American Economic Journal: Economic Policy 2023, 15(4): 255–294
https://doi.org/10.1257/pol.20210710

Effects of Maturing Private School Choice Programs on Public School Students†

By David N. Figlio, Cassandra M. D. Hart, and Krzysztof Karbownik*

Using a rich dataset that merges student-level school records with birth records, and leveraging a student fixed effects design, we explore how a Florida private school choice program affected public school students' outcomes as the program matured and scaled up. We observe growing benefits (higher standardized test scores and lower absenteeism and suspension rates) to students attending public schools with more preprogram private school options as the program matured. 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* H75, I21, I22, I28)

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 Washington, D.C., 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, and Urquiola 2017; Urquiola 2016). On the other hand, public school students could be harmed by private school choice programs

*Figlio: University of Rochester and National Bureau of Economic Research (email: david.figlio@rochester.edu); Hart: School of Education, University of California, Davis (email:cmdhart@ucdavis.edu); Karbownik: Department of Economics, Emory University, and National Bureau of Economic Research (email: krzysztof.karbownik@emory.edu). Erzo F.P. Luttmer was coeditor for this article. We are grateful to the Florida Departments of Education and Health for providing deidentified, matched data used in this analysis. Figlio acknowledges support from the National Science Foundation (grant #0338740), National Institute of Child Health and Human Development (grant #7R01HD054637), Bill and Melinda Gates Foundation (grant #OPP1029884), Institute for Education Sciences (CALDER grant #R305C120008), and Shelter Hill Foundation for assistance in building the dataset and/or conducting this research. We acknowledge support from the Smith Richardson Foundation (grant #2019-2082) in conducting this research. We thank Michael Dinerstein, Stefanie Fischer, Umut Özek, and seminar participants at the CESifo Economics of Education Meeting and University of New South Wales for helpful comments, suggestions, and resources. 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 nor those of our funders. All errors are our own.

†Go to https://doi.org/10.1257/pol.20210710 to visit the article page for additional materials and author disclosure statement(s) or to comment in the online discussion forum.

if the programs drain resources from the public schools or if the choice-induced sorting of students and teachers disadvantages those remaining in public schools (Epple, Romano, and 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, one 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 this is not a foregone conclusion: perhaps schools might begin to respond more when educators see that the program is not fly by night. A mature, sustained program may have different effects than the introduction of a program due to an increased sense of program permanence. Furthermore, 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 and Naylor 2020). Some 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. Thus, public schools may have relatively muted responses to the introduction of school choice programs in early years as they wait to see whether the programs will be sustained. Relatedly, 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 students' outcomes. 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.

At the same time, nearly every paper written in the US 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 mature and approach some longer-run steady state. The weight of the US 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 and Hart 2014; Figlio and Karbownik 2016; Greene and Winters 2007; Hoxby 2003; Rouse et al. 2013—see Urquiola 2016; Epple, Romano, and Urquiola 2017; and Egalite and Wolf 2016 for recent overviews of this literature). Similarly, Gilraine, Petronijevic, and Singleton (2021) provide evidence on the impacts of short-run (over the first two years) expansion of public school choice (charter) programs.

These studies generally focus on the very immediate short-run effects, evaluating the first one to four years after the initial introduction of school choice programs, when both the pros and cons of the program may be constrained due to the small

[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, and 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.

number of initial participants. They do not, however, consider what happens when the choice program matures, or as the number of students using private school vouchers grows to encompass a sizable fraction of the overall student body.[2] 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 maturing private school choice programs help or harm public schools.

Our paper complements this prior work and examines how the effects of a major statewide voucher program—the largest of its kind—changed over a time frame of over 15 years in the third-largest state in the United States. The program we investigate grew during our 15-year study period almost sevenfold from its original size and currently serves a participant population that is on average nearly 4 percent the size of the K–12 student population in Florida.[3] In some districts, the participation rate is over 10 percent. 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, a novel approach in the literature on school competition. Because the program expanded over time, it is challenging to disentangle possible maturation effects (e.g., resulting from an increased sense of program permanence) from the effects of program growth per se. However, to shine some light on this issue, we use several approaches to capture program growth and maturity, including time since program implementation, growth in the number of students participating, growth in program funding, growth in the total share of K–12 students enrolling in the voucher program, and growth in the number of participating private schools.

Empirically, we exploit differences in the initial competitive landscape faced by different schools—using measures of voucher competition introduced by Figlio and Hart (2014)—as well as aforementioned 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 growing and maturing 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

[2] While there are several studies in the international context that address the effects of programs that serve a larger share of the total student population, they generally look at slightly different questions, either covering total effects of private school competition on students in both the public and private sector (rather than focusing specifically on students who remain in the public school system—e.g., Hsieh and Urquiola 2006 in Chile and Böhlmark and Lindahl 2015 in Sweden), or looking at the early years of the voucher program rather than looking at dynamics over time (e.g., Sandström and Bergström 2005 in Sweden). Böhlmark and Lindahl (2015) also consider the effects on students in public schools, although they cannot measure outcomes at the individual level. Furthermore, Muralidharan and Sundararaman (2015) provide an experimental evaluation of the total effects of school choice in Andhra Pradesh, but they can only study short-run effects and consider sample sizes that are orders of magnitude smaller than those we examine in this paper. Many of the international voucher programs also vary significantly from the Florida context in ways that may provide less competitive pressure; for instance, Florida private schools are generally not allowed to practice selective admissions of applying students, unlike many major international programs (e.g., New Zealand or Chile; see Epple, Romano, and Urquiola 2017). Finally, Florida's voucher program is currently the largest of its kind in the United States in terms of both student enrollment and number of participating schools.

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

share—prepolicy fractions of private schools that we consider the most likely competitors to public schools—in the shift-share quasi-experimental research design (Bartik 1991) that we employ. We find evidence—described below—that our quasi-experimental setting satisfies the exogeneity-of-shares identifying assumptions recently detailed by Goldsmith-Pinkham, Sorkin, and Swift (2020).

We find that as public schools are more exposed to private school choice, their students experience increasing benefits as the program matures. In particular, higher levels of private school choice exposure are associated with lower rates of suspensions and absences, and with higher standardized test scores in reading and math. These results are not uniform: In our extensive heterogeneity analysis, we find that the public school students most positively affected by increased exposure to private school choice are comparatively low–socioeconomic status (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 the 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. We do find some evidence that higher levels of private school choice lead to changes in the characteristics of teachers remaining in public schools, but these factors do not appear to mediate the test score and behavioral gains. Thus, in our view, the increase in competitive pressure as the program matures and grows is the most plausible channel for the estimated gains in test scores and behavior.

I. Florida Context

We focus on the competitive effects of the Florida Tax Credit (FTC) scholarship program as it matured over a 15-year period. Announced in spring 2001, the FTC program provides dollar-for-dollar tax credits to corporations that donate to nonprofit 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). During the study period covered by this paper, in order to receive a scholarship for 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 2002–2003, 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, and families may supplement the scholarship as necessary to meet tuition bills. Initially, eligibility was restricted to students with a family income below 185 percent of the federal poverty line (see Florida Statute 220.187 (Florida Government 2001), and subsequent amendments), or \$49,025 for a family of four in 2021 dollars.

The program has expanded along several dimensions since 2002–2003, its first year of operation. Table 1 charts the expansion of the program in terms of the

Table 1 – Voucher Program Expansion

School year	Designated state funds (1)	Realized spending (2)	Number of scholarship enrollments (3)	Pctg. of K–12 public school enrollment (4)	Number of participating private schools (5)	Maximum annual family income allowed (6)	Maximum amount granted per student (7)
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

Notes: This table presents FTC scholarship program expansion between school years 2002–2003 and 2017–2018. Column 1 shows total amount of tax credits that may be granted in given year; column 2 shows realized spending in the program; column 3 shows the number of students enrolled through the scholarship program; column 4 shows the percentage of K–12 students in the state of Florida participating in the voucher program; column 5 shows the number of participating private schools; column 6 shows the maximum annual family income allowed; and column 7 shows the maximum amount of scholarship per student that can be awarded. Columns 1, 6, and 7 are based on Florida Statutes 220.187 for the years 2002–2003 to 2009–2010 and Florida Statutes 1002.395 for the years 2010–2011 to 2017–2018. Columns 2 to 5 are based on Florida Department of Education reports (Florida Department of Education 2009, 2013, 2018) and our own calculations.

Sources: Florida Government (2001, 2010); Florida Department of Education (2009, 2013, 2018)

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 2017–2018, 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 percent and 3.5 percent 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 2003–2004 to 1,818 schools participating in 2017–2018. 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, expanding the eligibility pool over time. In the 2006–2007 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 percent of the federal poverty line. In 2010–2011, this continuing-student eligibility threshold was raised to 230 percent of the federal poverty line. In 2016–2017, the eligibility restrictions were changed to allow partial scholarships for entering students with incomes between 185 percent and 260 percent of the federal poverty line.

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 $3,500 in 2002–2003 to $7,208 in 2017–2018. While state-level data on private school tuition are scarce, based on our calculations using national statistics from the National Center for Education Statistics, the $7,208 scholarship would cover 100 percent of the average tuition of a Catholic elementary school and about 83 percent of the average tuition of a non-Catholic religious elementary school.[4] These figures are, respectively, 67 percent and 40 percent of the average tuitions of the same institutional types at the high school level.

Given the substantial changes 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, online Appendix Figure A1 shows the district-level ratio of FTC participation to K–12 public school enrollment over four school years: 2005–2006, 2009–2010, 2013–2014, and 2017–2018. Here, darker gray 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 2005–2006 and 2017–2018. No district has more than 3 percent participation in 2005–2006, while roughly 45 percent of districts (representing nearly 60 percent of students) have participation rates in excess of 3 percent in 2017–2018.

Second, there is substantial spatial variation in FTC participation, particularly in later years. While 8 of Florida's 67 districts have less than 1 percent participation in 2017–2018, another 8 have participation rates of over 6 percent, while the district with the highest participation rate reaches over 10 percent. 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 in our empirical analyses. At the same time, 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 percent or more) in 2017–2018 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.

[4] Specifically, figures from the 2011–2012 Schools and Staffing Survey suggest that the national average tuition at Catholic, other religious, and nonsectarian 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 percent inflation rate, this suggests that a $7,208 scholarship for an elementary (high school) student would have covered 124 percent (67.5 percent), 83.1 percent (40.0 percent), or 36.4 percent (26.3 percent) of the average tuition at Catholic, other religious, or nonsectarian schools, respectively. Based on data from Florida Active Private School Directory, over 60 percent of FTC-participating private schools are religious (16 percent are Catholic, with 44 percent representing other religions).

II. Methods

A. *Data and Sample*

We draw upon data provided by the Florida Department of Education (n.d.d) and the Florida Department of Health (n.d). The Florida agencies merged data on all public school students, including test scores, absences, and suspension data for students in grades PK–12, with birth records for children born in Florida between 1992 and 2002, providing measures of families' SES at birth as well as neonatal outcomes such as birth weight. Because we also received birth record 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 (FLDOE), 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 and Hart 2014).[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.[6] We also require students to be present in grade 1 (G1) so that we can assign our competitive pressure measures, as explained in Section IIC. 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 to those with Florida birth certificates (i.e., 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 et al. 2014).[7] 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.[8] 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. Overall, our main analytic sample

[5] One of our measures also relies on a list of houses of worship maintained by ReferenceUSA.

[6] In Figlio, Hart, and Karbownik (2020), we demonstrate that our results are robust to including students present in any grade from 3 to 8 or those present in *all* grades from 3 to 8.

[7] 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 B, 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 versus less competitive areas, our estimates should not be biased. The main results further remain qualitatively unchanged—and, if anything, increase in magnitude—when we reweight the regressions with characteristics of students born in Florida who do not enter Florida public schools.

[8] 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.

includes student data for roughly 1.2 million unique students in the 2002–2003 to 2016–2017 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 the preprogram 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 dataset 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 the 1998–1999 and 2006–2007 school years. This analysis may further alleviate concerns regarding our preferred sample that is limited to matched birth-school records. It also allows us to execute an event study design analysis that illuminates the lack of prepolicy trends. On the other hand, this supplementary dataset has three major limitations: (i) 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; (ii) this dataset has only limited demographic information on students since we lack birth record data; and (iii) we do not have data on disciplinary outcomes. For these reasons we use it only as a supplemental data source supporting our identifying assumptions, and we come back to these issues when we discuss our event study analysis in Section III.

B. *Models*

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

$$Y_{isglt} = \beta Expansion_t \times Competition_{sl} + \theta_{il} + \delta_{gt} + \varepsilon_{isglt}, \tag{1}$$

where Y_{isglt} captures an outcome measure for student i who entered the FLDOE data in 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 elementary (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 G1 school, the first mandated grade in Florida. Note that the inclusion of the student-by-stage fixed effects also holds constant time-invariant factors affecting students throughout their careers at a given academic stage, such as prior parental investments into children's human capital.[9] This implicitly functions similarly to controlling for lagged test scores as a means of

[9] 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 the 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

capturing prior endowments and investments. The term δ_{gt} is a grade-by-year fixed effect. Robust standard errors (ε_{isglt}) are clustered by G1 school.

The coefficient of interest is β, which estimates the interaction between a measure, $Expansion_t$, that captures the degree of program maturity or utilization 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 preprogram 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 preprogram competitive landscapes, as explored in Figlio and Hart (2014), rather than contemporaneous measures of competition, 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 preprogram 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 facilitate the interpretation of effect sizes.

We rely on exogeneity of shares in the Bartik (1991) quasi-experimental shift-share research design for identification, a la Goldsmith-Pinkham, Sorkin, and 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.[10] We demonstrate below that that voucher program growth increases competitive pressure, implying that the relevance identifying assumption likely holds. We also show suggestive evidence that the parallel trends assumption is likely to hold, bolstering our confidence that the shares are exogenous to changes in outcomes so that the second identifying assumption is likely to be satisfied.[11]

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, are on the conservative side.

[10] To be precise, Goldsmith-Pinkham, Sorkin, and 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 1 industry (private schools), 15 periods (school years), and either approximately 1,700 (schools) or 67 (school districts) locations.

[11] An alternative approach is proposed by Borusyak, Hull, and Jaravel (2022) 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 share exogeneity is likely to hold in our application. We characterize the evidence on parallel trends as suggestive since we only have three prepolicy periods. Ideally, to increase our confidence in the assumption, we would want to observe a longer prepolicy period.

C. *Measures*

Outcomes.—Our main cognitive outcomes rely on math and reading scores for grade 3–8 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.[12] We use school years 2002–2003 to 2013–2014 for math (2016–2017 for reading). The years of study vary because more advanced math students were able to exercise more choice about which assessments to take starting in school year 2014–2015; for instance, students taking Algebra I in eighth grade could take an algebra-specific examination rather than a general examination on eighth grade math. 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 2002–2003 to 2013–2014.[13]

We also study behavioral outcomes—likelihood of being suspended and absence rates—observed through the 2011–2012 school year. 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.

Competition.—Building on our previous work (Figlio and Hart 2014), we create an index measure of competition that incorporates information from five measures that capture the degree of competitive pressure that each public school is likely to face. We provide the description of the underlying measures in online Appendix A. Briefly, they involve information on the proximity, density, and diversity of potential competitors; the religiosity of the community (Hoxby 1994; Dee 1998; Jepsen 2002); and private school enrollments.[14] Because presenting all five measures is unwieldy, especially for the purposes of robustness checks and heterogeneity analyses, we present our main results using a single composite "Competitive Pressure Index" measure based on a principal components analysis of the 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.

[12] 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 2010–2011, and then by the Florida Standards Assessments (FSA) in 2014–2015 (Florida Department of Education n.d.c).

[13] In the main regression for the years 2002–2003 to 2013–2014, 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 the years 2014–2015 to 2016–2017, 74.7 percent have both scores, 1.7 percent have only math, and 23.6 percent have only reading.

[14] Figlio, Hart, and Karbownik (2020) present versions of this analysis based on the underlying competition measures.

The component loadings generated by the principal components analysis for the first two components are documented in online 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 markets with poorer-quality public alternatives (Arsen and 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 (online Appendix Figure A2), we demonstrate that this empirical design meets the instrument relevance assumption described by Goldsmith-Pinkham, Sorkin, and 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 and the distance to the nearest school accepting an FTC scholarship according to the FLDOE's Active Private School Directory. Hence, initial private school penetration serves as a strong proxy for the FTC-accepting options available to students, even over a decade post program launch.[15]

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 prepolicy competitive landscape of students' 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 (i) the flow of students empirically observed in our data from each elementary school

[15] When we analyze the effects of voucher program maturity using our shift-share design on private school penetration rather than cognitive and behavioral outcomes, then we find positive effects on our density and distance measures. The point estimates are also larger when we consider FTC-accepting schools compared to all private schools. This means that (i) our estimated effects of program maturation could envelop broader effects of increasing private school competition, and (ii) postreform measures of competition should not be used in defining treatment, as they are likely endogenous. While we can confirm that the pre-FTC private school landscape is highly correlated with realized FTC competition in later years, we are unable to determine the precise dynamics that determine the spread of competition, beyond the original private school landscape; this is a limitation to our study.

(G1) to each middle school (G6) in Florida, and (ii) the preprogram 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 prepolicy 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 the competition measure to ease interpretation of the interaction terms. Thus, the competitive pressure indicator captures whether the student's projected school is above or below the median of the Competitive Pressure Index. 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 some extensions, we present results with the median split point calculated at the student level and confirm that our results are qualitatively similar if we use continuous measures of competition or if we characterize competition more flexibly based on quintiles of competition (also see Figlio, Hart, and Karbownik 2020).

Program Maturation and Expansion.—We use several measures of program expansion and maturation. First, we show a simple set of graphs showing changes in program effects by year, with 2003 serving as the baseline year. In these models, the interaction of year dummies and competition measure captures the differential change in outcomes for students in high-competition versus low-competition schools for the year in question, relative to the high-versus-low-competition-school gap in the baseline 2003 year.

We also use a variety of measures to specifically capture multiple aspects of the growth and expansion of the program. Our main measure of program expansion is 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 1 percent increase in the number of students served in schools initially facing an above-median degree of competitive pressure, compared to the effects of the increase in locations with lower competitive pressure. Alternative expansion measures include utilized funding for the program, number of participating private schools, ratio of utilized funds to number of participating students, and ratio of FTC participants to K–12 public students (drawing on data from Florida Department of Education 2009, 2013, 2018, n.d.f; Florida Government 2001, 2010; and National Center for Education Statistics n.d.).

Student Characteristics.—We have measures on 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 United States, mother's marital status at the time of birth, mother's years of education at the time of birth, and 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, 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 or reduced-price lunch through their public school career, and those that were ever designated eligible for free or reduced-price lunch.

Online Appendix Table A2 provides descriptive statistics for the full population of Florida births and our main empirical sample. The comparison between these two samples 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, with more children whose mothers are high school dropouts (24.9 percent versus 20.9 percent), and fewer whose mothers are college graduates (14.7 percent versus 20.2 percent). We are also more likely to observe Black children (23.3 percent versus 19.4 percent). At the same time, ethnicity, immigrant origin, maternal age at birth, maternal health and child's health at birth are comparable in these two samples.

In online Appendix Table A2, we also investigate whether characteristics of students differ based on the degree of the prepolicy competitive pressure faced. Students in locations facing less competitive pressure prior to the program's introduction are disproportionately White (68.0 percent versus 37.3 percent), are less likely to have mothers born outside of the United States (14.5 percent versus 29.7 percent), have somewhat higher average birth weights (3,322 grams versus 3,272 grams), and have other markers suggestive of higher SES (based on subsidized lunch use and parents' marital status at birth). Interestingly, however, composition of parental education is relatively similar across these high- and low-competition locations, while maternal health is better in the former sample. Overall, the results are consistent with lower-SES families being more likely to locate in densely populated areas, which also tend to have more private school penetration. It further 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, respectively, 9.1 and 10.8 percent of a standard deviation lower in the former sample as compared to the latter. 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 rollout of the voucher program, controlling for student fixed effects.

III. Results

A. *Main Results*

We find that as the voucher program grows and matures, students in areas with more preprogram competitive pressure experience significantly greater improvement in outcomes than do students in areas with lighter preprogram competitive pressure. In Figure 1, we show changes in the effects of the initial level of competitive pressure over time (as the program was expanding). The key terms here are the interactions of initial preprogram competitive pressure level by year. Panel A of Figure 1 shows the year-by-year marginal effects for being located in a more competitive market for each outcome. In order to present the different outcomes on a consistent scale, we standardize all five outcome variables to have mean 0 and standard deviation of 100 in their respective empirical samples.

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 2013–2014 relative to schools in markets with less competitive pressure. Results for reading scores, for which we have more years of data, continued to modestly grow in the following years. 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 relative to peers in schools facing less competitive pressure starting only in 2005–2006, while significant reductions for absences emerged starting in 2008–2009. We present corresponding graphs for each underlying competitive pressure measure and each outcome (unstandardized) separately in online Appendix Figure A3.

We reach the same conclusions using various measures of program expansion as well (Table 2). The key terms here are the interactions of initial preprogram competitive pressure level with different ways of measuring program expansion, rather than interactions of initial preprogram competitive pressure level by year as in Figure 1.[16] Panel A provides results using our preferred measure of program expansion, the log number of scholarship enrollments statewide in a given year. Across outcomes, the results are statistically significant and consistent in pattern: students in more highly competitive areas see relative improvements in test scores (columns 1 to 3) and relative reductions in suspensions (column 4) and absences (column 5) as the program expands. Recall that the outcomes are multiplied by 100, and thus estimates in panel A suggest that a 10 percent increase in the number of students participating in the voucher program is associated with a 0.5 percent of a standard deviation greater improvement in combined math/reading scores for students in schools with an above-median density of private competitors, compared to students in schools facing lower degrees of competitive pressure. This effect is larger for reading (about

[16] While we show results for the competitive pressure index here, results using each underlying measure of competition interacted with our preferred expansion measure—the logged number of FTC participants—are available in online Appendix Table A3 and follow similar patterns. In sensitivity tests available upon request, we also confirm that results are not sensitive to the decision to log transform expansion measures.

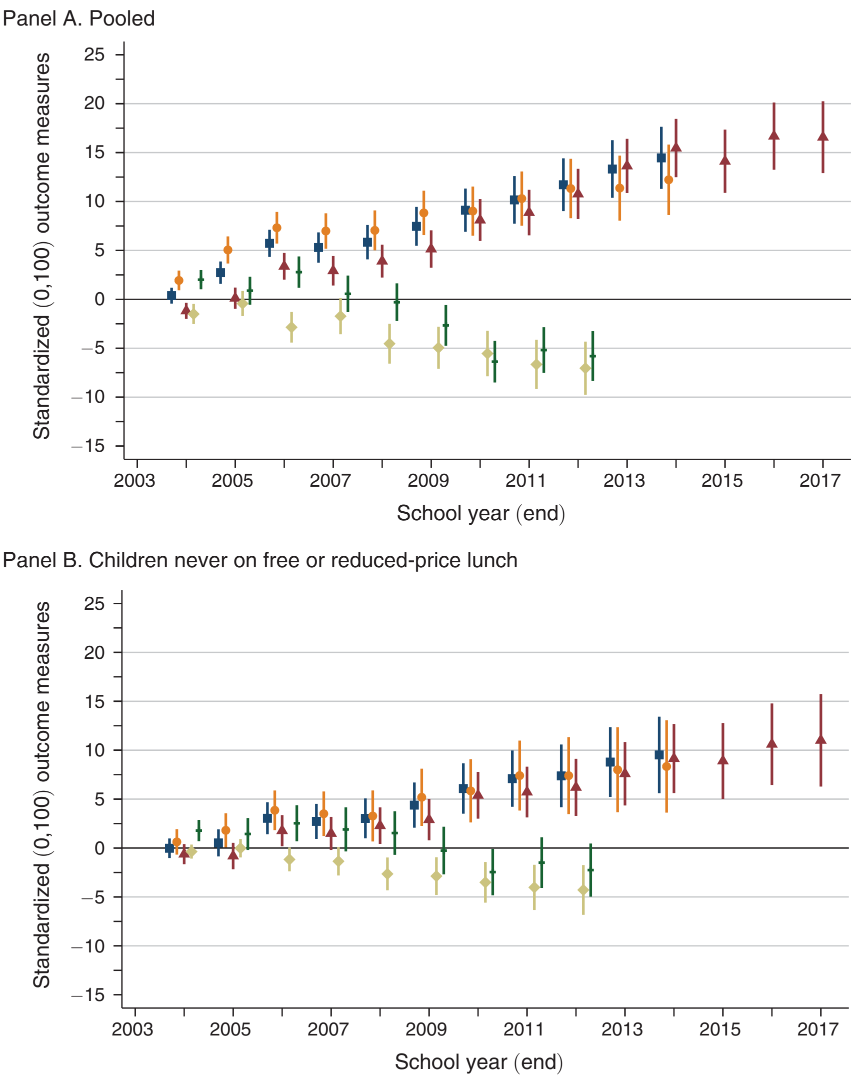


Figure 1. Effects of Voucher Expansion over School Years for Standardized Outcomes

(continued)

0.7 percent of a standard deviation) as compared to math (about 0.3 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 0.9 percent and the latter by 0.6 percent compared to their respective means. 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.

Panels B to E paint a very similar picture when we use alternative logged measures to capture program expansion, although magnitude-wise our preferred measure yields the more conservative effect sizes. Panel B measures expansion based on the

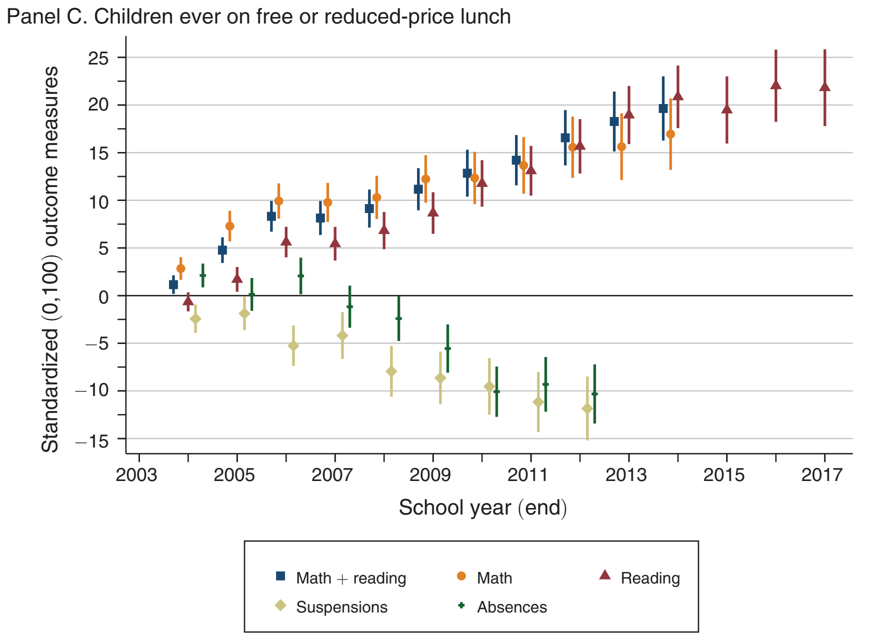


Figure 1. Effects of Voucher Expansion over School Years for Standardized Outcomes (*continued*)

Notes: This figure plots modified estimates from the main specification presented in panel A of Table 2 and from heterogeneity analyses presented in panels A and B of online Appendix Table A4, where, instead of an 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 (panel A of Table 2), while panels B and C divide the sample by the free or reduced-price lunch status of a child (online Appendix Table A4). 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 percent confidence intervals based on standard errors clustered at G1 school level.

realized expenditures on the program, panel C uses realized expenditures per FTC participant, panel D uses the ratio of FTC participants to the statewide enrollment of K–12 students in public schools, and panel E uses the number of private schools participating in the program. Sample sizes differ somewhat across panels because we do not have information on K–12 school enrollment and private schools in the early years of the program (see Table 1). Since the pattern of results is similar across expansion measures, we present results based on logged student participation going forward.

Results in panel A of Table 2 are also highly consistent with estimates presented in Figure 1, given that by 2013–2014, the program had expanded by nearly 300 percent compared to its original size. While it is difficult to disentangle program maturation from expansion, we note that the program effects seem to particularly accelerate around 2009. This is also coincident with a period in which students, funding, and participating private schools began to grow more rapidly, suggesting that pure program expansion effects may at least partly drive our results. Despite this focus on our expansion measure, we caution that both maturity and expansion mechanisms may be at play, as we discuss further in the conclusion.

Table 2 – Effects of Voucher Program Expansion on Student Cognitive and Behavioral Outcomes

	Math + reading (1)	Mathematics (2)	Reading (3)	Suspensions (4)	Absences (5)
Panel A. log number of scholarship enrollments expansion measure					
Expansion × above-median competition	5.11	2.64	7.39	−1.28	−0.28
	(0.59)	(0.74)	(0.61)	(0.27)	(0.05)
Panel B. log realized spending expansion measure					
Expansion × above-median competition	4.75	2.55	6.54	−1.36	−0.31
	(0.55)	(0.69)	(0.55)	(0.27)	(0.05)
Panel C. log realized spending per capita expansion measure					
Expansion × above-median competition	12.12	8.00	14.91	−4.67	−1.55
	(1.84)	(2.39)	(1.77)	(1.46)	(0.23)
Mean [SD] of *Y*	0.00 [93.08]	−0.04 [99.98]	−0.02 [99.98]	13.67 [34.35]	5.04 [5.79]
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
Panel D. log percentage of K–12 public school enrollment expansion measure					
Expansion × above-median competition	5.58	3.51	7.68	−1.57	−0.31
	(0.67)	(0.84)	(0.68)	(0.28)	(0.05)
Mean [SD] of *Y*	0.03 [93.08]	−0.04 [99.98]	−0.01 [99.98]	14.62 [35.33]	5.12 [5.92]
Observations	5,533,652	5,479,862	5,959,148	4,784,217	4,784,217
Children	1,207,762	1,207,416	1,209,862	1,213,690	1,213,690
Panel E. log number of participating private schools expansion measure					
Expansion × above-median competition	13.45	8.45	17.74	−1.26	−1.10
	(1.83)	(2.35)	(1.79)	(0.93)	(0.15)
Mean [SD] of *Y*	0.01 [93.08]	−0.04 [99.98]	−0.02 [99.98]	14.11 [34.81]	5.07 [5.85]
Observations	5,896,771	5,842,012	6,321,338	5,157,292	5,157,292
Children	1,215,814	1,215,530	1,217,917	1,220,815	1,220,815

Notes: Sample is based on individual-level observations in grades 3 to 8 for students attending Florida public schools between 2002–2003 and 2016–2017 and born between 1992 and 2002 in Florida. Each child has to be observed at least in G1 so that we can assign them school-level competition measures that 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 as logarithm of number of scholarships awarded based on column 3 of Table 1 (panel A), logarithm of realized spending based on column 2 of Table 1 (panel B), logarithm of realized spending per capita based on the ratio of columns 2 and 3 of Table 1 (panel C), logarithm of percentage of K–12 public school enrollment based on column 4 of Table 1 (panel D), and logarithm of number of participating private schools based on column 5 of Table 1 (panel E). These are measured at the annual level and are available between 2002–2003 and 2016–2017 for measures in panels A to C, between 2004–2005 and 2016–2017 for measure in panel D, and between 2003–2004 and 2016–2017 for measure in panel E. Test scores are based on FCAT developmental scores for the years 2002–2003 to 2013–2014 and on FSA developmental scores for the years 2014–2015 to 2016–2017, and we standardize them in-sample by year and grade to have mean 0 and standard deviation of 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 the years 2002–2003 to 2011–2012, and they are multiplied by 100. Each column represents a separate outcome variable. The competition measure is principal components analysis competition index ("Competitive Pressure Index") based on five measures presented in online Appendix Table A1. The 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 and grade-by-school-year fixed effects. School level is defined as indicator for grade 6 to 8 versus 3 to 5. Standard errors are clustered at G1 school level.

Source: Author calculations

B. *Benchmarking and Interpreting 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 online Appendix 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 (a gap of 9.1 percent of a standard deviation), reading scores (a gap of 10.8 percent of a standard deviation), and combined scores (a gap of 10.0 percent of a standard deviation), and a higher likelihood of suspensions (a gap of 1.4 percentage points, or 4.1 percent of a standard deviation). Given these figures and the effect sizes presented above, a 10 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 9.1 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 overrepresented 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 10 percent; the number of scholarship users was nearly 7 times higher in 2017–2018 than in 2002–2003—and thus the realized gains should actually be much larger. If we use our preferred estimates from Table 2 (panel A), 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 versus 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 outcome gaps between schools facing more or less competitive pressure.

It is also helpful to compare the magnitudes of our findings to other estimates in the education and human capital formation literature and, in particular, to those obtained using data from Florida. For instance, Figlio and Hart (2014), who looked at the introduction of the program, found that a 1 standard deviation increase in the preprogram 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 described later in the paper. 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. This implies that large program growth 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 sevenfold, suggesting that program maturity effects should 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 and Terrier (forthcoming) and

Table 3—Effects of Voucher Program Expansion on State Measures of School Quality

	School quality (FLDOE grades)			
	A school (1)	B school (2)	C school (3)	D or F school (4)
Expansion × above-median competition	2.40	−1.65	−2.81	2.06
	(1.05)	(1.01)	(1.06)	(0.79)
Mean of *Y*	48.68	20.33	21.95	9.05
Observations	42,897	42,897	42,897	42,897

Notes: Columns 1 to 4 present the effects of voucher program expansion on school quality measures published by the FLDOE. Column 1 is an indicator for A-grade school, column 2 is an indicator for B-grade school, column 3 is an indicator for C-grade school, and column 4 is an indicator for D- or F-grade school. All indicator variables are multiplied by 100. Each regression is based on cells aggregated to school in G1 by school level by school year level. The table displays the coefficient of interest, which is the interaction between the preferred competition and expansion measures from panel A of Table 2, and each regression includes school-in-G1-by-school-level fixed effects and year fixed effects. No additional controls are included, and data span school years 2002–2003 to 2016–2017. Standard errors are clustered at G1 school level.

Source: Author calculations

Gilraine, Petronijevic, and 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 percent of the birth order gap in reading scores (Breining et al. 2020). Note, however, that all these papers estimate the total effects of the treatment. By contrast, our estimates represent gains in high-competition areas relative to low-competition areas, so our results may reflect a lower bound on the total effect, a point we return to in the conclusion. These comparisons suggest that our estimated magnitudes appear meaningful from the perspective of economic and education policy.

C. *Effects on State School Quality Measures*

Since program maturity and scale-up affect student cognitive and behavioral outcomes, we can also ask whether they translate to state-measured indicators of public school quality. We explore this question using two approaches. First, an analysis (Table 3) using data on state-designated school accountability grades (Florida Department of Education n.d.e) suggests a “hollowing-out” effect, in which schools in above-median competition areas become more likely to earn “A” (top) grades but also become more likely to earn “D/F” (failing) grades as the program matures. Second, we explore this phenomenon by directly investigating the “grade points” that underlie the school grades. Because the accountability grading formula changed across years, we standardize it by the maximum number of points possible, so that the measure represents the percent of possible points (*pct*) earned by a school in a given year and ranges from 0 to 100. While the precise relationship between percent of points and school grades varies across years, schools that earned below 40 percent of the points never attain higher rating than a D, while schools with 70 percent of the points or more only rarely earned ratings lower than an A (online Appendix Figure A4).

Thus, we estimate a series of models capturing the probability of a school earning $pct = P$ points or higher for a series of target values P (online Appendix Figure A5). Since virtually no schools score below 15 or above 85 points, we truncate the graph at these values. Consistent with the "hollowing-out" effect, we find estimates that are negative and growing in magnitude for low values of P that reach a minimum at $P = 45$ and then reverse the trend to begin increasing up to $P = 70$. Past this threshold, the estimates decline again but remain positive and statistically significant. Therefore, we are more likely to observe schools scoring above $P = 60$ and below $P = 45$, implying that we are less likely to see schools scoring in the middle range of the distribution of P. As illustrated in online Appendix Figure A4, these score ranges closely correspond to A, D/F, and B/C grades, respectively.

Note that this hollowing-out effect is not necessarily related only to the performance of students born in Florida and remaining in the public school system at all points where they have outcomes. Unlike our main analyses, these school quality measures are also affected by changes in the composition of all students who remain behind in the public schools (and who therefore contribute test scores to school grade measures). While our student fixed effects approach guarantees that the results are not driven by changes in student composition across years affecting our sample of who is tested, changes in student composition may have peer effects on students. These peer effects may themselves be one channel through which the student performance of stayers is affected, a point that we return to in Section IV.

D. *Heterogeneity*

Returning to our main, student-level outcomes of interest, 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 results for all five outcome variables in panels B and C of Figure 1 and online Appendix Tables A4 and A5.[17]

Results are generally consistent in pattern across all subgroups; however, the exact magnitudes and statistical significance vary somewhat. Lower-SES students—whether measured by use of free or reduced-price lunch or by mother's education level—see larger effects across all outcomes than their high-SES peers. Within test score outcomes, these differences are more pronounced for reading than for math. However, notably, most effects for higher-SES students are still statistically significant. 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 SES, using a measure introduced for these data by Autor et al. (2019). The SES composite index is created

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

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 it uses these to construct a composite SES gradient index. We then separate the sample by SES deciles to observe scale-up effects across the SES gradient. Online Appendix Figure A6 presents coefficients associated with increasing competitive pressure by these SES deciles. Effects are strongest for families in the bottom six deciles, but increased competitive pressure is associated with statistically significant benefits for all families except for the very top SES decile.

We also observe differences in magnitudes of coefficients across racial and ethnic groups as well as by immigration status (online Appendix Table A5). First, test score gains are statistically significant and very similar for both Black and White children. The former group, however, does not experience statistically significant behavioral benefits in terms of absences or suspensions. Second, Hispanic students 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 unsurprising given that Hispanic children are disproportionately likely to have non-native-born mothers in our sample (68 percent of Hispanic children in our sample have a foreign-born mother, versus only 9 percent for non-Hispanic children).

E. *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 Table 4.[18] We reproduce our main Table 2 results in panel A to ease the comparisons. 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 (following Florida Association of Counties designation of regions). This specification produces results that are qualitatively similar to, albeit smaller in magnitude than, our main results, particularly for reading and suspensions. This suggests that part of the effects of competition may stem from differences between regions rather than solely at a hyperlocal level, a point to which we return in Section V. Panel C explores whether our results are an artifact of time-specific trends in performance gains by different student subgroups, some of which are differentially exposed to competition, or by schools of different initial quality levels. These specifications include interactions of our expansion

[18] A previous version of this paper presents many additional robustness checks (Figlio, Hart, and Karbownik 2020).

Table 4—Select Robustness Analyses of the Preferred Estimates

	Math + reading (1)	Mathematics (2)	Reading (3)	Suspensions (4)	Absences (5)
Panel A. Baseline					
Expansion × above-median competition	5.11	2.64	7.39	−1.28	−0.28
	(0.59)	(0.74)	(0.61)	(0.27)	(0.05)
Mean [SD] of *Y*	0.00 [93.08]	−0.04 [99.98]	−0.02 [99.98]	13.67 [34.35]	5.04 [5.79]
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 × above-median competition	3.60	2.55	4.09	−0.68	−0.25
	(0.61)	(0.79)	(0.57)	(0.28)	(0.05)
Mean [SD] of *Y*	0.00 [93.08]	−0.04 [99.98]	−0.02 [99.98]	13.67 [34.35]	5.04 [5.79]
Observations	6,160,525	6,104,889	6,584,014	5,427,985	5,427,985
Panel C. Including expansion times predetermined individual and school quality characteristics					
Expansion × above-median competition	3.97	2.42	5.30	−0.82	−0.18
	(0.54)	(0.73)	(0.52)	(0.28)	(0.05)
Mean [SD] of *Y*	−0.61 [93.11]	−0.64 [100.00]	−0.65 [100.01]	13.81 [34.50]	5.06 [5.83]
Observations	5,818,006	5,765,762	6,214,205	5,135,212	5,135,212
Panel D. Continuous competition measure					
Expansion × competition	1.45	0.62	2.30	−0.42	−0.09
	(0.16)	(0.20)	(0.16)	(0.07)	(0.01)
Mean [SD] of *Y*	0.00 [93.08]	−0.04 [99.98]	−0.02 [99.98]	13.67 [34.35]	5.04 [5.79]
Observations	6,160,525	6,104,889	6,584,014	5,427,985	5,427,985
Panel E. Competition measure assigned using observed first and sixth grades					
Expansion × above-median competition	4.34	1.95	6.53	−1.03	−0.28
	(0.57)	(0.72)	(0.59)	(0.25)	(0.05)
Mean [SD] of *Y*	00.00 [92.87]	00.00 [99.75]	−0.07 [99.84]	13.22 [33.87]	5.01 [5.71]
Observations	5,761,773	5,714,711	6,123,884	5,117,781	5,117,781
Panel F. Competition assigned at birth based on zip code of birth					
Expansion × above-median competition	7.49	5.49	9.52	−1.94	−0.41
	(0.57)	(0.70)	(0.60)	(0.24)	(0.04)
Mean [SD] of *Y*	0.09 [92.97]	−0.04 [99.98]	−0.01 [99.98]	12.12 [32.63]	4.82 [5.34]
Observations	4,930,478	4,880,567	5,340,120	4,189,326	4,189,326

Notes: Robustness checks based on estimates from panel A of Table 2. Panel A replicates the main result from panel A 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 adds as additional controls interactions between our expansion measure and a series of time-invariant student characteristics (student sex, race, and ethnicity; immigration status of student's mother; parental marital status; maternal education categories; maternal age at child's birth; and logarithm of child's birth weight) as well as prepolicy school quality measures defined by FLDOE for each school (indicators for B school, C school, D school, and F school, with A schools serving as reference group); panel D replaces the dummy indicator for above-median prereform competition with the continuous measure; panel E assigns the middle school prepolicy competition measures based on the actual grade 6 (middle) school initially attended by each student; and panel F assigns competition based on zip code and cohort of birth (leave-one-out measure of the average competition actually experienced across grades 1 through 8 for students in each zip-code-by-birth-cohort cell excluding the focal child). 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). All outcome variables are multiplied by 100. Standard errors are clustered at G1 school level.

Source: Author calculations

measure with a series of indicators for time-invariant student and school quality measures.[19] The pattern of results is similar to our main specification.

Our results are also largely robust to ways of characterizing competition. Panel D uses the underlying, continuous Competitive Pressure Index measure of preprogram competitive pressures in our interaction terms rather than the median-split term.[20] 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. In panel E, we assign the middle school prepolicy 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 also remain similar.

Another way to look at the influence of the prevoucher 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 presents behavioral outcomes. The effects of expansion on test scores are more pronounced for schools with higher levels of competition prepolicy, and this pattern is particularly evident for reading. Absences show similar patterns to the cognitive effects, with more competitive landscapes associated with greater improvements in the outcome across the span of the competition distribution. On the other hand, for suspensions, there is relatively little difference in the second and third quintiles of prepolicy competition compared to the schools with the least competitive landscapes, but pronounced differences emerge for schools in the fourth and fifth quintiles. In sum, we find greater advantages for students from schools in the most competitive areas as the program expands, roughly consistent with a dose-response relationship.

A remaining concern may be that families could select G1 schools strategically, in ways that result in positive correlation between expected student gains and the differential level of competition faced by schools. To address this concern, in panel F of Table 4, we use data from birth certificates to assign students to competition levels expected based on their reported zip codes at time of birth. Thus, we construct leave-one-out measures of the average competition actually experienced across grades 1 through 8 for students in each zip-code-by-birth-cohort cell, excluding the focal child. These measures then represent the expected competition that

[19] Specifically, the fixed student characteristics include indicators for student sex, race/ethnicity, an indicator for whether the child's mother was non-US-born, an indicator for parents' marital status at birth, indicators for high-school-graduate mother and college-graduate mother (high-school-dropout mother is the reference category), and variables capturing maternal age at birth as well as logarithm of infant birth weight. For schools, we include indicators for prepolicy school grades assigned by the FLDOE (indicators for B, C, D, and F schools, with A schools serving as the reference group, based on the lowest grade each school received in the three years prepolicy).

[20] This specification helps to address a potential limitation of our main results: the fact that above/below-median competition distinction is relatively coarse. We prefer the coarser measure in our main specifications because the interaction of two continuous variables is more challenging to interpret in a quantitatively meaningful way. However, we find it reassuring that our specifications using more flexible definitions of competition in panel D of Table 4 and in Figure 2 produce qualitatively similar results to the main specification.

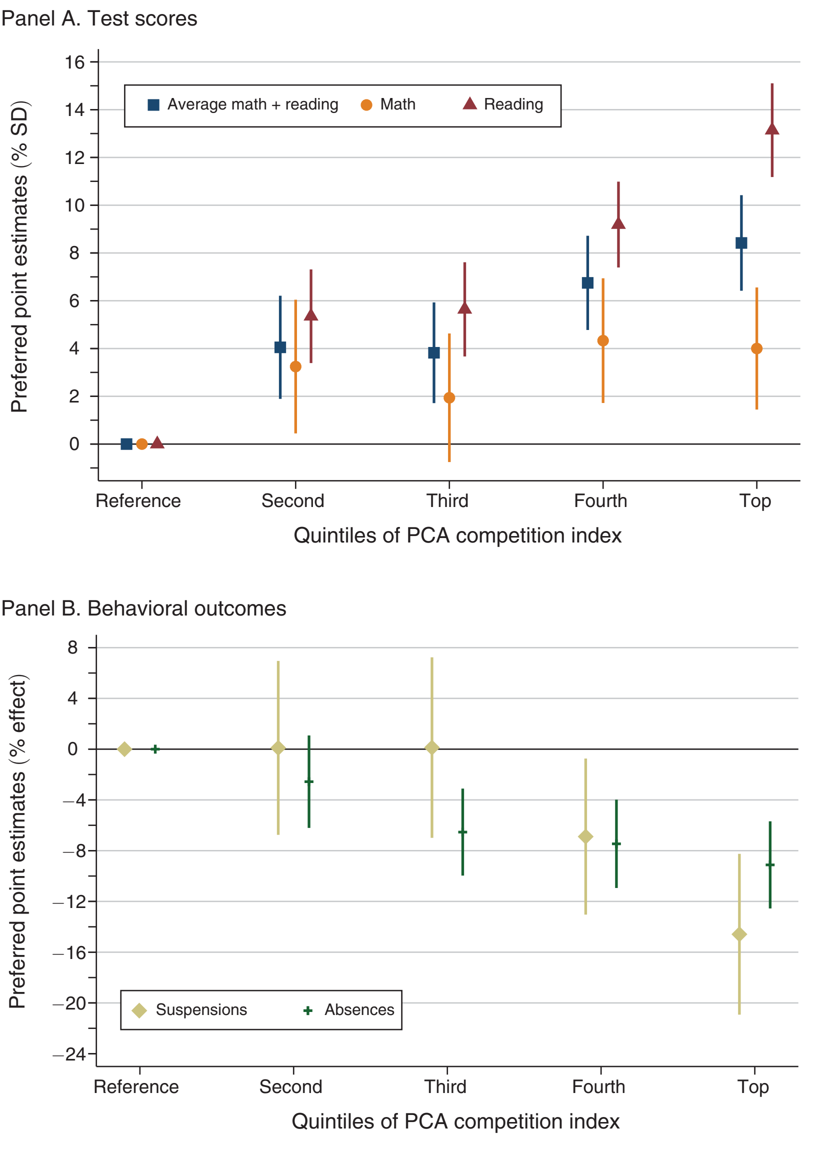


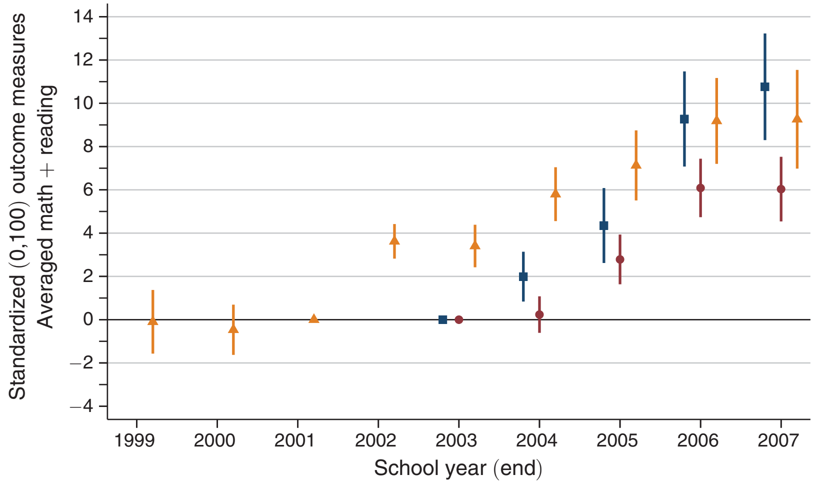
Figure 2. Quintiles of Competition

Notes: This figure presents estimates using specification and sample from panel A of Table 2 where, instead of median, we interact quintiles of PCA competition index with the log of expansion measure. The bottom quintile is a reference category. Outcome variables are averaged mathematics and reading test score (navy squares), mathematics test score (orange circles), and reading test score (maroon triangles) in panel A, as well as likelihood of being suspended (khaki diamonds) and absence rate (green pluses) in panel B. Spikes present 95 percent confidence intervals based on standard errors clustered at G1 school level.

each child's schools should face. The results are, if anything, slightly larger than in the main specification.

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 are occurring regardless of the voucher policies. In Figure 3, we

Panel A. Averaged mathematics and reading



Panel B. Mathematics

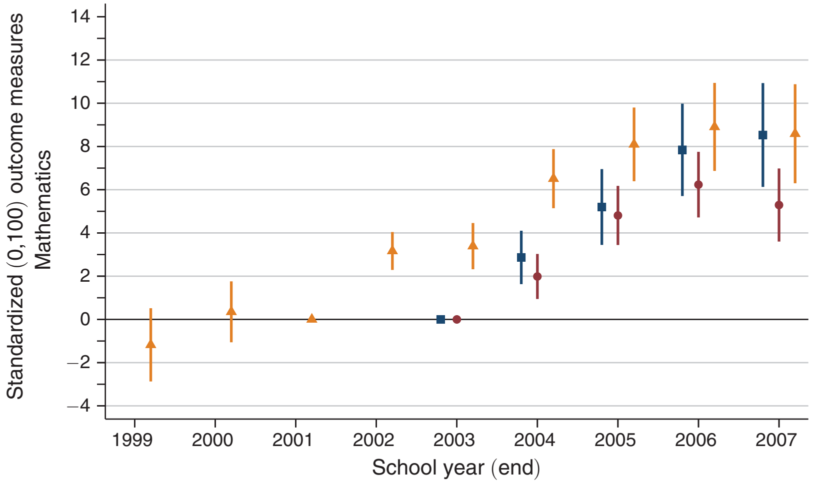


Figure 3. Event Studies (*continued*)

test this using data from a different sample based on all public school students but a limited set of years as described in Section IIA. Here, we do not have student outcome data in all grades in all years (specifically in the prepolicy years), and thus looking at within-student changes is not feasible to examine prepolicy 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 1 schools as a consequence of the policy. To ameliorate the concern that student sorting to grade 1 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 1 schools prima facie could not have been affected by the policy. Here, the relationship between competition and

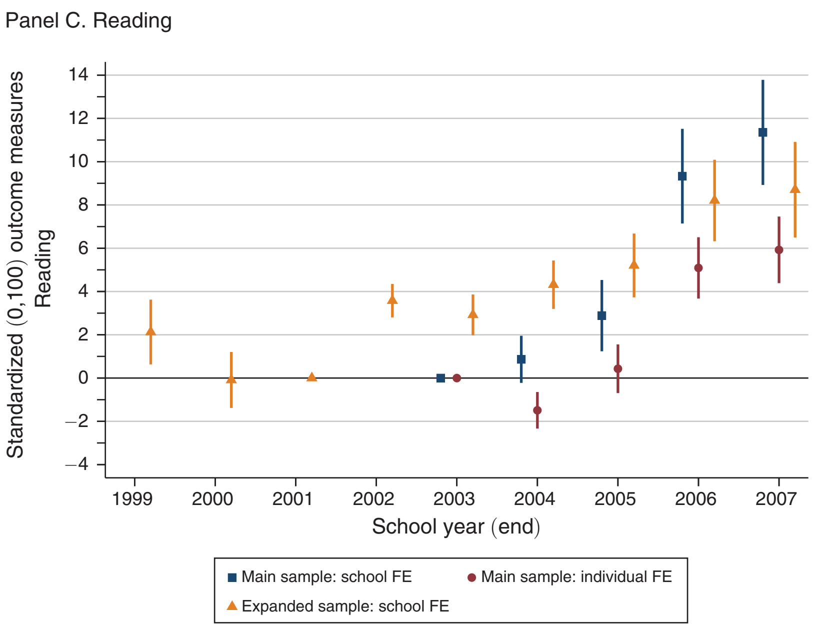


Figure 3. Event Studies (*continued*)

Notes: This figure presents 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 the composite competition index measure with school years. This measure is based on prereform competitive landscape. It is assigned at G1 school and G6 school in the main sample as in Table 2, while in the expanded sample it is assigned at the contemporaneously observed school. Navy and maroon scatter plots present analyses based on modified data used in Figure 1 (main sample), while the orange scatterplot uses a separate dataset described in Section IIA (expanded sample). Orange triangles present estimates from the expanded sample with contemporaneous school and grade-by-school year FE. Navy squares present estimates from the main sample with G1-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 C. Spikes present 95 percent confidence intervals based on standard errors clustered at the G1 school level in the main sample and the contemporaneous school level in the expanded sample.

student outcomes in each year is compared to the relationship in the omitted year of 2000–2001 (Figure 3, orange triangle markers). We provide additional information about these analyses in online Appendix C.

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 (1998–1999 to 2006–2007 rather than 2002–2003 to 2016–2017 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 were implemented in our preferred sample. We include point estimates that replicate our main estimation strategy (individual fixed effects) using 2002–2003 to 2006–2007 data (maroon circle markers), as well as a set of point estimates that uses our 2002–2003 to 2006–2007 sample but includes school-by-stage rather than individual-by-stage fixed effects (navy square markers). Both of these sets of estimates use an omitted year of 2002–2003, 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.

Importantly, Figure 3 shows that prepolicy competition is unrelated to changes in student scores before the voucher policy was announced in spring 2001. The coefficients for school years 1998–1999 and 1999–1900 are not significantly different from zero in five out of six cases, and the only statistically significant result (for reading in the 1998–1999 school year) suggests trends that, if anything, ran in the opposite direction of our postpolicy findings. Thus, our reading results can be considered conservative estimates.[21] Thereafter, as for the main results, we see that public schools in more competitive areas improve more quickly than public schools in less competitive areas, with positive and significant coefficients in each year from 2001–2002 onward and for each outcome. These figures suggest that there were no preexisting 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, and Swift (2020). Thus, to the extent that we can measure it, we believe that our results reflect the effects of the scale-up and the maturing of the program itself differentially affecting schools in higher-baseline competition areas, rather than reflecting any prior differential trends.

IV. Mechanisms

So far, we have suggested that our results are due to increased competitive pressure on public schools associated with program maturity and scale-up, which may induce schools to respond with productive academic changes that improve school performance (see, e.g., Jabbar 2015). However, there may be alternative explanations to these findings. For instance, voucher programs may change the composition of students or teachers remaining in the public schools over time, and these changes might be related to the degree of voucher competition that 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 in the public schools might plausibly go down. To the extent possible, we investigate these alternative explanations in this section.

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

[21] Our results are likely conservative for another reason. The Opportunity Scholarship Program (OSP), 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 eight 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 and 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.

Behind Act (NCLB) was under discussion at the same time that the FTC program was passed. Since schools with greater preprogram competition were lower performing (online Appendix 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 that this is unlikely for several reasons. First and foremost, it would have to be the case that the accountability pressure has been growing in the same way as the voucher program, and there is no evidence on that. Second, prior to NCLB, Florida already had a comparably stringent accountability law that, however, did not put substantial pressure on public schools (Rouse et al. 2013). Finally, our positive results extend, albeit to a lesser degree, to higher-SES students that are unlikely to be targeted by either the voucher program or NCLB, implicating increasing competition as a more likely driver of our results.

A. *Peer Composition*

First, we 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 the vouchers.[22]

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 to 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.[23] If we see that schools with more

[22] The benefits from changes in student composition could also accrue due to changes in teacher labor supply. For example, Karbownik (2020) documents that increase in student ability due to school choice–induced sorting in Sweden leads to declines in teacher turnover.

[23] 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, and abnormal conditions at birth and congenital anomalies, as well as mother's education, race, ethnicity, place of birth outside of United States, 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, at least for test scores, our explanatory variables have nontrivial predictive power for these outcomes. We then aggregate these predicted outcomes at the G1-school-by-stage-by-year level and link them to competition-by-expansion interaction measured at the G1-school-by-stage-by-year level since this is the level of variation in our preprogram competition measures, as explained in Section II. The results are very similar

Table 5—Mechanisms: Effects of Voucher Program Expansion on Peer Composition, Class Size, and Teacher Composition

	School-level peer effects					School-level class size	School-level teacher experience
	Math + reading (1)	Mathematics (2)	Reading (3)	Suspensions (4)	Absences (5)	(6)	(7)
Expansion × above-median competition	−0.40	−0.43	0.16	0.12	−0.01	−0.22	0.00
	(0.29)	(0.29)	(0.27)	(0.07)	(0.01)	(0.06)	(0.10)
Mean of Y	−2.24	−2.30	−1.60	14.42	5.15	16.50	10.22
Observations	37,880	37,875	44,685	31,334	31,334	32,340	30,863
	Teacher experience groups				Teacher racial/ethnic composition		
	0 to 2 years (8)	3 to 5 years (9)	6 to 12 years (10)	13+ years (11)	% White (12)	% Black (13)	% Hispanic (14)
Expansion × above-median competition	−1.23	−0.23	1.77	−0.09	−1.64	0.94	0.54
	(0.55)	(0.40)	(0.53)	(0.51)	(0.32)	(0.29)	(0.21)
Mean of Y	23.12	18.26	25.28	32.54	73.40	13.92	9.95
Observations	30,863	30,863	30,863	30,863	30,863	30,863	30,863

Notes: Columns 1 to 5 present the effects of voucher program expansion on 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. Columns 7 to 11 present the effects of voucher program expansion on school-level measures of teacher experience. We present effects on mean years of experience in column 7 and fractions of teachers in specific experience-range bins in columns 8 to 11. Columns 12 to 14 present the effects of voucher program expansion on school-level measures of teacher racial and ethnic composition. Each regression is based on cells aggregated to school in G1 by school level by school year level. The table displays the coefficient of interest, which is an interaction between the preferred competition and expansion measures from panel A of Table 2, and each regression includes school-in-G1-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 when prenatal care began, 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 predicted values are then aggregated at G1 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 FLDOE (http://www.fldoe.org/finance/budget/class-size/class-size-reduction-averages.stml) separately for grades prekindergarten (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 $CS = 0.5CS_{PK-3} + 0.5CS_{4-8} + 0CS_{9-12}$. Data on teacher experience and demographics are available for school years 2002–2003 to 2011–2012. Variables in columns 1 to 5 as well as 8 to 14 are multiplied by 100. Standard errors are clustered at G1 school level.

Source: Author calculations

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 of our main results, suggesting that schools in more competitive landscapes were more likely to enroll

if instead we aggregate both outcomes and competition at the G1-school-by-year level. They are likewise similar when weighted by number of students in each aggregated cell.

students with higher predicted suspension risk as the voucher program expanded. That would be akin to 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 effects in Table 5 are negligible in magnitude compared to our main results. For instance, 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 literature on peer effects in test scores in general suggests relatively small, if any, effects on students (Sacerdote 2014). 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 experience levels or student participation in special education (Lankford and Wyckoff 1995). 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 the 2006–2007 through 2016–2017 school years posted by the Florida Department of Education (n.d.a, g), which report the average class size for each school separately for students in grades PK–3, grades 4–8, and grades 9–12. 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 the 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 magnitude of this coefficient is too small to realistically explain away much of our main cognitive and behavioral effects. Our point estimate of -0.22 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 7-student reduction in class size effect implied by the Tennessee Student/Teacher Achievement Ratio experiment (Krueger 1999); this estimated effect is similar in magnitude to those found by Angrist and Lavy (1999); Lindahl (2005); Chetty et al. (2011); and Fredriksson, Öckert, and Oosterbeek (2013) in related studies. If we assume that 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 ($0.022 \times 22/7 = 0.07$ percent of a standard deviation). This would account for only about 14 percent of the effect of competition on combined math and reading test scores that we estimated in Table 2, where our coefficient implied a 0.51 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 and Rivkin 2009; see Chingos 2013 for a thorough review of the class size literature) estimate that class size reductions of about ten students produce improvements in test scores of between 0.05 and 0.22 standard deviations, and that other papers (e.g., Leuven and 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. Thus, while class size may be contributing somewhat to the observed positive effects of competitive pressure on cognitive and behavioral outcomes, it seems unlikely that changes in class size are a quantitatively meaningful driver of the findings.

C. *Teacher Composition*

Yet another way through which the scale-up of the voucher program could have generated the observed cognitive and behavioral gains is through changes in teacher quality and composition. In the short run, a period that most papers consider, this input into the school production function is unlikely to play a significant role due to costs of firing and hiring new teachers, and their relatively inelastic labor supply. In our application, however, we consider program expansion over 15 years, and thus it is conceivable that teacher sorting is an important mechanism, either through teacher labor responses to choice-induced student sorting (Karbownik 2020) or due to competitive effects as private schools try to recruit teachers (Hensvik 2012). In our data, since we are unable to link students with their teachers, we cannot calculate teacher value added. At the same time, we have information, aggregated at the school-by-year level, about an imperfect proxy of teacher quality—their experience (Harris and Sass 2011; Ladd and Sorensen 2017). Furthermore, we also observe teacher racial and ethnic composition, measured as the fractions of teachers who are White, Black, and Hispanic. This could be important since racial/ethnic match between teachers and students could improve students' achievement, disciplinary, and absence outcomes (Dee 2004; Lindsay and Hart 2017; Holt and Gershenson 2017; Gershenson et al. 2022; Grissom, Kern, and Rodriguez 2015; Bristol and Martin-Fernandez 2019).

Since the data are aggregated at the school-by-year level, we execute this analysis in the same way as peer composition and class size estimates, with the exception that teacher composition data are available only for school years 2002–2003 to 2011–2012. The results are presented in Table 5. Columns 7 to 11 show experience estimates, while columns 12 to 14 document the demographic estimates. In both instances we find statistically significant changes in teacher composition. Although there is no difference in average experience levels (column 7), we discover a substitution effect between relatively inexperienced teachers with up to 2 years of experience and those with 6 to 12 years of experience. There is no effect on the fraction of teachers with more than 13 years of experience. To the extent that more experienced teachers improve student outcomes, including test scores and behavior (Ladd and Sorensen 2017), this result could be a plausible mediator for our main findings presented in Section IIIA.

We further find that schools facing landscapes with more initial competitive pressures experienced changes in the demographic composition of their teachers as the program matured and scaled up. In particular, they employed relatively more minoritized (Black and Hispanic) teachers and relatively fewer White teachers. Recall from online Appendix Table A2 that schools located in above-median-competition locations have about twice as many Black and Hispanic students as schools located in relatively less competitive areas. Thus, if there is indeed a racial complementarity between students and their teachers, such sorting could be a mechanism at play.

Despite the fact that we find statistically significant effects on both measures of teacher experience and demographics, it is challenging to assess these magnitudes in the context of potential cognitive and behavioral gains. Most studies in the extant literature estimate gains from racial match between teachers and their student or effects per additional teacher in a specific experience category. On the other hand, results in Table 5 can only speak to relative changes in the fraction of teachers with specific characteristics. Furthermore, the effect sizes when compared to the means of each dependent variable imply relatively modest reshuffling, not exceeding 7 percent. To better understand the potential contribution of these changes in composition for our human capital outcomes, we explore this descriptively in our mediation analysis below.

V. District-Level versus 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, and 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 equally important as (or more important than) competition in 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 the school (e.g., class size) and district (e.g., teacher compensation schemes) level that might be changed in response to the increasing competitive pressures.

To examine this question, we create a district-level version of our competition measure, capturing whether the district-level competition that each child is expected to face, based on their G1 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 that has a roughly equal number of students in above- and below-median districts, we construct it using the level of competition faced by the median *student* as the cut point.[24] To make sure that the school- and district-level competition variables are consistent with each other, we also use the student-level median (rather than the school-level median) to stratify the school-level competition variable. As for the rest of the analyses, the school- and district-level competition measures are both based on prepolicy private school landscapes, and the above-median indicators are interacted with the year-specific measure of voucher program expansion. The Pearson correlation coefficient ($r = 0.62$) between the two measures suggests moderate correlation, 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 our main specifications but defines the median competition split at the student rather than the school level, 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 to E), the effect of district-level competition dominates the school-level measures. In all cases the school-level interaction terms also fall in magnitude compared to column 1, although with the exception of math scores, they retain both their signs and their 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 (online Appendix Table A6), we create separate categories capturing both the school-level and the district-level competition simultaneously. That is, we categorize whether each student is in a high-competition school/high-competition district (41 percent of sample), low-competition school/high-competition district (10 percent of sample), high-competition school/low-competition district (9 percent of sample), or low-competition school/low-competition district (40 percent 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 a school experiences lower

[24] Two small districts are narrowly classified as above median in terms of elementary school competition but below median in terms of middle school competition. 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 test scores in our data at 0.72.

Table 6–School versus District Competition and the Role of Time-Varying Characteristics

	Baseline		Additional time-varying control variables					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. Math + reading ($N = 6{,}160{,}525$; children $= 1{,}221{,}023$; mean of $Y = 0.00$; SD of $Y = 93.08$)								
Expansion × above-median competition (school)	5.34	2.02	1.78	2.08	2.02	2.05	2.27	2.09
	(0.58)	(0.68)	(0.68)	(0.69)	(0.68)	(0.68)	(0.68)	(0.67)
Expansion × above-median competition (district)		5.37	5.22	5.20	5.39	5.36	5.02	4.68
		(0.70)	(0.69)	(0.70)	(0.70)	(0.70)	(0.70)	(0.70)
Panel B. Mathematics ($N = 6{,}104{,}889$; children $= 1{,}220{,}753$; mean of $Y = -0.04$; SD of $Y = 99.98$)								
Expansion × above median competition (school)	2.93	0.33	−0.10	0.35	0.32	0.33	0.51	0.10
	(0.71)	(0.83)	(0.83)	(0.83)	(0.83)	(0.83)	(0.83)	(0.83)
Expansion × above median competition (district)		4.23	4.05	4.15	4.25	4.23	3.98	3.75
		(0.84)	(0.84)	(0.84)	(0.84)	(0.84)	(0.85)	(0.85)
Panel C. Reading ($N = 6{,}584{,}014$; children $= 1{,}223{,}123$; mean of $Y = -0.02$; SD of $Y = 99.98$)								
Expansion × above-median competition (school)	7.65	3.84	3.71	3.92	3.85	3.88	4.07	4.06
	(0.60)	(0.76)	(0.75)	(0.76)	(0.76)	(0.76)	(0.75)	(0.75)
Expansion × above-median competition (district)		6.14	5.88	5.91	6.15	6.13	5.72	5.20
		(0.77)	(0.77)	(0.77)	(0.77)	(0.77)	(0.77)	(0.76)
Panel D. Suspensions ($N = 5{,}427{,}985$; children $= 1{,}225{,}713$; mean of $Y = 13.67$; SD of $Y = 34.35$)								
Expansion × above-median competition (school)	−1.47	−0.53	−0.29	−0.51	−0.53	−0.53	−0.65	−0.38
	(0.26)	(0.27)	(0.27)	(0.27)	(0.27)	(0.27)	(0.27)	(0.27)
Expansion × above-median competition (district)		−1.52	−1.08	−1.57	−1.52	−1.51	−1.33	−0.91
		(0.28)	(0.29)	(0.28)	(0.28)	(0.28)	(0.28)	(0.29)
Panel E. Absences ($N = 5{,}427{,}985$; children $= 1{,}225{,}713$; mean of $Y = 5.04$; SD of $Y = 5.79$)								
Expansion × above-median competition (school)	−0.26	−0.18	−0.18	−0.17	−0.18	−0.18	−0.18	−0.17
	(0.05)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
Expansion × above-median competition (district)		−0.12	−0.13	−0.15	−0.12	−0.13	−0.13	−0.16
		(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
District-level magnets and charters			X					X
District-level average salaries				X				X
School-level class sizes					X			X
Teacher characteristics						X		X
Peer effects in all domains							X	X

Notes: All regressions include student-by-school level FE and grade-by-school-year FE. Column 1 presents modified estimates based on panel A of Table 2 where we define the median competition split at the student rather than the school level. Column 2 presents estimates from a horse race between competition measured at school and at school district level. District-level competition is the student-weighted average of the school-level competition collapsed at school district in G1 by school level by school year level. Columns 3 to 7 further add control variables that are time varying (at an annual level) at either the school or district level. These are assigned based on G1-school- or school-district-by-year level. Column 3 controls for the district-level number of charter schools per 1,000 students and number of magnet schools per 1,000 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 teacher characteristics, including the fraction of teachers with 0 to 2 years of experience, the fraction of teachers with 3 to 5 years of experience, the fraction of teachers with 6 to 12 years of experience, the fraction of teachers with 13 or more years of experience, the fraction of White teachers, the fraction of Black teachers, and the fraction of Hispanic teachers. Column 7 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 8 includes all controls from columns 3 to 7 jointly. Information on class size is available for the years 2006–2007 to 2016–2017, information on charter and magnet schools is available for the years 2002–2003 to 2016–2017, information for average salaries is available for the years 2004–2005 to 2016–2017, information on teacher characteristics is available for the years 2002–2003 to 2011–2012, and information on predicted peer effects is available for the years 2002–2003 to 2013–2014 for math and averaged math and reading, for the years 2002–2003 to 2016–2017 for reading, and for the years 2002–2003 to 2011–2012 for suspensions and absences. To maintain constant sample size, we perform the following imputations for variables with missing values due to differential coverage of years: (i) if available, impute mean school level values, and (ii) if school-level information is not available, impute sample average. Standard errors are clustered at G1 school level.

Source: Author calculations

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 5.4 to 80.2.

One concern that these results may raise is that different district-level efforts, rather than the expansion of school competition per se, may drive our results. In columns 3 to 8 of Table 6 (and in even-numbered columns of online Appendix Table A6), 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 preprogram private school competition. If so, these expanding public school choice 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 1,000 students in the district. Our results are little changed for most outcomes, suggesting that other district-level forms of competition do not explain away the findings. 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 nonsignificant.

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 also could 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 (Florida Department of Education n.d.b) has minimal effects on the coefficients for either school- or district-level competition.

Similarly, in columns 5 to 7, we explore whether the inclusion of, respectively, school-level class-size information, teacher characteristics, or predicted peer ability/behavior levels introduced in Table 5 alters the results. Adding 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 8, we include all district- and school-level variables simultaneously. For most outcomes, column 8 produces results very similar to column 2, again with the exception that the estimates on suspension become nonsignificant with respect to school-level competition that 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 the 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 exposed to more 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.

VI. 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 mature and grow is virtually nonexistent. Here, we investigate this question using data from the state of Florida, where, over the course of our sample period, voucher program participation increased nearly sevenfold. We build on past research in that, to date, this is the largest voucher expansion in the United States, it represents the largest school voucher program in the country, and we can study it over 15 years. By contrast, 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 matures and scales up, students in public schools that faced higher initial competitive pressure levels see greater gains 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. These results cannot be explained away by changes in student composition, teacher composition, or school resources, to the extent that these are measurable in our data. This is despite the fact that we do find statistically significant effects on class sizes, teacher experience, and teacher demographics. Our results are also consistent with past work showing modest benefits to the initial introduction of voucher programs (e.g., Hoxby 2003; Figlio and Hart 2014; Egalite 2016; Egalite and Wolf 2016; Figlio and Karbownik 2016) while extending upon these findings to show the persistence and growth of these positive effects as the program matured and 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 on student outcomes. Thus, future work that looks at district-level responses (or peer networks within a school district) as potential mechanisms may be fruitful. 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 FTC scholarship program. However, we also observe improvements in outcomes for students unlikely to be eligible for vouchers, suggesting that benefits may come partially through generalized school improvements rather than via improvements targeted solely at voucher-eligible students.

Another important interpretive point is that our results reflect relative changes in outcomes in more versus less competitive areas as the voucher program matures and scales up, but policymakers may also be interested in the total effects of voucher program expansion. On the one hand, our estimates may represent a lower bound on the total effect of voucher program scale-up: to the extent that competition generally produces positive effects for public school students, we miss the gains that may accrue to students in relatively low-competition areas whose schools nonetheless experience *some* increase in competitive pressure. On the other hand, it is plausible that our results are at least partially driven by *declines* for students in low-competition areas. This could occur if, say, schools in high-competition areas hired high-quality principals or teachers away from schools in low-competition areas, leading to declines in the latter set of schools. In the most worrisome scenario, our results could be driven entirely by declines in low-competition areas, with test scores staying relatively stagnant in high-competition areas. That said, Florida's gains in National Assessment of Educational Progress scores generally outstripped national averages over this period (Florida Department of Education n.d.h, i, j, k), suggesting that a story of redistribution of achievement gains across schools with no overall improvement in the state is unlikely.

Our analyses include several important limitations as well. Of course, our results may be specific to the policy and competitive environment present in Florida. In addition, we only look at effects of voucher maturation and scale-up on public schools. Policymakers considering scaling up voucher systems would also want to consider the important question of the total effects of the growth of voucher programs on the system as a whole, including students who ultimately participate in voucher programs and their peers in private schools, especially given that effects on voucher participants could be negative (Figlio and Karbownik 2016; Mills and Wolf 2017; Waddington and Berends 2018; Abdulkadiroglu, Pathak, and Walters 2015). That said, because the number of students affected by school choice through the competition channel is far larger than the number of students who enroll, negative effects on voucher users would have to be very large—or enrollment would have to expand considerably—to outweigh the positive effects for students who remain in public schools.

REFERENCES

Abdulkadiroglu, Atila, Parag A. Pathak, and Christopher R. Walters. 2015. "School Vouchers and Student Achievement: Evidence from the Louisiana Scholarship Program." NBER Working Paper 21839.

Angrist, Joshua D., and Victor Lavy. 1999. "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement." *Quarterly Journal of Economics* 114 (2): 533–75.

Arsen, David and Yongmei Ni. 2008. *The Competitive Effect of School Choice Policies on Traditional Public Schools*. Boulder, CO: National Education Policy Center.

Autor, David, David Figlio, Krzysztof Karbownik, Jeffrey Roth, and Melanie Wasserman. 2019. "Family Disadvantage and the Gender Gap in Behavioral and Educational Outcomes." *American Economic Journal: Applied Economics* 11 (3): 338–81.

Bartik, Timothy J. 1991. *Who Benefits from State and Local Economic Development Policies?* Kalamazoo, MI: W. E. Upjohn Institute.

Böhlmark, Anders, and Mikael Lindahl. 2015. "Independent Schools and Long-Run Educational Outcomes: Evidence from Sweden's Large-Scale Voucher Reform." *Economica* 82 (327): 508–51.

Borusyak, Kirill, Peter Hull, and Xavier Jaravel. 2022. "Quasi-Experimental Shift-Share Research Designs." *Review of Economic Studies* 89 (1): 181–213.

Breining, Sanni, Joseph Doyle, David Figlio, Krzysztof Karbownik, and Jeffrey Roth. 2020. "Birth Order and Delinquency: Evidence from Denmark and Florida." *Journal of Labor Economics* 38 (1): 95–142.

Bristol, Travis J., and Javier Martin-Fernandez. 2019. "The Added Value of Latinx and Black Teachers for Latinx and Black Students: Implications for Policy." *Policy Insights from the Behavioral and Brain Sciences* 6 (2): 147–53.

Chakrabarti, Rajashri. 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 (5–6): 1371–93.

Chetty, Raj, John N. Friedman, Nathaniel Hilger, Emmanuel Saez, Diane W. Schanzenbach, and Danny Yagan. 2011. "How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project STAR." *Quarterly Journal of Economics* 126 (4): 1593–660.

Chingos, Matthew M. 2013. "Class Size and Student Outcomes: Research and Policy Implications." *Journal of Policy Analysis and Management* 32 (2): 411–38.

Dee, Thomas S. 1998. "Competition and the Quality of Public Schools." *Economics of Education Review* 17 (4): 419–27.

Dee, Thomas S. 2004. "Teachers, Race, and Student Achievement in a Randomized Experiment." *Review of Economics and Statistics* 86 (1): 195–210.

Dhuey, Elizabeth, David Figlio, Krzysztof Karbownik, and Jeffrey Roth. 2019. "School Starting Age and Cognitive Development." *Journal of Policy Analysis and Management* 38 (3): 538–78.

EdChoice. 2019. "School Choice in America Dashboard." School Choice. https://www.edchoice.org/school-choice/school-choice-in-america/# (accessed on February 12 2019).

Egalite, Anna J. 2016. "Competitive Impacts of Means-Tested Vouchers on Public School Performance: Evidence from Louisiana and Indiana." Program on Education Policy and Governance Working Paper 14-05.

Egalite, Anna J., and Patrick J. Wolf. 2016. "A Review of the Empirical Research on Private School Choice." *Peabody Journal of Education* 91 (4): 441–54.

Epple, Dennis, Richard E. Romano, and Miguel Urquiola. 2017. "School Vouchers: A Survey of the Economics Literature." *Journal of Economic Literature* 55 (2): 441–92.

Figlio, David, Jonathan Guryan, Krzysztof Karbownik, and Jeffrey Roth. 2014. "The Effects of Poor Neonatal Health on Children's Cognitive Development." *American Economic Review* 104 (12): 3921–55.

Figlio, David N., and Cassandra M. D. Hart. 2014. "Competitive Effects of Means-Tested School Vouchers." *American Economic Journal: Applied Economics* 6 (1): 133–56.

Figlio, David N., Cassandra M. D. Hart, and Krzysztof Karbownik. 2020. "Effects of Scaling Up Private School Choice Programs on Public School Students." NBER Working Paper 26758.

Figlio, David N., Cassandra M. D. Hart, and Krzysztof Karbownik. 2023. "Replication Data for: Effects of Maturing Private School Choice Programs on Public School Students." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/E171381V1.

Figlio, David N., and Krzysztof Karbownik. 2016. *Evaluation of Ohio's EdChoice Scholarship Program: Selection, Competition, and Performance Effects*. Washington, DC: Thomas B. Fordham Institute.

Florida Department of Education. 2009. *FTC Scholarship Program Quarterly Report, October 2009*. Tallahassee, FL: Florida Department of Education.

Florida Department of Education. 2013. *Fact Sheet on the Florida Tax Credit Scholarship Program, September 2013*. Tallahassee, FL: Florida Department of Education.

Florida Department of Education. 2018. *Fact Sheet on the Florida Tax Credit Scholarship Program, September 2018*. Tallahassee, FL: Florida Department of Education.

Florida Department of Education. n.d.a. "Class Size Reduction–Averages." Florida Department of Education. https://www.fldoe.org/finance/budget/class-size/class-size-reduction-averages.stml (accessed May 22, 2022).

Florida Department of Education. n.d.b. "Data Publication and Reports: Teacher Salary, Experience, and Degree Level." Florida Department of Education. https://www.fldoe.org/accountability/data-sys/edu-info-accountability-services/pk-12-public-school-data-pubs-reports/archive.stml (accessed May 10, 2022).

Florida Department of Education. n.d.c. "FCAT Historical." Assessments and Publications Archive. http://www.fldoe.org/accountability/assessments/k-12-student-assessment/archive/fcat/ (accessed February 3, 2019).

Florida Department of Education. n.d.d. "Florida Education Research Data Files." Florida Department of Education. (accessed April 25, 2018).

Florida Department of Education. n.d.e. "Florida School Grades Archives." Florida Department of Education. https://www.fldoe.org/accountability/accountability-reporting/school-grades/archives.stml (accessed May 17, 2022).

Florida Department of Education. n.d.f. “Florida Tax Credit Scholarships: Quarterly Reports. Fall Report Used for Each Year 2005–2018.” Florida Department of Education. https://www.fldoe.org/schools/school-choice/k-12-scholarship-programs/ftc/quarterly-reports.stml (accessed May 2, 2022).

Florida Department of Education. n.d.g. “Master School ID Files.” Florida Department of Education. https://eds.fldoe.org/EDS/MasterSchoolID/ (accessed May 22, 2022).

Florida Department of Education. n.d.h. “NAEP Florida, 1992–2015 NAEP Scores: Math, Grade 4.” Florida Department of Education. https://cdn.fldoe.org/core/fileparse.php/7667/urlt/Trends9215Gr4Math.pdf (accessed February 26, 2022).

Florida Department of Education. n.d.i. “NAEP Florida, 1992–2015 NAEP Scores: Math, Grade 8.” Florida Department of Education. https://www.fldoe.org/core/fileparse.php/7667/urlt/Trends9015Gr8Math.pdf (accessed February 26, 2022).

Florida Department of Education. n.d.j. “NAEP Florida, 1992–2015 NAEP Scores: Reading, Grade 4.” Florida Department of Education. https://cdn.fldoe.org/core/fileparse.php/7667/urlt/Trends9215Gr4Read.pdf (accessed February 26, 2022).

Florida Department of Education. n.d.k. “NAEP Florida, 1992–2015 NAEP Scores: Reading, Grade 8.” Florida Department of Education. https://www.fldoe.org/core/fileparse.php/7667/urlt/Trends9815Gr8Read.pdf, (accessed February 26, 2022).

Florida Department of Education. n.d.l. “Student Enrollment: State Level.” Florida Department of Education. https://edstats.fldoe.org/SASWebReportStudio/gotoReportSection.do?sectionNumber=1 (accessed January 31, 2019).

Florida Department of Health. n.d. “Florida Birth Record Data Files.” Florida Department of Health. (accessed April 25, 2018).

Florida Government. 2001. *Florida Statute 220.187: Credits for Contributions to Nonprofit Scholarship-Funding Organizations*. Tallahassee, FL: Florida Government.

Florida Government. 2010. *Florida Statute 1002.395: Florida Tax Credit Scholarship Program*. Tallahassee, FL: Florida Government.

Fredriksson, Peter, Björn Öckert, and Hessel Oosterbeek. 2013. “Long-Term Effects of Class Size.” *Quarterly Journal of Economics* 128 (1): 249–85.

Gershenson, Seth, Cassandra M. D. Hart, Joshua Hyman, Constance A. Lindsay, and Nicholas W. Papageorge. 2022. “The Long-Run Impacts of Same-Race Teachers.” *American Economic Journal: Economic Policy* 14 (4): 300–42.

Gilraine, Michael, Uros Petronijevic, and John D. Singleton. 2021. “Horizontal Differentiation and the Policy Effect of Charter Schools.” *American Economic Journal: Economic Policy* 13 (3): 239–76.

Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift, H. 2020. “Bartik Instruments: What, When, Why, and How?” *American Economic Review* 110 (8): 2586–624.

Greene, Jay P., and Marcus A. Winters. 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.

Grissom, Jason A., Emily K. Kern, and Luis A. Rodriguez. 2015. “The ‘Representative Bureaucracy’ in Education: Educator Workforce Diversity, Policy Outputs, and Outcomes for Disadvantaged Students.” *Educational Researcher* 44 (3): 185–92.

Harris, Douglas N., and Tim Sass. 2011. “Teacher Training, Teacher Quality, and Student Achievement.” *Journal of Public Economics* 95 (7–8): 798–812.

Hensvik, Lena. 2012. “Competition, Wages, and Teacher Sorting: Lessons Learned from a Voucher Reform.” *Economic Journal* 122 (561): 799–824.

Holt, Stephen B., and Seth Gershenson. 2017. “The Impact of Demographic Representation on Absences and Suspensions.” *Policy Studies Journal* 47 (4): 1069–99.

Hoxby, Caroline Minter. 1994. “Do Private Schools Provide Competition for Public Schools?” NBER Working Paper 4978.

Hoxby, Caroline Minter. 2003. “School Choice and School Productivity: Could School Choice Be a Tide That Lifts All Boats?” In *The Economics of School Choice*, edited by Caroline Minter Hoxby, 287–341. Chicago: University of Chicago Press.

Hsieh, Chang-Tai, and Miguel Urquiola. 2006. “The Effects of Generalized School Choice on Achievement and Stratification: Evidence from Chile’s Voucher Program.” *Journal of Public Economics* 90 (8–9): 1477–503.

Jabbar, Huriya. 2015. “‘Every Kid is Money’: Market-Like Competition and School Leader Strategies in New Orleans.” *Educational Evaluation and Policy Analysis* 37 (4): 638–59.

Jepsen, Christopher. 2002. "The Role of Aggregation in Estimating the Effects of Private School Competition on Student Achievement." *Journal of Urban Economics* 52 (3): 477–500.

Jepsen, Christopher, and Steven Rivkin. 2009. "Class Size Reduction and Student Achievement: The Potential Tradeoff between Teacher Quality and Class Size." *Journal of Human Resources* 44 (1): 223–50.

Karbownik, Krzysztof. 2020. "The Effects of Student Composition on Teacher Turnover: Evidence from an Admission Reform." *Economics of Education Review* 75: 101960.

Krueger, Alan B. 1999. "Experimental Estimates of Education Production Functions." *Quarterly Journal of Economics* 114 (2): 497–532.

Ladd, Helen F., and Lucy C. Sorensen. 2017. "Returns to Teacher Experience: Student Achievement and Motivation in Middle School." *Education Finance and Policy* 12 (2): 241–79.

Lankford, Hamilton, and James Wyckoff. 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, Edwin, and Sturla A. Løkken. 2020. "Long-Term Effects of Class Size in Compulsory School." *Journal of Human Resources* 55 (1): 309–48.

Lindahl, Mikael. 2005. "Home versus School Learning: A New Approach to Estimating the Effect of Class Size on Achievement." *Scandanavian Journal of Economics* 107 (2): 375–94.

Lindsay, Constance A., and Cassandra M. D. Hart. 2017. "Teacher-Student Race Match and Student Disciplinary Outcomes for Black Students in North Carolina." *Educational Evaluation and Policy Analysis* 39 (3): 485–510.

Mills, Jonathan N., and Patrick J. Wolf. 2017. "Vouchers in the Bayou: The Effects of the Louisiana Scholarship Program on Student Achievement after 2 Years." *Educational Evaluation and Policy Analysis* 39 (3): 464–84.

Muralidharan, Karthik, and Venkatesh Sundararaman. 2015. "The Aggregate Effect of School Choice: Evidence from a Two-Stage Experiment in India." *Quarterly Journal of Economics* 130 (3): 1011–66.

National Center for Education Statistics. 2013. "Table 205.50. Private Elementary and Secondary Enrollment, Number of Schools, and Average Tuition, by School Level, Orientation, and Tuition: Selected years, 1999–2000 through 2011-12." Digest of Education Statistics. https://nces.ed.gov/programs/digest/d13/tables/dt13_205.50.asp (accessed April 7, 2020).

National Center for Education Statistics. n.d. "Elementary/Secondary Information System, Common Core of Data." Institute of Education Sciences. https://nces.ed.gov/ccd/elsi/ (accessed May 10, 2022).

Nickell, Stephen. 1981. "Biases in Dynamic Models with Fixed Effects." *Econometrica* 49 (6): 1417–26.

Ridley, Matthew, and Camille Terrier. Forthcoming. "Fiscal and Education Spillovers from Charter School Expansion." *Journal of Human Resources*.

Rouse, Cecilia Elena, Jane Hannaway, Dan Goldhaber, and David Figlio. 2013. "Feeling the Florida Heat? How Low-Performing Schools Respond to Voucher and Accountability Pressure." *American Economic Journal: Economic Policy* 5 (2): 251–81.

Sacerdote, Bruce. 2014. "Experimental and Quasi-Experimental Analysis of Peer Effects: Two Steps Forward?" *Annual Review of Economics* 6: 253–72.

Sandström, F. Michael, and Fredrik Bergström. 2005. "School Vouchers in Practice: Competition Will Not Hurt You." *Journal of Public Economics* 89 (2–3): 351–80.

Totenberg, Nina, and Brian Naylor. 2020. "Supreme Court: Montana Can't Exclude Religious Schools from Scholarship Program." *National Public Radio*, June 30.

Urquiola, Miguel. 2016. "Competition among Schools: Traditional Public and Private Schools." In *Handbook of the Economics of Education*, edited by Eric A. Hanushek, Stephen Machin, and Ludger Woessmann, 209–37. Amsterdam: Elsevier

Waddington, R. Joseph, and Mark Berends. 2018. "Impact of the Indiana Choice Scholarship Program: Achievement Effects for Students in Upper Elementary and Middle School." *Journal of Policy Analysis and Management* 37 (4): 783–808.