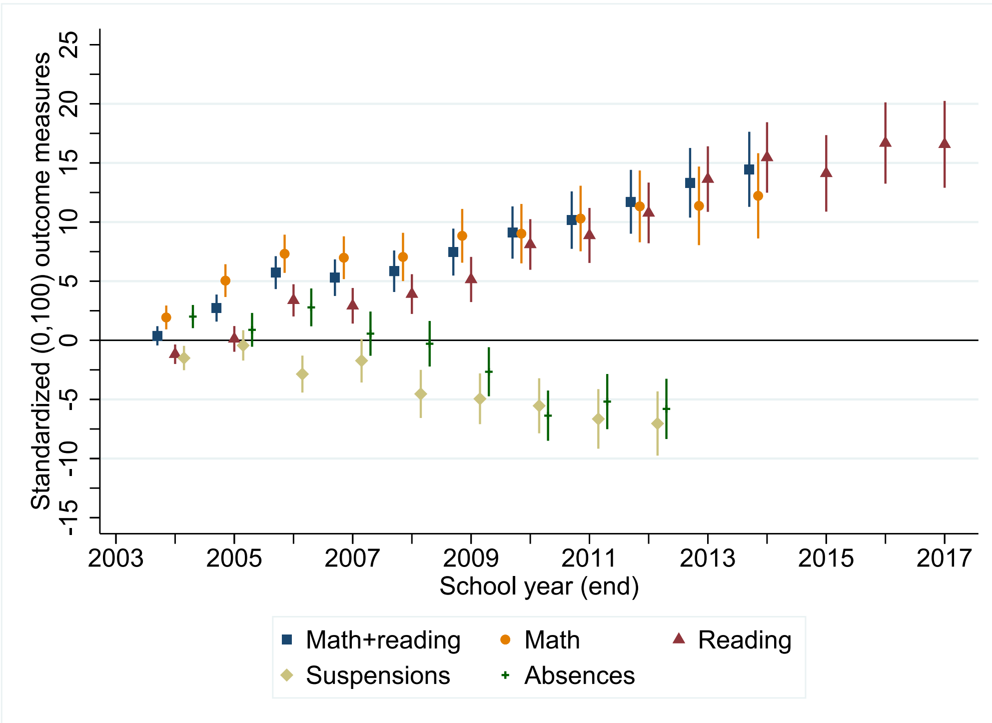
Totenberg, N., & Naylor, B. (2020). Supreme Court: Montana can't exclude religious schools from scholarship program. *National Public Radio,* June 30, 2020. Accessed at: https://www.npr.org/2020/06/30/883074890/supreme-court-montana-cant-exclude-religious-schools-from-scholarship-program

Urquiola, M. (2016). Chapter 4. Competition among schools: Traditional public and private schools. In E. A. Hanushek, S. Machin, & L. Woessmann, *Handbook of the Economics of Education* (pp. 209-237). Oxford: Elsevier.

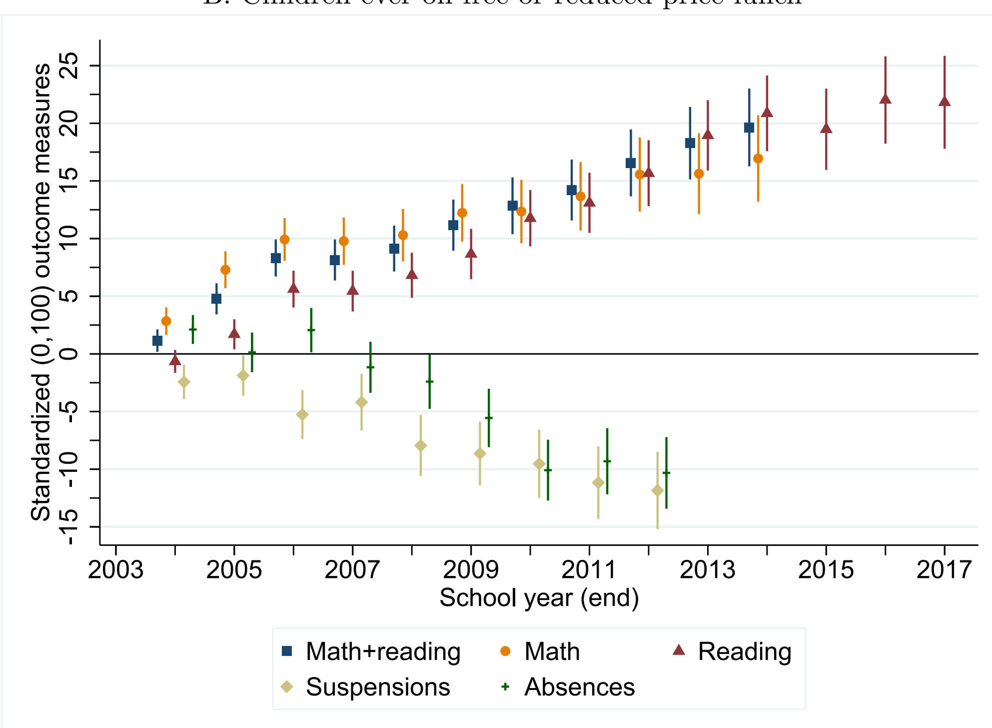
Figures

Figure 1: Effects of voucher expansion over school years for standardized outcomes

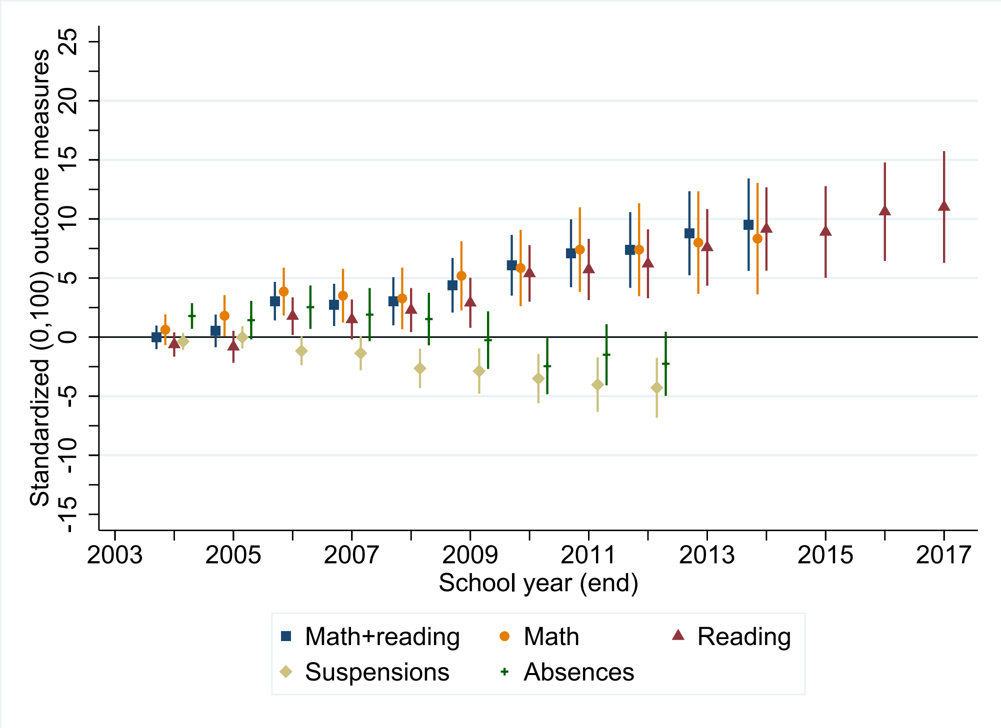
A. Pooled



B. Children ever on free or reduced price lunch



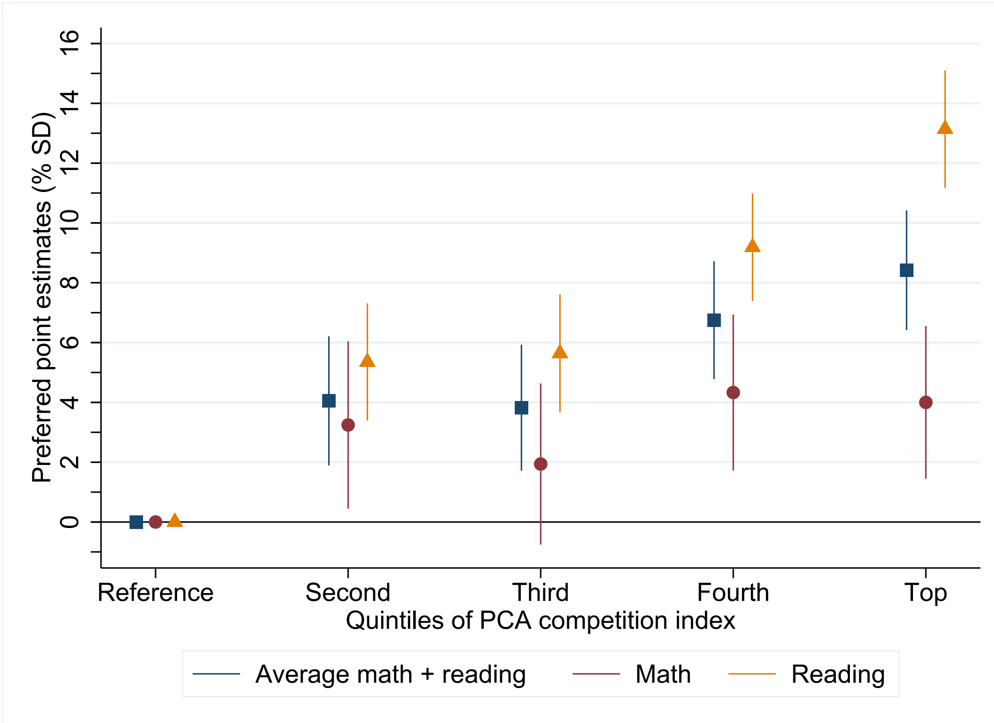
C. Children never on free or reduced price lunch



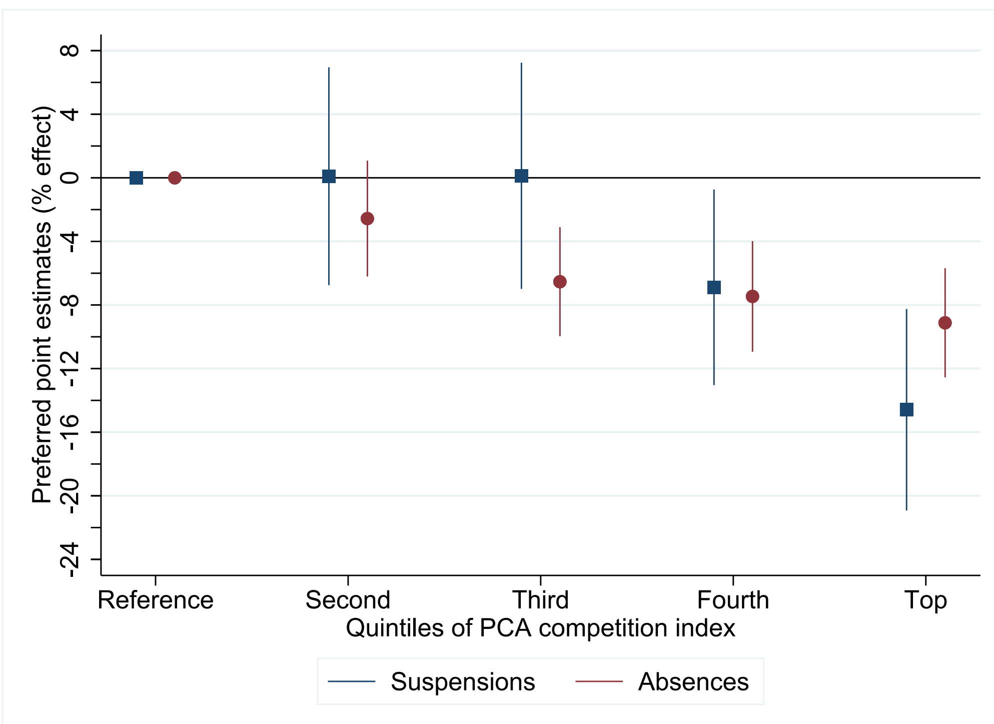
Note: These figures plot modified estimates from the main specification estimated in panel F of Table 2 and from heterogeneity analysis from panels A and B of Table A3, where instead of interaction between composite competition measure and log number of scholarships we plot composite competition measure interacted with school years, and with baseline omitted year 2002/2003. Panel A presents results for pooled sample (Table 2) while panels B and C divide the sample by free and reduced price lunch status of a child (Table A3). Outcomes are averaged test scores in mathematics and reading (navy squares); mathematics test scores (orange circles); reading test scores (maroon triangles); likelihood of being suspended (khaki diamonds); and absence rate (green pluses). Each outcome variable is standardized in its empirical sample to have mean zero and standard deviation of 100. Spikes present 95% confidence intervals based on standard errors clustered at grade one school level.

Figure 2: Quintiles of competition

A. Test scores



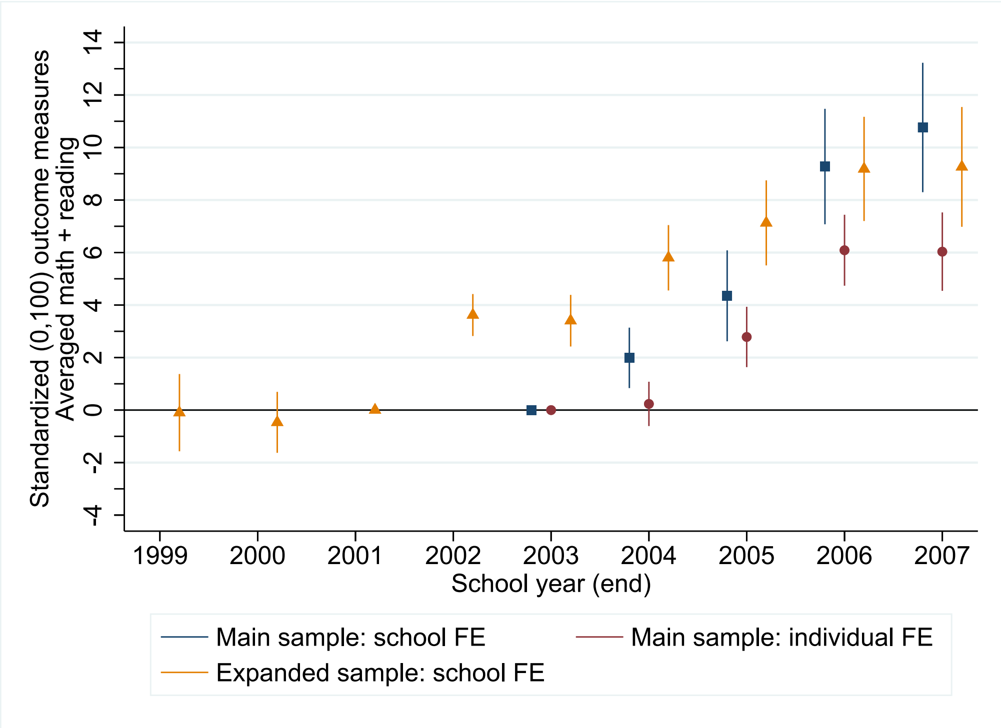
B. Behavioral outcomes



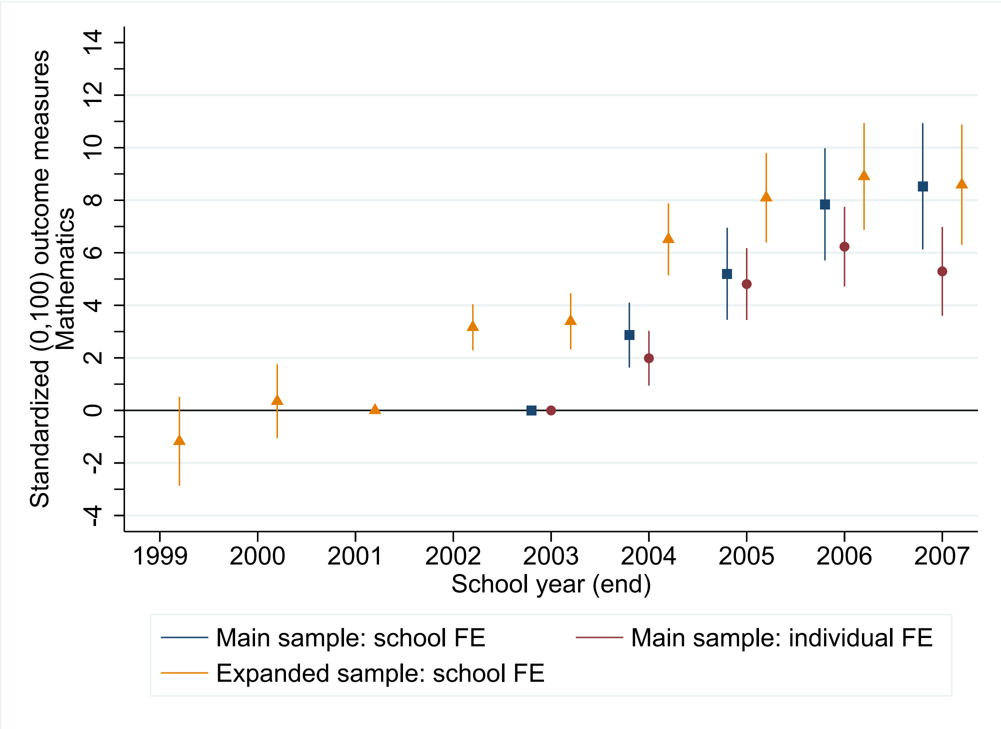
Note: These figures present estimates using specification and sample from panel F of Table 2 where instead of median we interact quintiles of PCA competition index with log of expansion measure. Bottom quintile is a reference category. Outcome variables are averaged mathematics and reading test score (navy squares), mathematics test score (maroon circles), and reading test score (orange triangles) in Panel A as well as likelihood of being suspended (navy squares), and absence rate (maroon circles) in Panel B. Spikes present 95% confidence intervals based on standard errors clustered at grade one school level.

Figure 3: Event studies

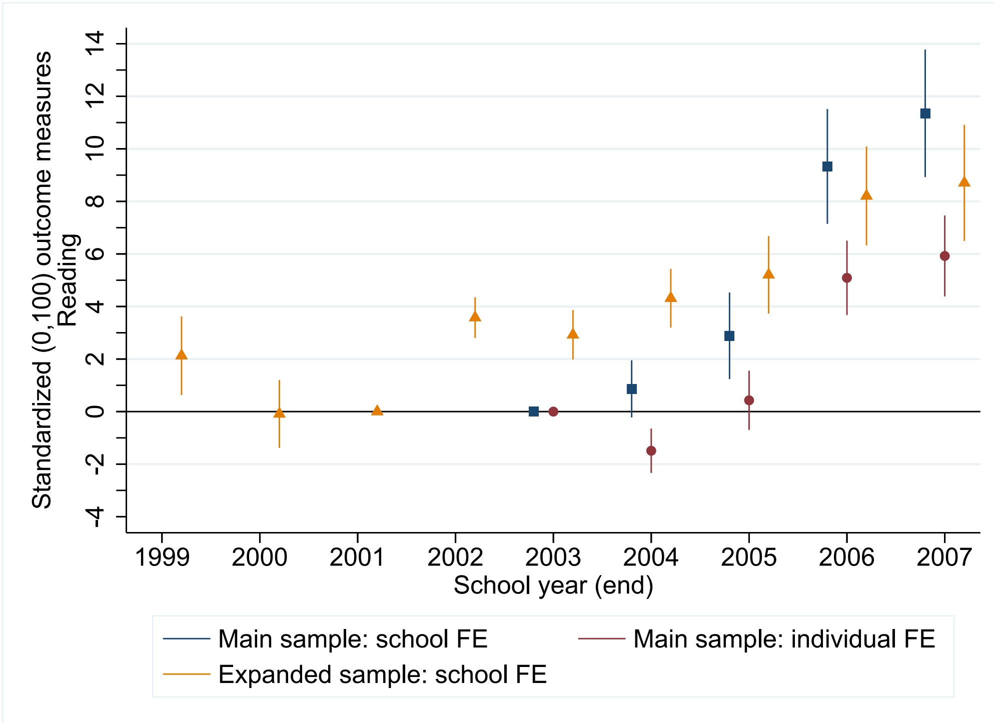
A. Averaged math + reading



B. Mathematics



C. Reading



Note: These figures present multiple event studies for averaged mathematics and reading test scores (panel A), mathematics test scores (panel B), and reading test scores (panel C). In each case we interact composite (PCA index) competition measure with school years. This measure is based on pre-reform competitive landscape. It is assigned at grade one school and grade six school in the main sample as in Table 2 while in the expanded sample it is assigned at contemporaneously observed school. Navy and maroon scatterplots present analyses based on modified data used in Figure 1 (main sample) while orange scatterplot uses a separate data set described in Section II.A. (expanded sample). It is based on all students attending Florida public schools between 1999 and 2007 who are born before September 1st 1994. Furthermore, we restrict the preferred matched sample used throughout the paper to students starting grade one prior to school year 2000/2001. In both cases it ensures that we only include students who started school in pre-reform period. In the main sample test scores are available to us only from school year 2002/2003 (first post-reform year) and we treat it as a reference group. In the expanded sample we also observe earlier test scores starting from school year 1998/1999, however, here we cannot execute individual fixed effects estimation strategy because unlike in the main sample students are not tested in every grade and subject. Thus, in the expanded sample we use last pre-reform school year, 2000/2001, as a reference period. Orange triangles present estimates from expanded sample with contemporaneous school and grade-by-school year FE. Navy squares present estimates from main sample with grade one school-by-school level FE and grade-by-school year FE. Maroon circles present estimates from main sample with student-by-school level FE and grade-by-school year FE. Additional controls in regressions with school rather than individual FE include dummies for gender, race, ethnicity, current free or reduced price lunch participation, month of birth and year of birth. Further details on these analyses are provided in Online Appendix B. Spikes present 95% confidence intervals based on standard errors clustered at grade one school level in the main sample and contemporaneous school level in expanded sample.

Tables

Table 1: Program expansion

School year	(1) Designated state funds	(2) Realized spending	(3) Number of scholarship enrollments	(4) Pctg. of K-12 public school enrollment	(5) Number of participating private schools	(6) Maximum annual family income allowed	(7) Maximum amount granted per student
2002/03	50,000,000	50,000,000	15,585	N/A	N/A	< 185% FPL	3,500
2003/04	88,000,000	40,000,000	11,550	N/A	924	< 185% FPL	3,500
2004/05	88,000,000	36,655,500	10,549	0.48	973	< 185% FPL	3,500
2005/06	88,000,000	46,745,482	15,123	0.55	895	< 185% FPL	3,500
2006/07	88,000,000	59,300,655	17,819	0.63	948	< 200% FPL	3,750
2007/08	88,000,000	73,450,691	21,493	0.83	933	< 200% FPL	3,750
2008/09	118,000,000	88,626,463	24,871	0.94	1,002	< 200% FPL	3,950
2009/10	118,000,000	106,049,940	28,927	1.18	1,033	< 200% FPL	3,950
2010/11	140,000,000	129,474,868	34,550	1.44	1,114	< 230% FPL	4,106
2011/12	175,000,000	147,481,308	40,248	1.92	1,216	< 230% FPL	4,011
2012/13	229,000,000	206,974,102	51,075	2.29	1,338	< 230% FPL	4,335
2013/14	286,000,000	274,495,570	59,822	2.57	1,429	< 230% FPL	4,880
2014/15	357,800,000	344,887,372	69,950	2.89	1,533	< 230% FPL	5,272
2015/16	447,000,000	418,693,458	78,664	3.43	1,602	< 230% FPL	5,677
2016/17	559,000,000	539,252,526	98,936	3.75	1,733	< 260% FPL	5,886
2017/18	698,000,000	641,024,651	108,098	3.60	1,818	< 260% FPL	7,208

Note: This table presents Florida Tax Credit (FTC) Scholarship Program expansion between school years 2002/03 and 2017/18. Column 1 shows total amount of tax credits which may be granted in given year; column 2 shows realized spending in the program; column 3 shows number of students enrolled through the scholarship program; column 4 shows percentage of K-12 students in the state of Florida participating in the voucher program; column 5 shows number of participating private schools; column 6 shows maximum annual family income allowed; and column 7 shows maximum amount of scholarship per student that can be awarded. Columns 1, 6 and 7 are based on Florida Statues 220.187 for years 2002/03 to 2009/10 and Florida Statues 1002.395 for years 2010/11 to 2017/18. Columns 2 to 5 are based on Florida Department of Education reports (Choice Facts, 2008; 2010; 2011; 2014; 2018) and our own calculations.

Table 2: Effects of voucher expansion by baseline competition measures

	(1) Math + Reading	(2) Mathematics	(3) Reading	(4) Suspensions	(5) Absences
			Panel A. Diversity		
Expansion × above median competition	4.233***	1.937***	6.539***	-0.504*	-0.265***
	(0.599)	(0.736)	(0.618)	(0.267)	(0.052)
			Panel B. Density		
Expansion × above median competition	5.293***	2.817***	7.566***	-1.109***	-0.258***
	(0.586)	(0.728)	(0.608)	(0.268)	(0.052)
			Panel C. Distance		
Expansion × above median competition	1.648***	-0.308	3.542***	-0.430*	-0.151***
	(0.590)	(0.712)	(0.622)	(0.261)	(0.052)
			Panel D. Houses of worship nearby		
Expansion × above median competition	3.917***	1.643**	5.966***	-1.428***	-0.223***
	(0.598)	(0.727)	(0.626)	(0.265)	(0.051)
Mean [SD] of Y	0.027 [93.104]	0.000 [100.000]	0.000 [100.000]	13.633 [34.314]	5.039 [5.786]
Observations	6,187,563	6,131,878	6,611,067	5,453,653	5,453,653
# children	1,222,165	1,222,912	1,223,799	1,228,461	1,228,461
			Panel E. Slots per grade		
Expansion × above median competition	6.064***	3.361***	8.684***	-1.749***	-0.290***
	(0.590)	(0.732)	(0.604)	(0.266)	(0.052)
			Panel F. Competitive Pressure Index		
Expansion × above median competition	5.111***	2.639***	7.389***	-1.282***	-0.281***
	(0.593)	(0.738)	(0.611)	(0.267)	(0.052)
Mean [SD] of Y	0.000 [93.085]	-0.038 [99.977]	-0.017 [99.984]	13.666 [34.349]	5.041 [5.790]
Observations	6,160,525	6,104,889	6,584,014	5,427,985	5,427,985
# children	1,221,023	1,220,753	1,223,123	1,225,713	1,225,713

Note: Sample is based on individual-level observations in grades 3 to 8 for students attending Florida public schools between 2002/03 to 2016/17 and born between 1992 and 2002. Each child has to be observed at least in grade 1 so that we can assign them school-level competition measures which are based on Figlio and Hart (2014); these are assigned to individuals for the schools they attend in grades 1 and 6. Thus, there are up to two values of competition observed for each individual. Expansion is measured at annual level between 2002/03 and 2016/17 as logarithm of number of scholarships awarded. Test scores are based on FCAT developmental scores for years 2000/2001 to 2013/2014 and on FSA developmental scores for years 2014/2015 to 2016/2017, and we standardize them in-sample by years and grade to have mean 0 and standard deviation of 1. These standardized scores are then multiplied by 100. Averaged mathematics and reading as well as mathematics test scores are available up to school year 2013/2014 while reading test scores are available up to school year 2016/2017. Suspensions (indicator for ever being suspended in a given year) and absences (absence rate in a given year net of suspension days) are measured for years 2002/03 to 2011/2012, and they are multiplied by 100. Each column represents a separate outcome variable. Competition measures are: number of denominational types represented (panel A); number of local private schools (panel B); miles to nearest private school competitor (panel C); number of churches, synagogues, and mosques (panel D); number of private school slots per grade (panel E); and principal components analysis competition index ("Competitive Pressure Index") based on five measures from panels A to E (panel F). Regression table presents interactions between competition measure (dummy for competition above median in the full sample of schools) and log of expansion measure, and all regressions include student-by-school level FE and grade-by-school year FE. School level is defined as indicator for grade 6 to 8 vs. 3 to 5. Standard errors are clustered at grade one school level. Point estimates marked ***, **, and * are statistically significant at the 1, 5, and 10 percent levels, respectively.

Table 3: Robustness of the preferred estimates: Measurement of competiton and expansion

	(1) Math + Reading	(2) Mathematics	(3) Reading	(4) Suspensions	(5) Absences
			Panel A. Baseline		
Expansion × above median competition	5.111***	2.639***	7.389***	-1.282***	-0.281***
	(0.593)	(0.738)	(0.611)	(0.267)	(0.052)
Mean [SD] of Y	0.000 [93.085]	-0.038 [99.977]	-0.017 [99.984]	13.666 [34.349]	5.041 [5.790]
Observations	6,160,525	6,104,889	6,584,014	5,427,985	5,427,985
			Panel B. Median split at student rather than school level		
Expansion × above median competition	5.340***	2.931***	7.653***	-1.470***	-0.260***
	(0.576)	(0.713)	(0.601)	(0.262)	(0.052)
Mean [SD] of Y	0.000 [93.085]	-0.038 [99.977]	-0.017 [99.984]	13.666 [34.349]	5.041 [5.790]
Observations	6,160,525	6,104,889	6,584,014	5,427,985	5,427,985
			Panel C. Continuous competition measure		
Expansion × competition	1.453***	0.620***	2.296***	-0.422***	-0.086***
	(0.157)	(0.196)	(0.161)	(0.068)	(0.014)
Mean [SD] of Y	0.000 [93.085]	-0.038 [99.977]	-0.017 [99.984]	13.666 [34.349]	5.041 [5.790]
Observations	6,160,525	6,104,889	6,584,014	5,427,985	5,427,985
			Panel D. Log funding expansion measure		
Expansion × above median competition	4.747***	2.547***	6.536***	-1.358***	-0.311***
	(0.555)	(0.695)	(0.552)	(0.268)	(0.052)
Mean [SD] of Y	0.000 [93.085]	-0.038 [99.977]	-0.017 [99.984]	13.666 [34.349]	5.041 [5.790]
Observations	6,160,525	6,104,889	6,584,014	5,427,985	5,427,985
			Panel E. Competition measure unweighted with elementary to middle school flows		
Expansion × above median competition	4.337***	1.955***	6.528***	-1.031***	-0.280***
	(0.574)	(0.722)	(0.588)	(0.250)	(0.048)
Mean [SD] of Y	-0.003 [92.865]	-0.002 [99.755]	-0.072 [99.844]	13.223 [33.874]	5.011 [5.706]
Observations	5,761,773	5,714,711	6,123,884	5,117,781	5,117,781
			Panel F. Weights based solely on pre-program transitions between grades 5 and 6		
Expansion × above median competition	5.633***	3.111***	8.209***	-1.333***	-0.228***
	(0.578)	(0.718)	(0.594)	(0.265)	(0.052)
Mean [SD] of Y	0.002 [93.084]	-0.036 [99.975]	-0.014 [99.984]	13.653 [34.336]	5.038 [5.793]
Observations	6,071,801	6,016,952	6,487,847	5,351,967	5,351,967
			Panel G. Competition measure assigned only to grade 1 school		
Expansion × above median competition	4.711***	2.236***	6.962***	-1.342***	-0.306***
	(0.608)	(0.741)	(0.637)	(0.268)	(0.052)
Mean [SD] of Y	-0.023 [93.064]	-0.064 [99.959]	-0.034 [99.967]	13.634 [34.315]	5.038 [5.785]
Observations	6,137,574	6,082,372	6,557,631	5,409,447	5,409,447

Note: Robustness checks based on estimates from panel F of Table 2. Panel A replicates the main result from panel F of Table 2; panel B uses the median split at individual rather than school level (i.e., we divide the sample into above vs below median competition based on school competition measures weighted with student population); panel C replaces dummy indicator for above median pre-reform competition with continuous measure; panel D replaces logarithm of number of scholarships expansion measure with logarithm of funding; panel E assigns the middle-school pre-policy competition measures based on the actual grade 6 (middle) school initially attended by each student; panel F generates expected competitive pressure measures for middle school-aged students using only pre-policy announcement flow between elementary and middle schools; and panel G assigns school competition only to grade one school rather than grade one and grade six schools. Outcome variables are averaged mathematics and reading test score (column 1), mathematics test score (column 2), reading test score (column 3), likelihood of being suspended (column 4), and absence rate (column 5). Standard errors are clustered at grade one school level. All outcome variables are multiplied by 100. Point estimates marked ***, **, and * are statistically significant at the 1, 5, and 10 percent levels, respectively.

Table 4: Robustness of the preferred estimates: Samples and econometric specifications

	(1) Math + Reading	(2) Mathematics	(3) Reading	(4) Suspensions	(5) Absences
			Panel A. Baseline		
Expansion × above median competition	5.111***	2.639***	7.389***	-1.282***	-0.281***
	(0.593)	(0.738)	(0.611)	(0.267)	(0.052)
Mean [SD] of Y	0.000 [93.085]	-0.038 [99.977]	-0.017 [99.984]	13.666 [34.349]	5.041 [5.790]
Observations	6,160,525	6,104,889	6,584,014	5,427,985	5,427,985
			Panel B. Including contemporaneous region-by-year FE		
Expansion × competition	3.602***	2.548***	4.094***	-0.679**	-0.245***
	(0.610)	(0.790)	(0.570)	(0.280)	(0.053)
Mean [SD] of Y	0.000 [93.085]	-0.038 [99.977]	-0.017 [99.984]	13.666 [34.349]	5.041 [5.790]
Observations	6,160,525	6,104,889	6,584,014	5,427,985	5,427,985
			Panel C. Limiting the sample to schools with at least one competitor within 5 miles		
Expansion × above median competition	4.271***	2.221***	6.131***	-1.129***	-0.193***
	(0.618)	(0.767)	(0.642)	(0.284)	(0.052)
Mean [SD] of Y	0.003 [93.124]	-0.042 [99.976]	-0.019 [99.984]	13.636 [34.317]	5.008 [5.781]
Observations	5,703,761	5,650,965	6,092,805	5,030,501	5,030,501
			Panel D. School years 2002/03 to 2011/12		
Expansion × above median competition	4.584***	2.644***	6.578***	-1.282***	-0.281***
	(0.589)	(0.738)	(0.615)	(0.267)	(0.052)
Mean [SD] of Y	-0.219 [92.915]	-0.040 [99.974]	-0.019 [99.981]	13.666 [34.349]	5.041 [5.790]
Observations	5,336,140	5,323,917	5,323,137	5,427,985	5,427,985
			Panel E. Balanced panel (6-years)		
Expansion × above median competition	5.463***	3.496***	6.933***	-1.427***	-0.364***
	(0.660)	(0.825)	(0.683)	(0.306)	(0.055)
Mean [SD] of Y	0.095 [92.933]	-0.037 [99.974]	-0.015 [99.986]	14.524 [35.234]	4.969 [5.551]
Observations	3,958,889	3,919,656	5,303,632	2,845,185	2,845,185
			Panel F. In sample standardized test scores		
Expansion × above median competition	4.825***	2.484***	7.165***		
	(0.586)	(0.733)	(0.603)	N/A	N/A
Mean [SD] of Y	0.061 [92.587]	0.190 [99.823]	-0.069 [99.805]		
Observations	5,756,691	5,756,691	5,756,691		
			Panel G. In population standardized test scores		
Expansion × above median competition	4.913***	2.729***	7.097***		
	(0.555)	(0.700)	(0.573)	N/A	N/A
Mean [SD] of Y	4.669 [89.064]	4.657 [96.371]	4.681 [95.610]		
Observations	5,756,691	5,756,691	5,756,691		

Note: Robustness checks based on estimates from panel F of Table 2. Panel A replicates the main result from panel F of Table 2; panel B adds region-by-year fixed effects based on contemporaneously attended school (we divide Florida into six regions based on Florida Association of Counties classification; these are Northwest, Northeast, West Central, East Central, Southwest, and Southeast); panel C limits the initial sample to only schools with at least one competitor within 5 miles; panel D restricts the sample to school years 2002/03 to 2011/12 where we observe all five outcomes; panel E restricts the sample to 6-year panel of observations starting with grade 3 and within school years available for a given variable; and panels F and G restrict the sample to school years 2002/03 to 2012/13 where we observe test scores that are standardized for the full population of Florida students - panel F presents our in-sample standardization while panel G population-level standardization. Outcome variables are averaged mathematics and reading test score (column 1), mathematics test score (column 2), reading test score (column 3), likelihood of being suspended (column 4), and absence rate (column 5). Standard errors are clustered at grade one school level. All outcome variables are multiplied by 100. Point estimates marked ***, **, and * are statistically significant at the 1, 5, and 10 percent levels, respectively.

Table 5: Alternative explanations: Effects of voucher program expansion on peer composition and class size

	(1)	(2)	(3)	(4)	(5)	(6)
	School-level peer effects					
	Math + Reading	Math	Reading	Suspensions	Absences	School-level class size
Expansion × above median competition	-0.399	-0.430	0.163	0.122*	-0.005	-0.221***
	(0.290)	(0.288)	(0.271)	(0.071)	(0.010)	(0.061)
Mean of Y	-2.2	-2.3	-1.6	14.4	5.1	16.5
Observations	37,880	37,875	44,685	31,334	31,334	32,340

Note: Columns 1 to 5 present the effects of voucher program expansion on potential school-level peer effects where the dependent variables are predicted rather than actual test scores (columns 1 to 3), suspensions (column 4), and absences (column 5). Column 6 presents the effects of voucher program expansion on school-level class size information. Each regression is based on cells aggregated to school in grade one by school-level by school year level. Table displays coefficient of interest which is interaction between the preferred competition measure from panel F of Table 2 (dummy for competition above median in the full sample of schools) and log of expansion measure, and each regression includes school in grade one by school-level fixed effects and year fixed effects. Predicted test scores and disciplinary outcomes are based on predicted values from a regression of actual test scores or disciplinary outcomes on year and month of birth dummies, gender, birth weight, maternal years of education dummies, gestational age dummies, marital status, mother's place of birth, race, ethnicity, maternal age at birth, prior number of births to mother, month prenatal care, complications of labor and delivery, abnormal conditions at birth, congenital anomalies, maternal health problems and Medicaid paid birth. R-squares from these regressions are 0.240, 0.205, 0.214, 0.100, and 0.071 for averaged math and reading, math, reading, suspensions, and absences respectively. These are then aggregate at grade one school by school level by year level. Data on class size for school years 2006/2007 to 2016/2017 are based on reports provided by Florida Department of Education (http://www.fldoe.org/finance/budget/class-size/class-size-reduction-averages.stml) separately for grades PK to 3, 4 to 8 and 9 to 12. For each school and year we weight these reported class sizes according to actual grades served e.g., if school is serving grades PK to 8 then we compute school-level class size as $CS = 0.5CS_{PK-3} + 0.5CS_{4-8} + 0CS_{9-12}$. Standard errors are clustered at grade one school level. Point estimates marked ***, **, and * are statistically significant at the 1, 5, and 10 percent levels, respectively.

Table 6: School versus district competition and the role of time-varying characteristics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Baseline		Additional time-varying control variables				
	Panel A. Math+Reading (N = 6,160,525; # children = 1,221,023; mean of Y = 0.000; SD of Y = 93.085)						
Expansion × above median competition (school)	5.340***	2.020***	1.779***	2.081***	2.017***	2.268***	2.078***
	(0.576)	(0.684)	(0.679)	(0.687)	(0.684)	(0.682)	(0.678)
Expansion × above median competition (district)		5.371***	5.217***	5.196***	5.388***	5.017***	4.688***
		(0.700)	(0.695)	(0.703)	(0.699)	(0.700)	(0.699)
	Panel B. Mathematics (N = 6,104,889; # children = 1,220,753; mean of Y = -0.038; SD of Y = 99.977)						
Expansion × above median competition (school)	2.931***	0.327	-0.097	0.354	0.324	0.506	0.109
	(0.713)	(0.833)	(0.829)	(0.835)	(0.833)	(0.833)	(0.829)
Expansion × above median competition (district)		4.230***	4.051***	4.150***	4.252***	3.978***	3.746***
		(0.843)	(0.844)	(0.844)	(0.843)	(0.848)	(0.849)
	Panel C. Reading (N = 6,584,014; # children = 1,223,123; mean of Y = -0.017; SD of Y = 99.984)						
Expansion × above median competition (school)	7.653***	3.838***	3.712***	3.925***	3.846***	4.068***	4.034***
	(0.601)	(0.759)	(0.754)	(0.763)	(0.759)	(0.753)	(0.752)
Expansion × above median competition (district)		6.141***	5.876***	5.914***	6.146***	5.717***	5.207***
		(0.771)	(0.767)	(0.775)	(0.770)	(0.766)	(0.766)
	Panel D. Suspensions (N = 5,427,985; # children = 1,225,713; mean of Y = 13.666; SD of Y = 34.349)						
Expansion × above median competition (school)	-1.470***	-0.534**	-0.292	-0.511*	-0.533**	-0.653**	-0.391
	(0.262)	(0.269)	(0.270)	(0.269)	(0.269)	(0.269)	(0.269)
Expansion × above median competition (district)		-1.519***	-1.083***	-1.566***	-1.515***	-1.335***	-0.932***
		(0.278)	(0.288)	(0.279)	(0.278)	(0.277)	(0.289)
	Panel E. Absences (N = 5,427,985; # children = 1,225,713; mean of Y = 5.041; SD of Y = 5.790)						
Expansion × above median competition (school)	-0.260***	-0.183***	-0.176***	-0.173***	-0.183***	-0.180***	-0.164***
	(0.052)	(0.057)	(0.057)	(0.057)	(0.057)	(0.057)	(0.057)
Expansion × above median competition (district)		-0.125**	-0.134**	-0.145***	-0.123**	-0.126**	-0.156***
		(0.055)	(0.056)	(0.055)	(0.055)	(0.056)	(0.056)
District-level magnets & charters			X				X
District-level average salaries				X			X
School-level class sizes					X		X
Peer-effects in all domains						X	X

Note: All regressions include student-by-school level FE and grade-by-school year FE. Column 1 replicates the results from panel B in Table 3. Column 2 presents estimates from a horse-race between competition measured at school and at school district level. In each case we interact competition index (dummy for competition above median in the full sample of schools) and log of expansion measure. District level competition is student-weighted average of the school-level competition collapsed at school district in grade one by school level by school year level. Columns 3 to 7 further add control variables that are time-varying (at annual level) at either school or district level. These are assigned based on grade one school or school district by year level. Column 3 controls for district-level number of charter schools per 1000 students and number of magnet schools per 1000 students. Column 4 controls for district-level average public school teachers salaries. Column 5 controls for school level average class size. Column 6 controls for school-level peer effects (based on predicted outcomes) in averaged math and reading test scores, math test scores, reading test scores, suspensions, and absences. Column 7 includes all controls from columns 3 to 6 jointly. Information on class size is available for years 2006/2007 to 2016/2017, information on charter and magnet schools is available for years 2002/2003 to 2016/2017, information for average salaries is available for years 2004/2005 to 2016/2017, and information on predicted potential peer effects is available for years 2002/2003 to 2013/2014 for math and averaged math and reading, for years 2002/2003 to 2016/2017 for reading, and for years 2002/2003 to 2011/2012 for suspensions and absences. To maintain constant sample size we perform following imputations for variables with missing values due to differential coverage of years: (1) if available impute mean school level values and (2) if school-level information not available impute sample average. Standard errors are clustered at grade one school level. Point estimates marked ***, **, and * are statistically significant at the 1, 5, and 10 percent levels, respectively.

Online Appendix (not for publication)

Online Appendix A. Composition of Students in Florida Public Schools

Here we explore the extent to which competitive pressures affect the composition of students ever appearing in Florida public schools. Recall that, during this paper's study period, voucher participants must either have spent the previous year in Florida public schools or been entering kindergarten or first grade, and the latter would never be observed in the sample. Since the empirical strategy in our paper relies on student fixed effects, would-be peers never observed in the public schools will not contribute to changes in students' schooling environments and thus our coefficients should remain unbiased. Who is in the sample, however, could affect the external validity and interpretation of our results.

To address this question, we analyze whether the voucher program's roll-out affected which children were present in the birth records but not in the school records. To do so, we proxy for the zip code of birth's level of competition pressure in any given year by re-weighting our measures of competition pressure (introduced in Section II.C.) for birth cohorts expected to enter first grade after the program started (September 1995 to December 2002 births) with empirically observed flows of students born in any given zip code to all possible grade one (G1) schools as observed for birth cohorts entering schooling before the program started (January 1994 to August 1995 births). Table A6 shows how the voucher program roll-out affected the probability that a child would ultimately appear in the Florida public school data, both overall (panel A) and stratified for samples with a given characteristic (e.g. child of high school dropout mother in panel D or child of immigrant mother in panel J).

We observe that, unsurprisingly, as the program expanded fewer students born in communities with greater competitive pressures ended up in public schools, meaning that locales with more competition pressure straightaway were the places sending more children to private schools as the voucher program expanded. These results are concentrated in the set of children whose births were funded by Medicaid and those with relatively poorly-educated mothers which makes sense since the program supports vouchers for low-income families. At the same time, as the program expanded, Black children and the children of immigrants were disproportionately likely to never appear in public schooling in communities with greater competitive pressures. It is also worth highlighting that the estimates from Table A6 are very modest in magnitude with effect sizes not exceeding 6 percent of sample mean. In summary, the voucher program led to a public school sector that is modestly more affluent with higher parental education. These composition changes, albeit small in magnitude, underscore the importance of gauging heterogeneity in the effects of competition pressure, as we report in Section III.C. On the one hand, such selection could reduce the estimated competitive effects if higher-SES individuals are less responsive to the effects of competition. On the other hand, it could increase them if there is complementarity between school-level student ability and competitive pressure. Assuming that student fixed effects account for time-invariant characteristics related to these selection processes and there are no time-varying covariates differentially correlated with scale up in more vs. less competitive areas, in neither case the estimates will be biased. We view these results as additional contribution to prior literature that, due to data limitations, was not able to examine selection processes of that sort. Nonetheless, this could affect external validity of our findings to a small degree.

Online Appendix B. Analyses Presented in Figure 3

Event studies in Figure 3 are based on two samples. First is a subset of our matched birth-school records restricted to school years 2002/03 to 2006/07 and students who started grade one (G1) school in the last pre-policy year 2000/01. These students are generally born prior to September 1st 1994. We execute two regression analyses in this sample based on school fixed effects (equation 1) and based on individual fixed effects (equation 2):

$$Y_{isglt} = \beta_t \sum_{t=2004}^{2007} Year_t \times Competition_{sl} + \gamma_{sl} + \delta_{gt} + \pi X_{it} + \varepsilon_{isglt} \tag{1}$$

$$Y_{isglt} = \beta_t \sum_{t=2004}^{2007} Year_t \times Competition_{sl} + \theta_{il} + \delta_{gt} + \varepsilon_{isglt} \tag{2}$$

where Y_{isglt} captures an outcome measure for student i who entered the FLDOE data in grade one (G1) school s, observed in grade g corresponding to academic stage (elementary or middle school) l in year t, θ_{il} is a student-by-stage fixed effect, γ_{sl} is a school-by-stage fixed effect, and δ_{gt} is a grade-by-year fixed effect. Control variables (X_{it}) in equation 1 include gender, racial and ethnic categories, free and reduced price lunch status (time varying) as well as birth year and birth month dummies. School year 2002/03 serves as a reference category in this event study. Robust standard errors (ε_{isglt}) are clustered at student's G1 school level.

Our second sample is based on all public school students who were tested between 1998/99 and 2006/07 school years and born prior to September 1st 1994. Unlike in the pervious sample here students are not tested in each subject in each grade and therefore we cannot execute our individual fixed effects strategy. We estimate the following equation:

$$\begin{aligned} Y_{isgt} =& \beta_t^{pre} \sum_{t=1999}^{2000} Year_t \times Competition'_s + \beta_t^{post} \sum_{t=2002}^{2007} Year_t \times Competition'_s + \\ & \omega_s + \delta_{gt} + \pi X_{it} + \varepsilon_{isgt} \end{aligned} \tag{3}$$

where Y_{isgt} captures an outcome measure for student i in school s, observed in grade g in year t, ω_s is a school fixed effect, and δ_{gt} is a grade-by-year fixed effect. Control variables (X_{it}) in equation 3 include gender, racial and ethnic categories, free and reduced price lunch status (time varying) as well as birth year and birth month dummies. School year 2000/01, last pre-policy year, serves as a reference category in this event study. Robust standard errors (ε_{isgt}) are clustered at school level.

In equation 3 we denote variable $Competition'_s$ with a prime because we assign it to currently attended school rather than to grade one (G1) school (weighted with middle school flows). Thus, our variation here is defined at school-by-year level rather than at G1 school-by-stage-by-year level. We are forced to make this adjustment because in the expanded data we can only observe students in grades in which they are being tested and testing in Florida commences in grade 3 at the earliest. Furthermore, until school year 2000/2001 students were only tested in grades four, five, and eight. Therefore, we do not know either the school a student was attending in grade one or their transitions between elementary and middle school stages.

Appendix Tables

Table A1: Construction of Competitive Pressure Index based on principal components analysis

	(1)	(2)	(3)	(4)
	Grades 1 to 5		Grades 6 to 8	
	First component	Second component	First component	Second component
Diversity	0.474	0.144	0.474	0.169
Density	0.499	-0.214	0.494	-0.215
Distance	0.295	0.903	0.331	0.873
Number of houses of worship	0.460	-0.229	0.455	-0.271
Number of slots	0.477	-0.257	0.463	-0.301
Eigenvalue	3.614	0.813	3.829	0.733

Note: This table reports the results of a principal components analysis of number of denominational types represented (diversity), number of local private schools (density), miles to nearest private school competitor (distance), number of churches, synagogues, and mosques, and number of private school slots per grade. The eigenvectors associated with the first (columns 1 and 3) and second (columns 2 and 4) components are reported separately for grades 1 to 5 and 6 to 8, as well as their associated eigenvalues.

Table A2: Descriptive statistics

	(1)	(2)	(3)	(4)
			Competition index	
	All births	Empirical sample	Above median	Below median
		Panel A. Sociodemographic characteristics		
White	55.4	50.9	37.3	68.0
African-American	19.4	23.3	30.2	14.5
Hispanic	22.8	23.9	30.5	15.6
Mother foreign born	23.2	23.0	29.7	14.5
Male	51.2	51.1	51.1	51.1
Mother HS dropout	20.9	24.9	25.0	24.6
Mother HS graduate	58.8	60.4	60.7	60.1
Mother college graduate	20.2	14.7	14.3	15.2
Mother age at birth	27.1	26.6	26.6	26.5
Parents married at birth	64.9	59.2	54.6	65.1
Ever on FRPL	N/A	71.8	75.8	66.6
		Panel B. Competition measures		
Diversity	N/A	5.1	6.8	2.9
Density		15.6	24.0	5.0
Distance		-1.9	-1.0	-3.0
Number of houses of worship		143.0	207.9	61.0
Number of slots		2.9	4.7	0.7
Competition index (PCA)		0.3	1.6	-1.5
		Panel C. Outcomes		
Math+reading score	N/A	0.0	-4.4	5.6
Math score		0.0	-4.1	5.0
Reading scores		0.0	-4.8	6.0
Likelihood suspended		13.7	14.3	12.9
Absence rate		5.0	5.0	5.2
Maximum # observations	2,028,798	6,971,914	3,890,161	3,081,753
Maximum # children	2,028,798	1,255,084	755,254	609,646

Note: Panel A presents means of sociodemographic variables (all indicator variables multiplied by 100); panel B presents means of competition measures with distance reverse coded (more positive values indicate higher competition); panel C presents outcome variables (all multiplied by 100). Column 1 presents characteristics of full sample of births between 1992 and 2002; column 2 presents characteristics of our preferred empirical sample for school years between 2002/03 to 2016/17; columns 3 and 4 divide sample from column 2 into two mutually exclusive categories based on median of the PCA competition index.

Table A3: Heterogeneity in the effects of voucher expansion: Socioeconomic status measures

	(1) Math + Reading	(2) Mathematics	(3) Reading	(4) Suspensions	(5) Absences
	Panel A. Ever on free or reduced price lunch				
Expansion × above median competition	6.504***	3.660***	9.187***	-1.946***	-0.431***
	(0.624)	(0.770)	(0.666)	(0.338)	(0.063)
Mean [SD] of Y	-21.409 [89.396]	-21.402 [97.315]	-21.010 [96.385]	17.305 [37.829]	5.587 [6.314]
Observations	4,362,211	4,324,143	4,696,426	3,803,417	3,803,417
	Panel B. Never on free or reduced price lunch				
Expansion × above median competition	3.971***	2.781***	4.860***	-0.917***	-0.144***
	(0.759)	(0.970)	(0.771)	(0.254)	(0.055)
Mean [SD] of Y	51.933 [80.552]	51.840 [86.435]	52.216 [89.054]	5.147 [22.095]	3.763 [4.044]
Observations	1,798,314	1,780,746	1,887,588	1,624,568	1,624,568
	Panel C. Mother high school dropout				
Expansion × above median competition	5.256***	2.956***	7.504***	-2.159***	-0.338***
	(0.762)	(0.953)	(0.852)	(0.479)	(0.096)
Mean [SD] of Y	-43.261 [88.539]	-42.496 [97.764]	-43.506 [95.362]	21.774 [41.271]	6.460 [7.175]
Observations	1,504,461	1,492,865	1,609,399	1,334,914	1,334,914
	Panel D. Mother high school graduate				
Expansion × above median competition	5.675***	2.900***	8.297***	-1.113***	-0.337***
	(0.597)	(0.752)	(0.620)	(0.287)	(0.052)
Mean [SD] of Y	1.324 [87.359]	0.979 [94.291]	1.434 [94.682]	12.655 [33.247]	4.870 [5.415]
Observations	3,739,944	3,709,186	3,989,457	3,304,238	3,304,238
	Panel E. Mother college graduate				
Expansion × above median competition	3.163***	1.863	3.812***	-0.776***	-0.060
	(0.902)	(1.146)	(0.940)	(0.292)	(0.057)
Mean [SD] of Y	65.640 [82.760]	65.993 [88.952]	65.155 [91.140]	4.182 [20.017]	3.356 [3.792]
Observations	916,120	902,838	985,158	788,833	788,833

Note: Specifications are based on those in panel F of Table 2 with the baseline sample split by child's free or reduced price lunch history (panels A and B) and maternal education (panels C to E). Outcome variables are averaged mathematics and reading test scores (column 1), mathematics test scores (column 2), reading test scores (column 3), likelihood of being suspended (column 4), and absence rate (column 5). All outcomes are multiplied by 100. Standard errors are clustered at grade one school level. Point estimates marked ***, **, and * are statistically significant at the 1, 5, and 10 percent levels, respectively.

Table A4: Heterogeneity in the effects of voucher expansion: Demographic characteristics

	(1) Math + Reading	(2) Mathematics	(3) Reading	(4) Suspensions	(5) Absences
			Panel A. White mother		
Expansion × above median competition	4.682***	3.803***	5.269***	-1.116***	-0.228***
	(0.677)	(0.860)	(0.657)	(0.295)	(0.064)
Mean [SD] of Y	22.894 [88.067]	22.368 [93.933]	23.256 [96.293]	10.145 [30.193]	5.257 [5.730]
Observations	3,156,514	3,132,289	3,348,248	2,815,187	2,815,187
			Panel B. African-American mother		
Expansion × above median competition	4.319***	3.693***	5.277***	-0.121	-0.021
	(0.844)	(1.122)	(0.834)	(0.650)	(0.094)
Mean [SD] of Y	-48.816 [88.572]	-49.807 [98.517]	-47.479 [94.430]	25.486 [43.578]	5.311 [6.483]
Observations	1,414,642	1,403,085	1,513,010	1,248,570	1,248,570
			Panel C. Hispanic mother		
Expansion × above median competition	5.488***	1.557	9.126***	-2.452***	-0.170**
	(0.989)	(1.274)	(1.093)	(0.423)	(0.084)
Mean [SD] of Y	-6.380 [88.165]	-4.972 [95.255]	-7.608 [95.334]	10.518 [30.678]	4.441 [5.190]
Observations	1,469,512	1,451,296	1,593,065	1,261,615	1,261,615
			Panel D. Mother born in the U.S.		
Expansion × above median competition	3.607***	2.142***	4.744***	-0.898***	-0.288***
	(0.591)	(0.759)	(0.564)	(0.296)	(0.059)
Mean [SD] of Y	-1.281 [93.900]	-1.947 [100.685]	-0.724 [100.883]	14.870 [35.579]	5.374 [6.008]
Observations	4,739,008	4,701,848	5,049,671	4,198,015	4,198,015
			Panel E. Foreign born mother		
Expansion × above median competition	2.690***	-0.945	6.103***	-1.631***	-0.127*
	(0.926)	(1.223)	(1.018)	(0.390)	(0.071)
Mean [SD] of Y	4.272 [90.182]	6.359 [97.294]	2.311 [96.932]	9.558 [29.402]	3.907 [4.805]
Observations	1,421,517	1,403,041	1,534,343	1,229,970	1,229,970

Note: Specifications are based on those in panel F of Table 2 with the baseline sample split by race/ethnicity (panels A to C) and maternal immigration status (panels D and E). Outcome variables are averaged mathematics and reading test scores (column 1), mathematics test scores (column 2), reading test scores (column 3), likelihood of being suspended (column 4), and absence rate (column 5). All outcomes are multiplied by 100. Standard errors are clustered at grade one school level. Point estimates marked ***, **, and * are statistically significant at the 1, 5, and 10 percent levels, respectively.

Table A5: School and district competition measures split by the median

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Math + Reading		Mathematics		Reading		Suspensions		Absences	
Expansion × school above and	5.293***	5.142***	3.401***	2.862**	6.818***	7.023***	-0.761	-0.620	-0.202***	-0.168**
district below median competition	(1.047)	(1.022)	(1.285)	(1.259)	(1.054)	(1.045)	(0.464)	(0.463)	(0.075)	(0.074)
Expansion × school below and	7.953***	7.090***	6.665***	5.913***	8.486***	7.548***	-1.698***	-1.112***	-0.140*	-0.160*
district above median competition	(0.928)	(0.937)	(1.143)	(1.150)	(1.062)	(1.063)	(0.360)	(0.364)	(0.084)	(0.084)
Expansion × school and district	7.441***	6.809***	4.599***	3.891***	10.019***	9.283***	-2.057***	-1.327***	-0.308***	-0.320***
both above median competition	(0.649)	(0.643)	(0.814)	(0.815)	(0.660)	(0.652)	(0.308)	(0.329)	(0.059)	(0.061)
Mean of Y	0.000		-0.038		-0.017		13.666		5.041	
SD of Y	93.085		99.977		99.984		34.349		5.790	
F-statistic	48.9	41.0	15.2	11.5	80.2	70.2	16.2	5.7	9.2	9.3
# children	1,221,023		1,220,753		1,223,123		1,225,713		1,225,713	
Observations	6,160,525		6,104,889		6,584,014		5,427,985		5,427,985	
Time varying controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes

Note: All regressions include student-by-school level FE and grade-by-school year FE based on modified estimates from panel F of Table 2. This table presents results for interactions between voucher expansion and the following four competition measures: school- and district-level competition below median (reference category), school-level competition below while district-level competition above median, school-level competition above while district-level competition below median, and both school- and district-level competition above median. Columns 1 and 2 present results for averaged mathematics and reading test scores, columns 3 and 4 present results for mathematics test scores, columns 5 and 6 present results for reading test scores, columns 7 and 8 present results for suspensions, and columns 9 and 10 present results for absences. All outcomes are multiplied by 100. Odd-numbered columns present results without any additional controls while even-numbered columns present results controlling for time-varying school- and district-level variables akin to column 7 in Table 6. F statistics tests for joint equality of the interaction terms. Information on class size is available for years 2006/2007 to 2016/2017, information on charter and magnet schools is available for years 2002/2003 to 2016/2017, information for average salaries is available for years 2004/2005 to 2016/2017, and information on predicted potential peer effects is available for years 2002/2003 to 2013/2014 for math and averaged math and reading, for years 2002/2003 to 2016/2017 for reading, and for years 2002/2003 to 2011/2012 for suspensions and absences. To maintain constant sample size we perform following imputations for variables with missing values due to differential coverage of years: (1) if available impute mean school level values and (2) if school-level information not available impute sample average. Standard errors are clustered at grade one school level. Point estimates marked ***, **, and * are statistically significant at the 1, 5, and 10 percent levels, respectively.

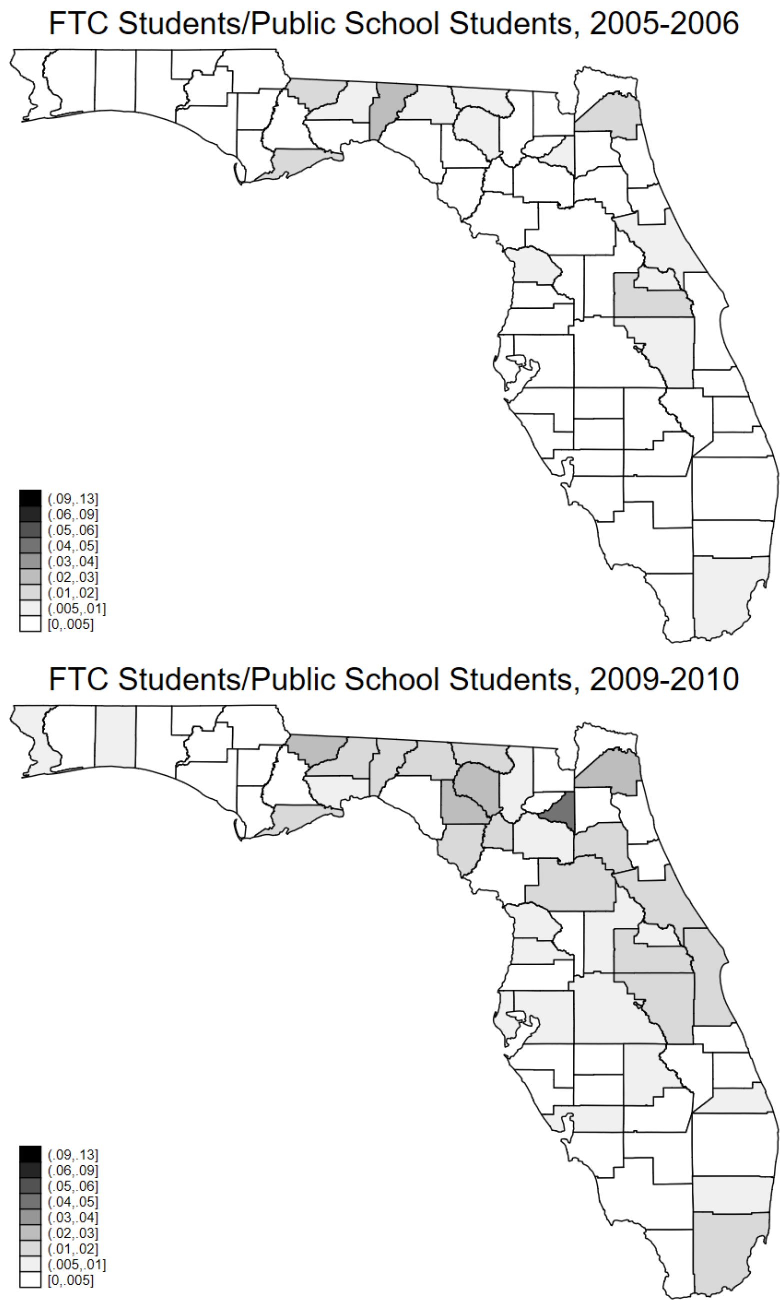
Table A6: Voucher program expansion and likelihood of being observed in matched birth-public school records

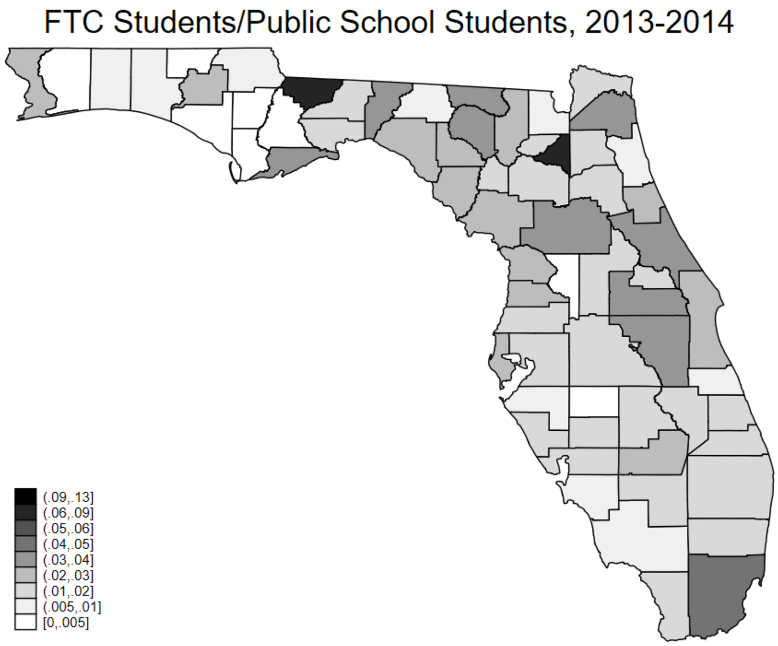
	(1)	(2)	(3)	(4)	(5)	(6)
	Outcome: Probability of being matched to school records (*100)					
Competition measures	PCA index	Diversity	Density	Distance	Houses of worship	Slots
	Panel A: Overall (Mean = 79.8; N = 1,279,009)					
Expansion × above median weighted competition	-1.055***	-1.218***	-1.238***	-1.364***	-0.963***	-1.185***
	(0.356)	(0.351)	(0.354)	(0.385)	(0.359)	(0.356)
Implied % effect	-1.3	-1.5	-1.6	-1.7	-1.2	-1.5
	Panel B. Non-Medicaid paid birth (Mean = 74.7; N = 709,570)					
Expansion × above median weighted competition	-0.115	-0.301	-0.258	-1.090**	-0.290	-0.125
	(0.505)	(0.494)	(0.506)	(0.525)	(0.502)	(0.506)
Implied % effect	-0.2	-0.4	-0.3	-1.5	-0.4	-0.2
	Panel C. Medicaid paid birth (Mean = 86.1; N = 569,438)					
Expansion × above median weighted competition	-2.342***	-2.433***	-2.435***	-1.687***	-2.036***	-2.506***
	(0.385)	(0.384)	(0.384)	(0.415)	(0.390)	(0.378)
Implied % effect	-2.7	-2.8	-2.8	-2.0	-2.4	-2.9
	Panel D. Mother high school dropout (Mean = 87.1; N = 250,565)					
Expansion × above median weighted competition	-1.451***	-1.714***	-1.581***	-1.114*	-1.370**	-1.428***
	(0.536)	(0.528)	(0.534)	(0.597)	(0.547)	(0.529)
Implied % effect	-1.7	-2.0	-1.8	-1.3	-1.6	-1.6
	Panel E. Mother high school graduate (Mean = 82.2; N = 746,382)					
Expansion × above median weighted competition	-0.769**	-0.863**	-0.818**	-1.256***	-0.686*	-0.765*
	(0.389)	(0.381)	(0.387)	(0.409)	(0.388)	(0.391)
Implied % effect	-0.9	-1.0	-1.0	-1.5	-0.8	-0.9
	Panel F. Mother college graduate (Mean = 66.9; N = 282,062)					
Expansion × above median weighted competition	-0.189	-0.397	-0.421	-0.641	-0.322	-0.683
	(0.853)	(0.817)	(0.834)	(0.776)	(0.773)	(0.834)
Implied % effect	-0.3	-0.6	-0.6	-1.0	-0.5	-1.0
	Panel G. White, non-Hispanic, non-immigrant (Mean = 77.4; N = 640,193)					
Expansion × above median weighted competition	0.034	-0.083	-0.265	-0.704	-0.232	-0.053
	(0.433)	(0.435)	(0.436)	(0.452)	(0.440)	(0.433)
Implied % effect	0.0	-0.1	-0.3	-0.9	-0.3	-0.1
	Panel H. Black, non-Hispanic, non-immigrant (Mean = 89.5; N = 213,720)					
Expansion × above median weighted competition	-2.230***	-2.566***	-2.374***	-2.439***	-1.351**	-2.255***
	(0.597)	(0.573)	(0.590)	(0.609)	(0.620)	(0.578)
Implied % effect	-2.5	-2.9	-2.7	-2.7	-1.5	-2.5
	Panel I. Hispanic, non-immigrant (Mean = 81.4; N = 107,344)					
Expansion × above median weighted competition	-0.742	-0.639	0.095	-0.493	-0.031	-1.252
	(1.049)	(1.031)	(1.038)	(0.856)	(1.022)	(1.053)
Implied % effect	-0.9	-0.8	0.1	-0.6	0.0	-1.5
	Panel J. Immigrant mother (Mean = 77.6; N = 317,752)					
Expansion × above median weighted competition	-4.296***	-4.289***	-4.231***	-2.830***	-3.508***	-4.335***
	(0.815)	(0.810)	(0.825)	(0.872)	(0.830)	(0.779)
Implied % effect	-5.5	-5.5	-5.5	-3.6	-4.5	-5.6
	Panel K: Females (Mean = 80.2; N = 624,677)					
Expansion × above median weighted competition	-1.290***	-1.313***	-1.466***	-1.326***	-0.998**	-1.221***
	(0.419)	(0.414)	(0.418)	(0.444)	(0.426)	(0.420)
Implied % effect	-1.6	-1.6	-1.8	-1.7	-1.2	-1.5
	Panel L. Males (Mean = 79.5; N = 654,332)					
Expansion × above median weighted competition	-0.816*	-1.118**	-1.006**	-1.400***	-0.910**	-1.125**
	(0.444)	(0.438)	(0.445)	(0.460)	(0.442)	(0.438)
Implied % effect	-1.0	-1.4	-1.3	-1.8	-1.1	-1.4

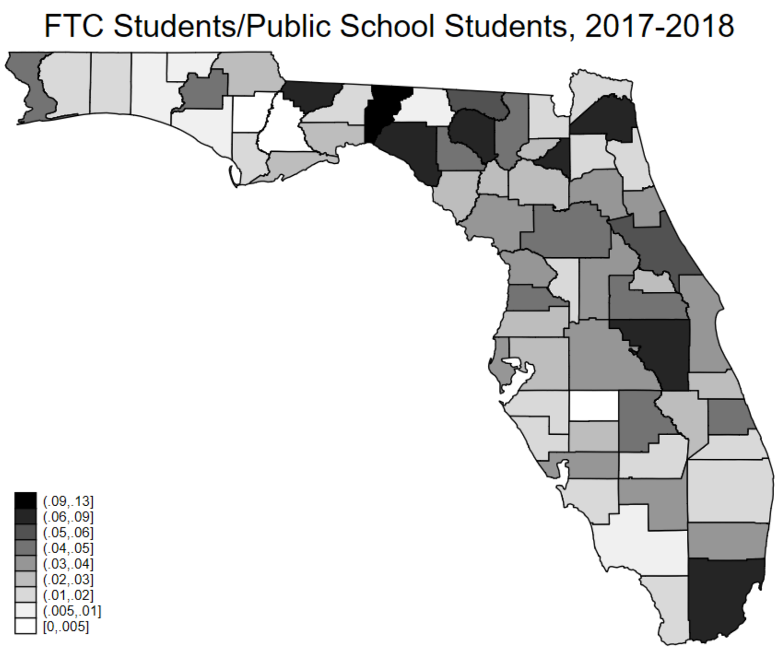
Note: This table presents estimates where the outcome variable is an indicator for being matched between birth and school records multiplied by 100. Panel A presents overall probability while panels B to L present results for various subsamples. Independent variable of interest is an interaction between annual voucher expansion and weighted competition at zip code level. Analysis is based on data for cohorts entering grade one after the program stated (September 1995 and later) while weights are created based on pre-program grade one cohorts (January 1994 to August 1995). Weighting is based on observed flows of individuals born in a given zip code to all possible schools. Models further include zip code level and year fixed effects. Standard errors are clustered at zip code level. Additional details on this analysis are provided in Online Appendix A. Point estimates marked ***, **, and * are statistically significant at the 1, 5, and 10 percent levels, respectively.

Appendix Figures

Figure A1: Spatial and time variation in voucher utilization



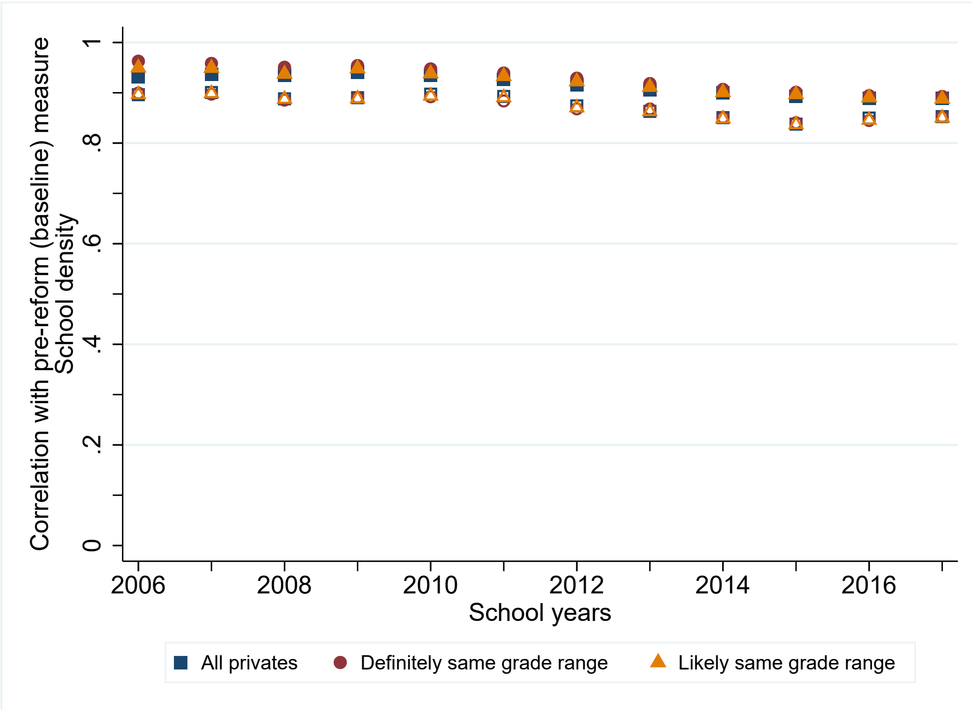




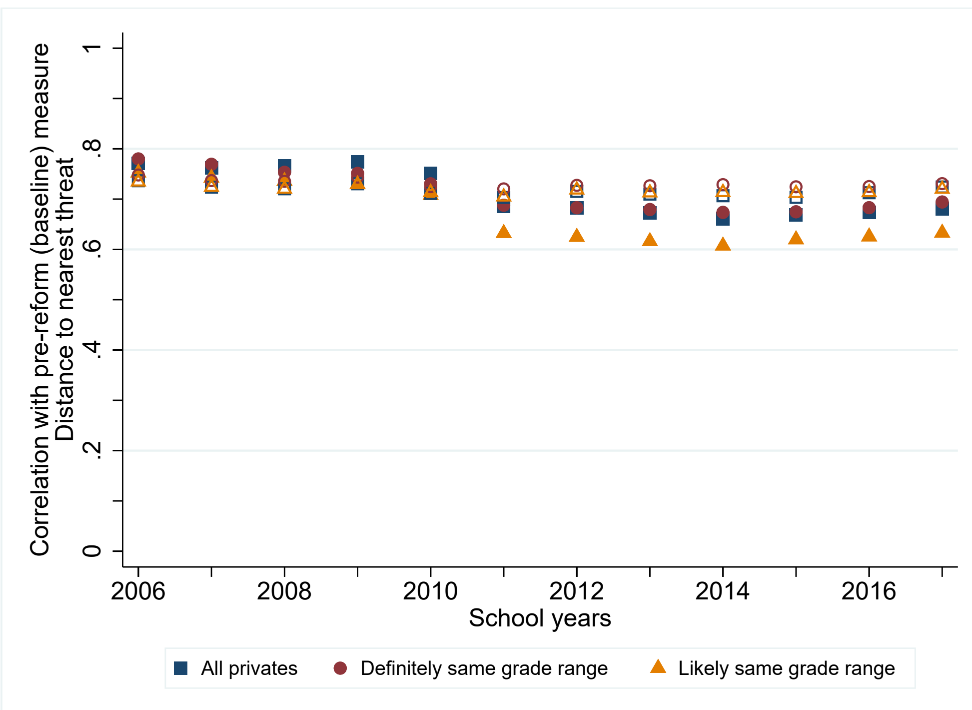
Note: District-level enrollment figures in the FTC program were drawn from quarterly reports produced by the Florida Department of Education (http://www.fldoe.org/schools/school-choice/k-12-scholarship-programs/ftc/quarterly-reports.stml). FTC enrollment figures for each district were taken from September reports, and were standardized by the number of K-12 students reported in NCES Common Core of Data reports. Students in certain types of specialized schools (special education, vocational education, or adult schools) in the NCES data were dropped.

Figure A2: Correlations in competition measures

A. Density



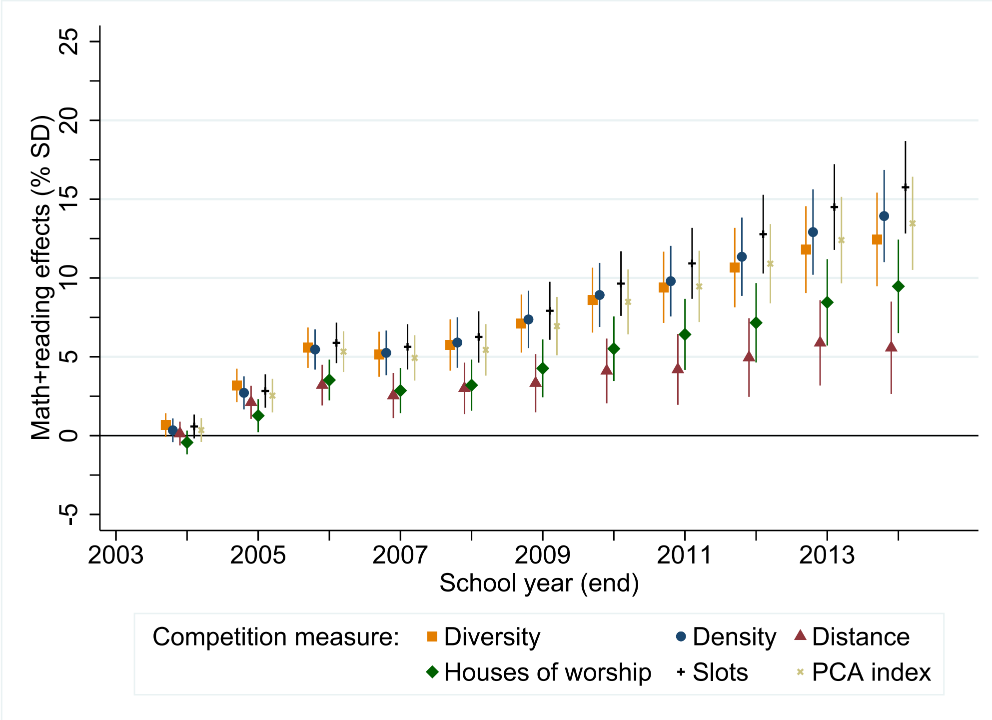
B. Distance



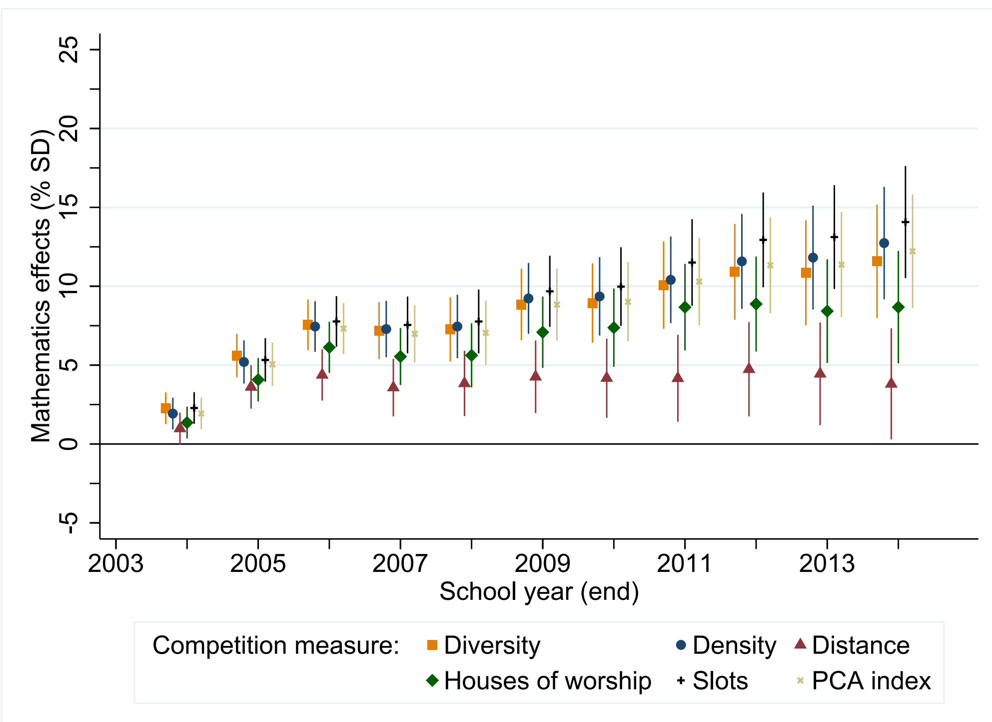
Note: These figures present Pearson correlation coefficients over time for two of our competition measures: school density (panel A) and distance to nearest competitor (panel B). In that we correlate competition measure defined for each year with our initial pre-program competition measure defined in spring semester of school year 1999/2000. There are six series of coefficients presented in each graph depending on how we define competing schools. These include all private schools (navy squares), private schools where we are certain that they are serving the same grade range as public school in question (maroon circles), and private schools where we are quite certain ("likely") that they are serving the same grade range as public school in question (orange triangle). We define being quite certain ("likely") if (a) we see evidence that they definitely do serve same grades based on FLDOE or (b) we can match the FLDOE Private School Directory data to NCES data at a high level of confidence in a fuzzy match (> 85% of similarity) of district and school name and see evidence in NCES that the school serves that grade level. Furthermore, solid markers consider all private schools within the defined categories while hollow markers only consider private school participating in Florida Tax Credit Program.

Figure A3: Effects of voucher expansion over school years

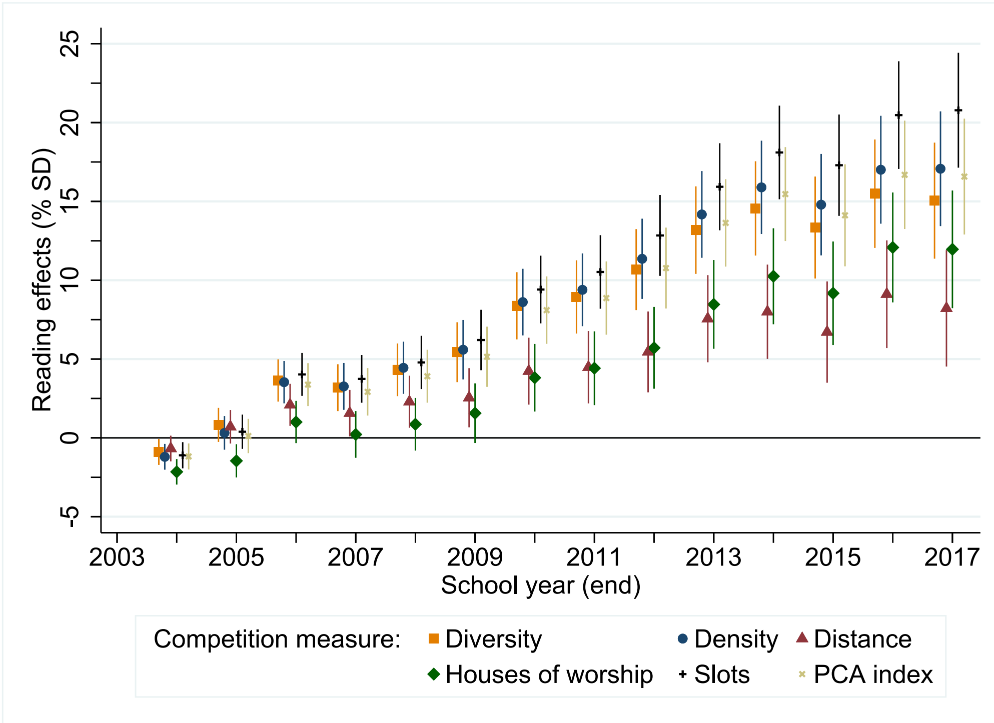
A. Averaged math + reading



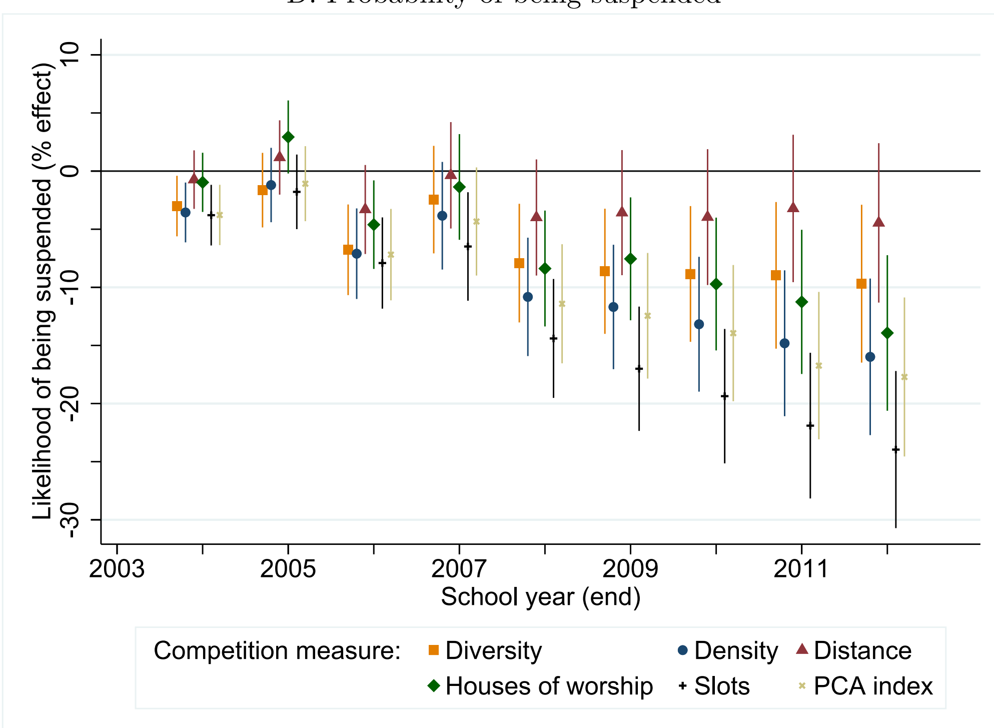
B. Mathematics



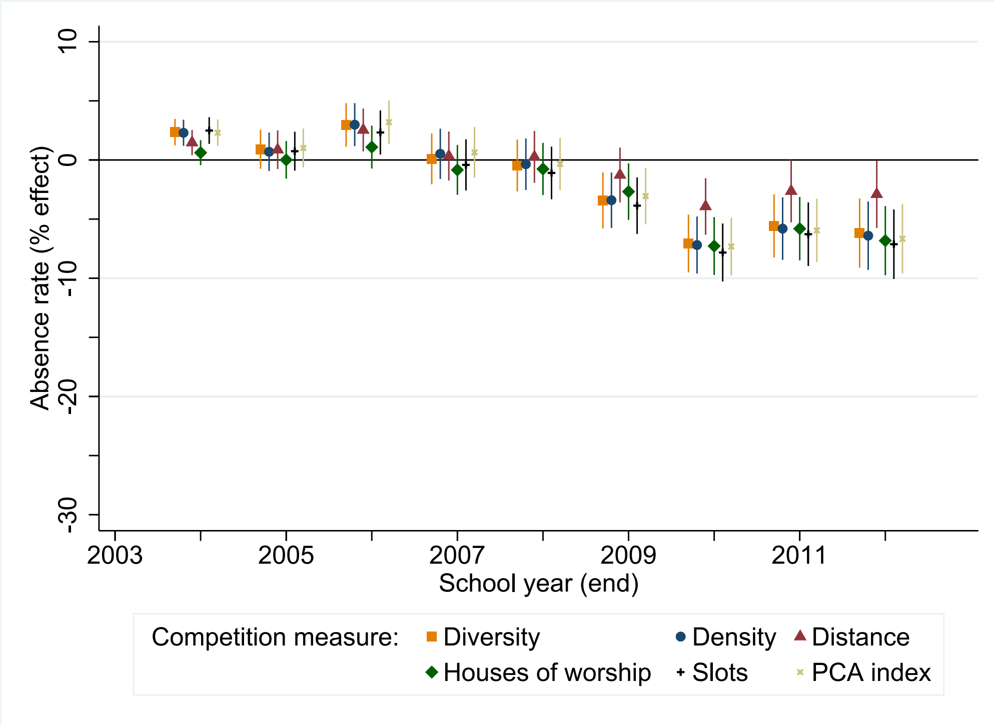
C. Reading



D. Probability of being suspended

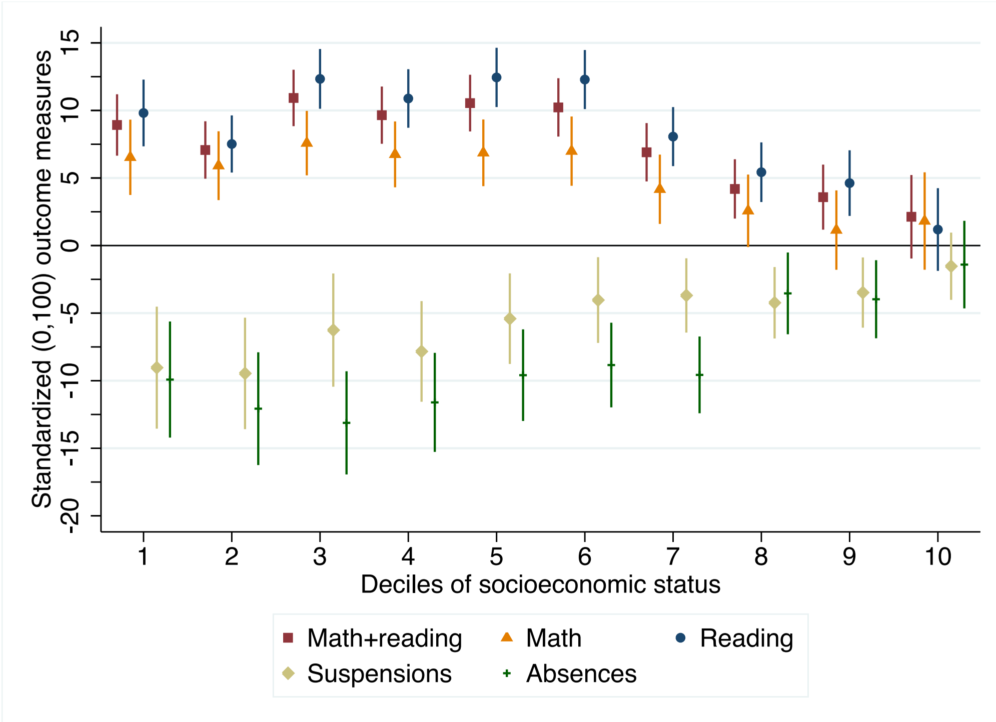


E. Absence rate



Note: These figures plot estimates from the specifications estimated in panels A to F of Table 2 where instead of interaction between competition measures and log number of scholarships we plot competition measures interacted with school years, and with baseline omitted year 2002/2003. Outcomes are averaged test scores in mathematics and reading (panel A); mathematics test scores (panel B); reading test scores (panel C); likelihood of being suspended (panel D); and absence rate (panel E). Competition measures are: number of denominational types represented (orange square); number of local private schools (navy circle); miles to nearest private school competitor (maroon triangles); number of churches, synagogues, and mosques (green diamonds); number of private school slots per grade (black pluses); and composite index of all five measures (khaki exes). Spikes present 95% confidence intervals based on standard errors clustered at grade one school level.

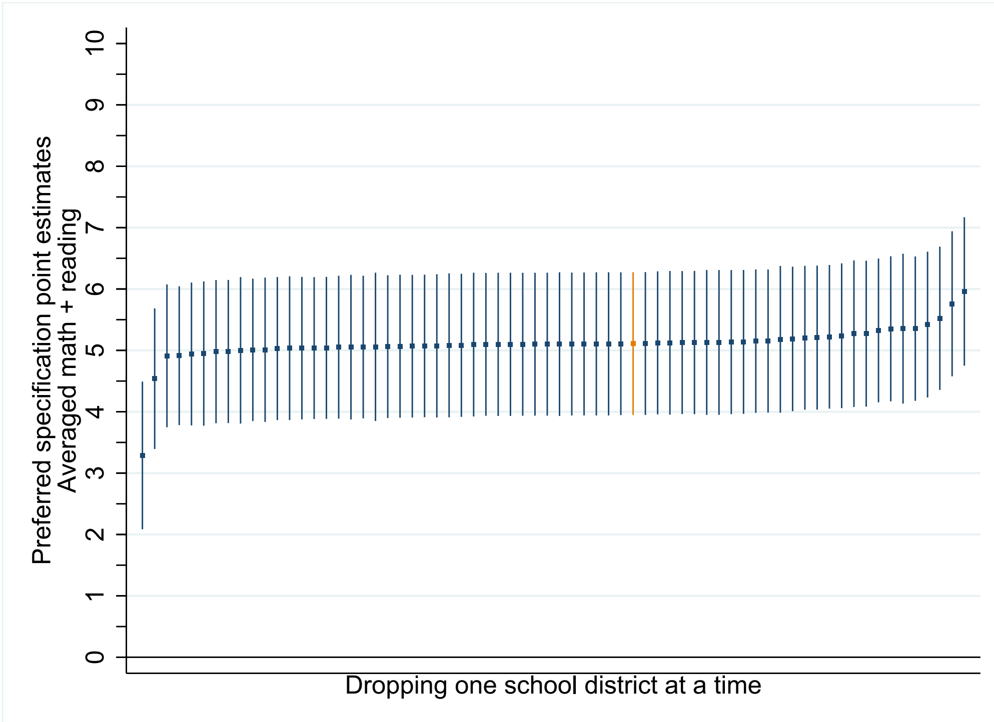
Figure A4: Effects of voucher expansion: Heterogeneity by socioeconomic status index



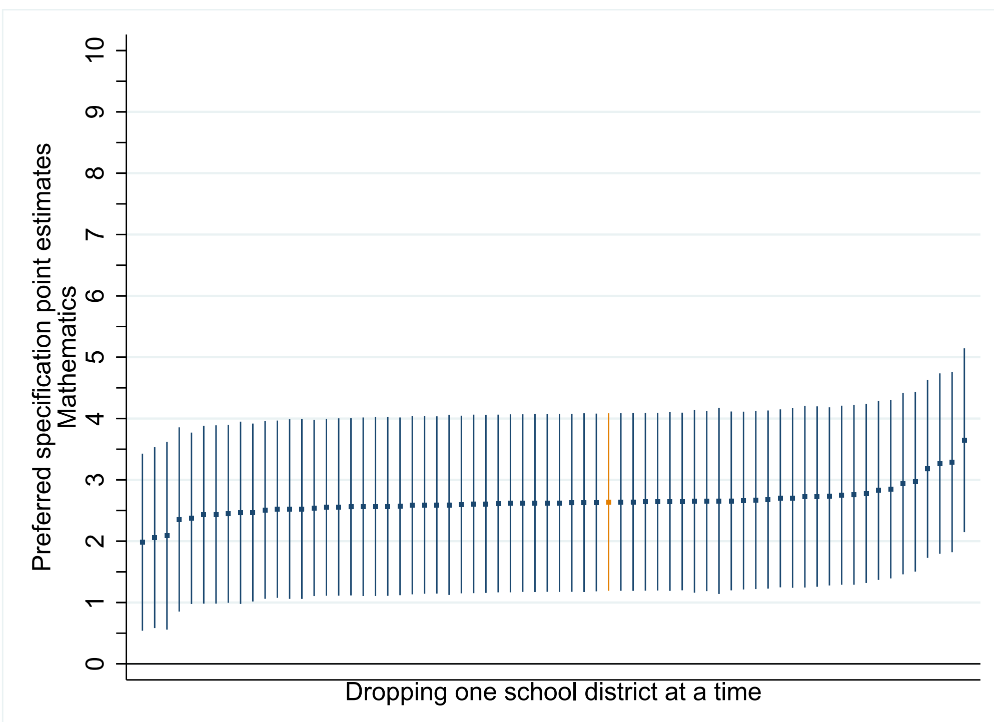
Note: This figure plots heterogeneity estimates for the main specification estimated in panel F of Table 2. These are computed separately for each outcome and each decile of socioeconomic status distribution. SES index is computed as first component from Principal Components Analysis (PCA) of maternal years of education, marital status, maternal age at birth, indicator for Medicaid paid birth, and zip code neighborhood income at the time of birth. Sample is restricted to births between 1994 and 2002. Outcomes are averaged test scores in mathematics and reading (maroon squares); mathematics test scores (orange triangles); reading test scores (navy circles); likelihood of being suspended (khaki diamonds); and absence rate (green pluses). Each outcome variable is standardized in its empirical sample to have mean zero and standard deviation of 100. Spikes present 95% confidence intervals based on standard errors clustered at grade one school level.

Figure A5: Robustness: Drop one school district at a time

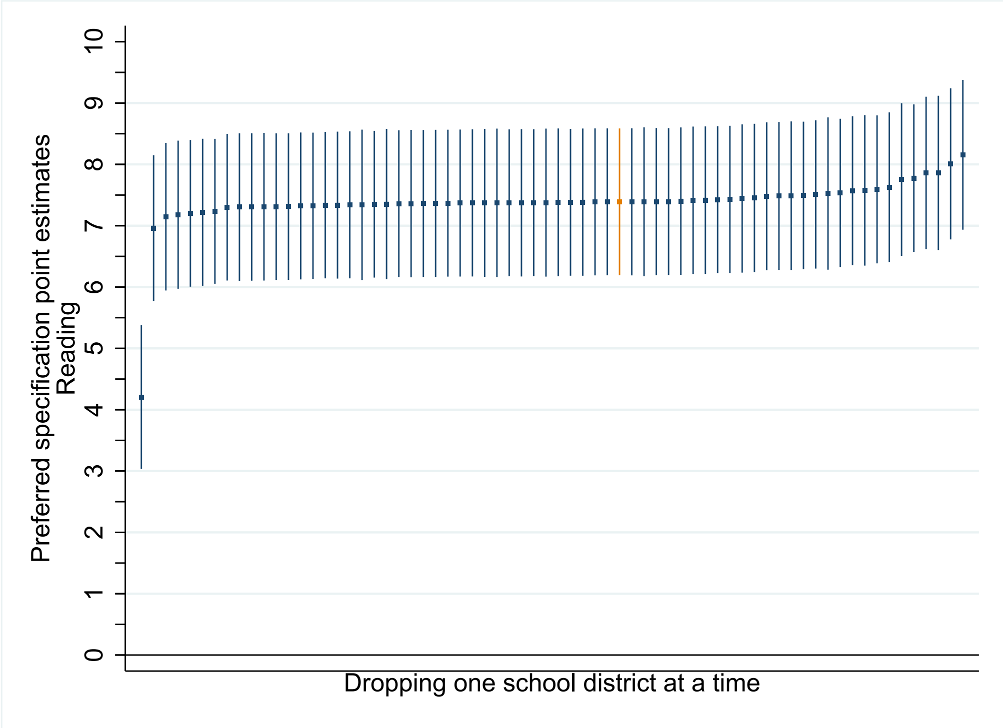
A. Averaged math + reading



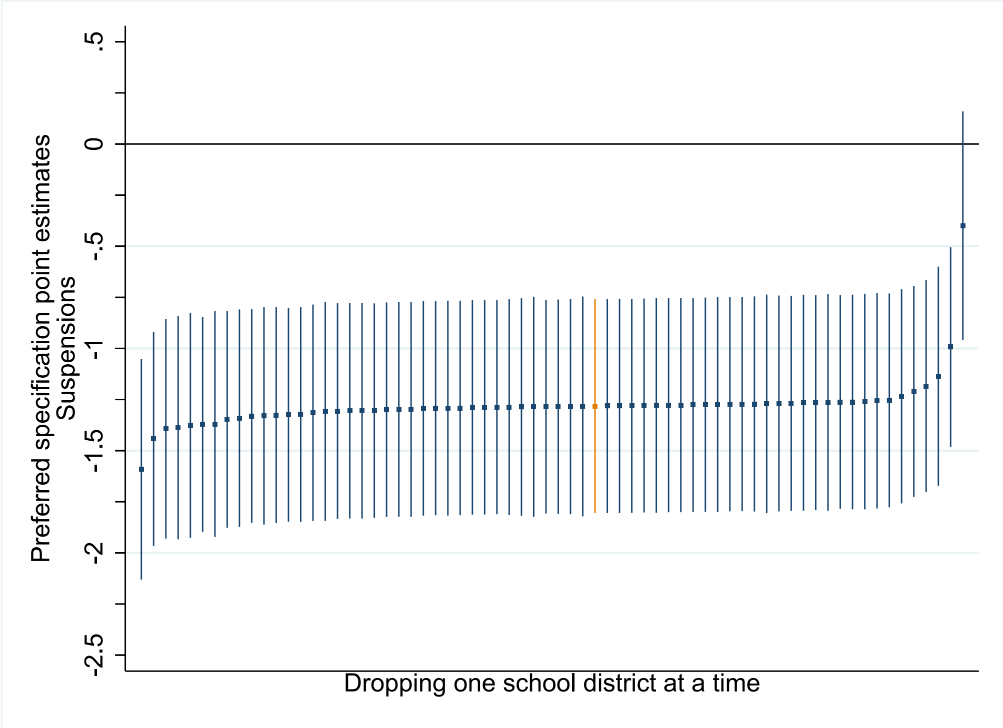
B. Mathematics



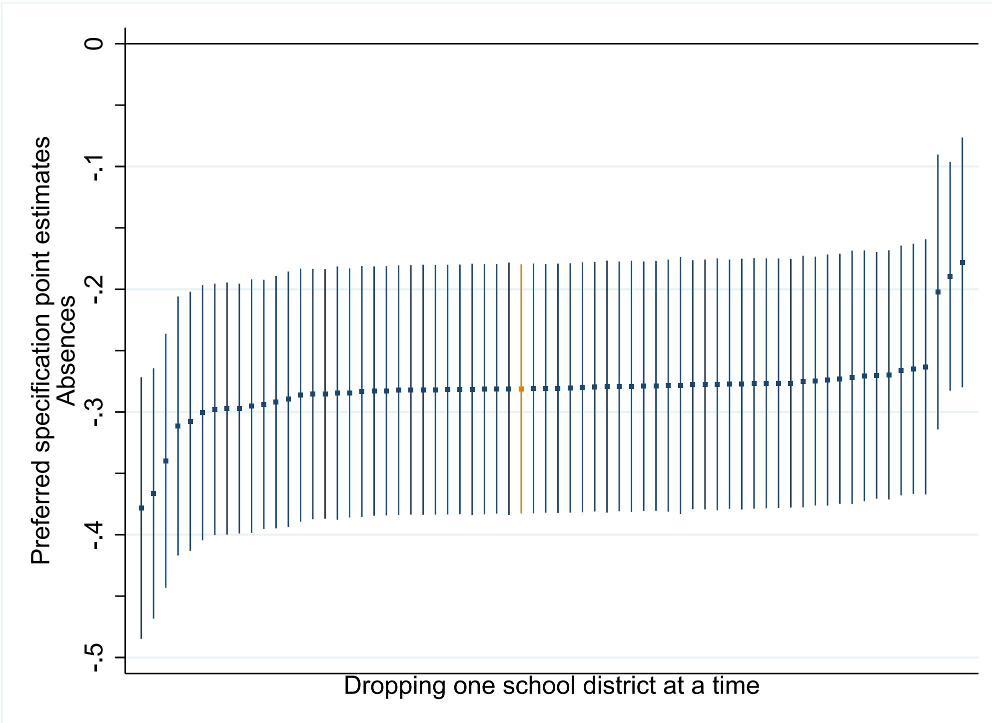
C. Reading



D. Probability of being suspended



E. Absence rate



Note: These figures plot estimates from analyses presented in panel F of Table 2 where we drop one school district at a time. Outcome variables are averaged mathematics and reading test score (panel A), mathematics test score (panel B), reading test score (panel C), likelihood of being suspended (panel D), and absence rate (panel E). Point estimates are ordered from the smallest to the largest and orange-highlighted estimate comes from a full sample as in Table 2. Spikes present 95% confidence intervals based on standard errors clustered at grade one school level.