Overall Effects of Targeted School Vouchers on College Enrollment and Graduation: Evidence from Ohio *

Matthew Chingos, Urban Institute

David N. Figlio, University of Rochester and NBER

Krzysztof Karbownik, Emory University and NBER

November 24, 2025

Abstract

Using a regression discontinuity design and administrative data from Ohio, we study the effects of eligibility for a school voucher program on college attendance and degree attainment. We find that voucher eligibility boosted the likelihood of students' college attendance and graduation, including at four-year and selective institutions. The results are most positive for the students who, based on observables, were relatively unlikely to attend or graduate from college. We also find some modest negative effects for students who were most likely to graduate from college.

JEL: H75, I21, I22, I28

Keywords: school choice, school vouchers, school competition

*Chingos: mchingos@urban.org; Figlio: david.figlio@rochester.edu; Karbownik: krzysztof.karbownik@emory.edu. This is an update of the findings initially presented in an unpublished Urban Institute research report entitled "The Effects of Ohio's EdChoice Voucher Program on College Enrollment and Graduation". We are grateful to the Ohio Department of Education and Workforce, and especially Eben Dowell, for providing us with the confidential individual-level data used in this study. We acknowledge research support from the Bill and Susan Oberndorf Foundation and the Shelter Hill Fondation. We thank conference and seminar participants at 2025 CesIfo Economics of Education Meeting, Emory University and Harvard University, as well as Breno Braga and Minseon Park, for helpful comments and feedback. Rachel Lamb provided excellent research assistance. The findings and opinions expressed herein are exclusively our own, and do not reflect the positions of the research funders or the Ohio Department of Education and Workforce. All errors are our own.

1 Introduction

There has been a tremendous amount of policy discussion in recent years in the USA regarding school vouchers, with 18 states having implemented universal school choice programs at the time of writing, most within the past few years, and the One Big Beautiful Bill Act passed in July 2025 promising a federal subsidy for school choice for families with income below 300 percent of the median gross income in their local area. At the time of writing, over one million children in the USA attend private schools under public school choice programs, and that number will surely increase considerably in the coming years as the recently-implemented policies take effect and mature. This policy development has far outpaced the research evidence, and to date we know relatively little about what happens to student outcomes in communities with increased private school choice, especially in the long-run. To the extent to which fewer financial resources affect student outcomes, this could hurt students staying in the public schools if funding is tied to pupil counts but costs are relatively constant. Voucher programs could also affect school sorting, which might lead to improved outcomes if there are better student-school matches with increased choice, but could lead to reduced outcomes for some depending upon the nature of peer effects. Finally, voucher programs could stimulate competition that might lead to better student outcomes in both public and private schools.

In this paper we study the overall effects on students – regardless of the sector in which they are educated – of Ohio's EdChoice school voucher program, at the time one of the largest targeted school voucher programs in the United States. Specifically, we ask the following research question: are college outcomes of students quasi-randomly assigned to being eligible for private school voucher better or worse from those of students ineligible for a voucher. This analysis is noteworthy in two ways: First, the design of Ohio's school voucher program lends itself to a regression discontinuity design and analysis of outcomes for all students eligible for the voucher. This is the policy relevant effect of vouchers as it asks whether it is an overall net positive to let vouchers into specific communities. Second,

the Ohio Department of Education and Workforce merged individual student records with data from the National Student Clearinghouse, permitting the first US study of the effects of targeted school voucher programs on longer-run outcomes through college completion, as opposed to K-12 outcomes or college attendance measures.

Our empirical approach is a regression discontinuity design based on Ohio's EdChoice Voucher Program eligibility rules. The program – enacted in 2005 with the first cohort of students using vouchers to attend private schools in 2006-2007 academic year – is an ideal environment for studying the overall effects of school vouchers for two reasons. First, the formula used to determine school-level voucher eligibility – accountability score below a specific threshold in two out of the three years – lends itself to a regression discontinuity design. Second, every student enrolled in a treated public school was immediately eligible for a school voucher. Thus, we compare college outcomes of *all* students in schools that barely missed the cutoff and become voucher eligible to *all* students in schools that barely made the cutoff and were not voucher eligible, irrespective if individual students decided to exercise choice or stay in the public sector.

We find evidence that Ohio's voucher program led to increased college attendance and graduation of voucher-eligible students, relative to just-ineligible students from schools that barely missed the cutoff based on their second-best PI measure. These results are present for all colleges, as well as four-year and selective institutions. The estimated effects of vouchers are driven by high school students rather than elementary and middle school students. The effects are most positive among students predicted to be least likely to attend/graduate from college based on their observables; indeed, we find some negative effects on college graduation for the students predicted most likely to graduate college. The results are robust to using alternative estimators, different bandwidths, donut specifications, and pass a placebo check.

Our work fills two knowledge gaps in the literature. First, this is the first study that speaks to the overall effects of school vouchers in the USA.Second, even among the US studies considering students who use vouchers very few look beyond high school outcomes

and no prior work investigated college outcomes of students who remain in public schools threatened by voucher competition, which in all programs is still the vast majority of the population. Furthermore, with the exception of Erickson et al. (2021) who use random lottery assignment, no prior work was able to use a credible quasi-experimental design such a regression discontinuity to follow students past high school.[1] Prior work either considers the effects of vouchers on the K-12 outcomes of public school students or compare voucher participants to non-participants. Figlio and Hart (2014) and Figlio et al. (2023) find that targeted school vouchers in Florida boosted public school student test scores, while Figlio and Karbownik (2016) find smaller positive effects in Ohio. Results vary in Indiana, where Adger and Felegi (2025) find positive test score effects for public school students, while Egalite and Catt (2025) find no effects on test scores or high school graduation. Another strand of literature presents mixed findings regarding outcomes of *participants* who use vouchers to attend private schools, where studies from Indiana, Louisiana, and Ohio find negative effects on participants' test scores (Abdulkadiroglu et al. (2018); Erickson et al. (2021); Figlio and Karbownik (2016); Mills and Wolf (2017); Waddington and Berends (2018)) but either no statistically significant effects on participants' college enrollment in Louisiana (Erickson et al. (2021)) or positive effects on their enrollment and degree attainment in Florida and Milwaukee (Chingos et al. (2019)). We build upon this literature by studying college outcomes – enrollment, persistence, and degree attainment – for *all students eligible for a school voucher*, regardless of the sector in which they are educated. Given that the Ed Choice voucher was universal from the start for a given population, as opposed to other large targeted programs such as Florida and Indiana where eligibility was based upon family

[1]There exists some long-run international evidence from Chile and Sweden documenting overall effects. Bravo et al. (2010) present the results from a structural life-cycle model and survey evidence suggesting that Chile's universal voucher increased educational attainment but had no effect on earnings, while Correa et al. (2021) find that a subsequent targeted voucher in Chile boosted initial earnings. Böhlmark and Lindahl (2015) show that Sweden's voucher program increased educational attainment and college going while Edmark et al. (2024) find evidence that it affected employment rates and occupational choice. International evidence might not generalize to the US given that in the US relatively smaller share of students use vouchers and the US schools are generally not allowed to practice selective admissions of applying students. This institutional differences are above and beyond differences in culture, wealth, income inequality, or provision of public goods.

income before they became universal, it is uniquely suited to studying the policy-relevant overall effects of voucher where all students in a geography become eligible.

2 Background and Methods

2.1 Ohio's EdChoice Voucher Program

Ohio enacted what is now called the Educational Choice Scholarship Program (EdChoice) in 2005 and students first used scholarships from this program to attend private schools beginning in the 2006-07 school year.[2] The program was directly tied to school accountability system in Ohio which was introduced much earlier in 1997. Specifically the voucher eligibility criteria were tied to school performance index (PI score), one of the accountability metrics available, which assigns scores to each school based on all tested students with differential weights for students of different performance levels. The index is on a scale of 0 to 120 points with 0 implying that no students are tested.

The voucher program has expanded significantly over the years. At first, eligibility was limited to students who previously attended public schools that received the lowest rating under the state's accountability system for three years in a row. In 2006, the legislature expanded eligibility to students attending schools that received either of the two lowest ratings in at least two out of the three years preceding the academic year in which the determination was made.[3] The EdChoice program expanded rapidly over the years, from

[2]See Figlio and Karbownik (2016) for a more detailed history of the EdChoice program and citations to original sources.

[3]This means that the three relevant PI scores determining eligibility for school year 2007-08 are from school years 2003-04, 2004-05, and 2005-06. There are several exceptions from this rule. First, students in a given school are ineligible for vouchers despite their school failing the accountability index cutoff in two of the three preceding years if the school in the current year is nonetheless labeled as excellent or effective (school years 2007-08 to 2011-12) or received letters A or B on PI score and VA progress (school years 2012-2013 to 2015-16) or received letters A or B on PI score and letter A on VA progress (school year 2016-17). There is a separate exception provision related to high schools introduced in SY 2012-13 that requires at least 75% 4-year graduation rate. Second, students in a given underperforming school are likewise ineligible if their school district has overall designation of excellent or effective (school years 2007-08 to 2011-12) or letters A or B on PI and VA progress (school years 2012-2013 to 2015-16) or letters A or B overall and letter A on VA progress (school years 2016-17). In our empirical sample XX% of expected-to-be-treated schools qualified

3,071 students in 2006-07 to a total of 62,289 in the original and expansion programs in 2022-23 and 129,541 in 2023-24 following the elimination of means testing. As of spring 2014, the last year covered by our college graduation analysis, 20,263 students received a voucher through the original EdChoice program. This is also the year that Ohio launched EdChoice Expansion, which provides scholarships to economically disadvantaged students regardless of their public school's accountability rating. This program gradually expanded (one grade per year) from kindergarten students in 2013-14 to all K-12 students in 2020-21 but even by 2022-23 it enrolled fewer students than the original EdChoice program (see Figure A1).[4] We make two main sample restrictions in our analysis. First, we exclude Cleveland, Ohio's second-largest city, from the analysis because all K-12 students who reside in the Cleveland Metropolitan School District are eligible for a voucher, and students must take the Cleveland scholarship if they are deemed eligible rather than Ohio's statewide EdChoice scholarship. Second, we drop students participating in two small voucher programs for students with special needs, the Autism and Jon Peterson Special Needs Scholarship Programs.

During our time period of interest, students became eligible for EdChoice vouchers if their schools were rated as Academic Emergency or Academic Watch in two of the prior three academic years.[5] One the the school was flagged for eligibility all its students had the right to a voucher, however, students had to first find a private school that was accepting vouchers before they could use it.[6] Schools could be rated as "Continuous Improvement" or better if they met the state's standard for adequate yearly progress, met a majority of the school's standards, or achieved a "performance index" (PI) at or above 80 (out of 120 points), based on the degree to which students in the school moved up the proficiency spectrum on

for either of these exemptions.

[4]Since 2023-24 all students in Ohio are eligible for EdChoice Expansion regardless of income, though families with incomes above 450 percent of the federal poverty level receive a smaller scholarship.

[5]Between 2003-04 and 2011-12 school and districts were rated as Academic Emergency, Academic Watch, Continuous Improvement, Effective, Excellent, and Excellent with Distinction. Starting in 2012-13 this scale transitioned to letter grades of F, D, C, B, A with A combining both Excellent and Excellent with distinction categories. Therefore post 2011-12 eligibility was based on F and D vs. C grades. Throughout the paper we use the initial terminology to describe schools above and below the cutoff.

[6]This is different from e.g., Florida or XX where students first get a voucher and then look for private schools to enroll in.

state accountability tests. In practice, most schools met "Continuous Improvement" through the PI standard. For example, only 23.7 percent of (rated) schools with PI between 80 and 84.9 in 2007 met adequate yearly progress and 4.6 percent met a majority of a school's standards, but all were designated as "Continuous Improvement". We make use of the PI score as a running variable for our analysis.

2.2 Data and Sample

We make use of individual-level student records from the Ohio Department of Education and Workforce mateched to data from the National Student Clearinghouse (NSC). The data file includes all students in K-12 public schools and all students who used vouchers to attend private schools (but not students who attended private schools without a voucher or who were educated at home). The K-12 data include test scores from grades 3-8, demographic information, and geographic information about schools attended. The NSC match, which was last conducted in fall 2020, includes information about student enrollment and graduation from nearly all US colleges and universities.

We focus on students first attending public schools in grades 3-12, and we limit our sample to students observed at least once in public school between 2007-08 and 2016-17 (2013-14 for college graduation measures), so we can measure outcomes as (1) college enrollment within two years of expected high school graduation and (2) college graduation within six years of expected high school graduation. Our ultimate sample sizes at our preferred bandwidth (12 points on the PI) are 412,557 students (including 6,616 voucher users) for studying college enrollment and 137,455 students (including 1,533 voucher users) for studying graduation.[7] Based on these numbers, it is clear that the vast weight of our estimated effects of voucher eligibility will be among students educated exclusively in public schools.

[7]In some of our model specifications, we control for a set of covariates, including baseline test scores, which are missing for 9% of the elementary/middle school grade sample and approximately 13% of the high school grade sample. We show results that both include and exclude controls, as well as those that limit the analysis to students for whom we observe baseline test scores, in order to ascertain robustness to the treatment of baseline test scores. Our results are robust.

Schools at the margin of eligibility for EdChoice vouchers are different from Ohio schools overall. Table A1 presents descriptive statistics for the college attendance (columns 1 to 4) and college graduation (columns 5 to 8) samples.[8] We further divide these samples into schools with running variable below our preferred bandwidth (columns 2 and 6), in our preferred bandwidth (columns 3 and 7), and above our preferred bandwidth (columns 4 and 8). Compared to the full population, schools we consider are more likely to educate minority students, impoverished students, and lower performing students. On the other hand, the composition of schools we consider are representative of the overall Ohio population with respect to sex, English Language Learner and disability status appear. Consistent with the pattern that lower-income and lower-performing students are less likely to attend and complete college than the population as a whole, college outcomes in our sample are worse compared to all Ohio schools with attendance at and graduation from 4-year institutions lower by 18 and 16 percentage points, respectively. At the same time, these outcomes are higher by factors of 2.1 and 4.6, respectively, compared to school with PI score < 68.

2.3 Models

The institutional setting described in Section 2.1 lends itself to a regression discontinuity design evaluation with the second-best PI score used as a running variable. It allows us to compare college outcomes of students in eligible schools with college outcomes of students in ineligible schools irrespective of whether they use the voucher or not. In this setting, assuming that regression discontinuity identifying assumption holds, the latter group forms a valid quasi-experimental counterfactual to the treated group and computes overall total effects of voucher availability. Note that this is different from prior work that varied eligibility within schools and thus was better suited for separately estimating either effects on takers or competitive effects.

[8] Even the full sample with valid PI scores is somewhat positively selected based on test scores. The reason for this is that we require schools to be operational and assessed for at least 5 years in order to compute the second-best PI score, our running variable, and these schools are slightly positively selected compared to all schools with a least one assessment.

Given our setting, we estimate the following regressions:

$$Y_{is} = \alpha + \beta_1 1(\text{PI} < 80)_{st} + \beta_2 \text{PI}_{st} + \beta_3 1(\text{PI} < 80)_{st} \times \text{PI}_{st} + \omega_{gt} + [\Delta X_{it}] + \varepsilon_{ist} \qquad (1)$$

where Y_{is} are college outcomes (attendance or graduation) of an individual i who attended school s, PI_{st} is the second-best PI score of school s in year t, and $1(\text{PI} < 80)_{st}$ is an indicator variable that takes value 1 if the PI score of school s in year t is below 80 which was a threshold for being placed on "Academic Watch" and turning on students' eligibility for vouchers.[9] Note that in this regression we observe the outcome for each student only once but the treatment is stacked at school-by-year level and students can be treated over multiple years and in different schools. For this reason, we also include grade-by-year fixed effects ω_{gt}. In select specifications we further add individual level controls, X_{it}, including indicators for race/ethnicity, gender, poverty, being an English language learner, and having a disability, as well as pre-treatment test scores in reading and mathematics. Standard errors, ε_{ist}, are clustered at the school level, which is the most conservative approach, but in Section 3.2 we explore alternative adjustments.

Our coefficient of interest in Equation 1 is β_1 which represents the average difference in college attendance/graduation rates of students attending schools below versus above the voucher eligibility cutoff. This is a reduced form estimate which we estimate using our preferred bandwidth of $-12 \leq \text{PI} \leq 12$ points, an upper bound of optimal bandwidth indicated by the data-driven procedure of Calonico et al. (2014) across multiple outcomes.

An alternative would be to perform an instrumental variables estimation where we instrument school eligibility with $\text{PI} < 80$ cutoff in the same bandwidth of $-12 \leq \text{PI} \leq 12$ points. This amounts to a scaling exercise with a first stage in our application ranging from

[9]Recall that not all schools with $PI < 80$ are on "academic watch", as schools can be considered in "Continuous Improvement" if they meet adequate yearly progress goals or (in very rare instances) otherwise meet the majority of standards expected of the school, so this is necessarily a fuzzy regression discontinuity design.

0.371 (high school grades for college graduation outcomes) to 0.540 (elementary and middle school grades for college attendance outcome). These results are presented in Table A2. We present graphical evidence for the first stage in Figure A5.

The identifying assumption for Equation 1 is that potential outcomes are smooth across the PI cutoff of 80. Although this is not testable we present two sets of results in support of this assumption: (1) smoothness of the density distribution and (2) balance of covariates at the cutoff. First, Figure A2 plots the density of schools around PI score = 80 – the eligibility cutoff – and suggests that distribution of schools is smooth across the cutoff with p-value of 0.170 based on the test proposed by Calonico et al. (2014). Second, Table 1 presents covariate balance analyses based on Equation 1 where the dependent variables are socio-demographic characteristics (columns 1 to 7), prior test scores (columns 8 and 9), and an omnibus index (column 9). Out of 60 estimates we find 6 statistically significant at least at 10% level which is in line with what we'd expect to find by change. Nonetheless, we acknowledge that estimates for an indicator for being Black and an indicator for being Hispanic are large even if not always statistically significant. We address this potential problem in two ways. First, we show that our estimates are largely stable whether we include covariates or not and irrespective of what covariates we include. Second, we show that our results are similar when we exclude predominantly Black schools from the sample which generate the imbalance in the aforementioned covariates. Beyond the racial/ethnic indicators our test supports the smoothness of covariates assumption with estimates for other sociodemographic characteristics being small and the predicted index never exceeding an effect size of 5.4% imbalance. Overall, we conclude that our identifying assumption is likely to hold.

3 Results

3.1 Main results

Table 2 presents the main results for our preferred model specification. We estimate that voucher eligibility increases the probability of college attendance within two years by 3.4 percentage points (7 percent of the control mean) if we pool voucher eligibility across all grades in our model in which we control for covariates.[10] These results are driven completely by voucher eligibility in high school grades: The estimated effect of voucher eligibility in elementary/middle school grades is negligible (across all outcome variables), but the estimated effect of voucher eligibility in high school is 8.8 percentage points (20 percent). The positive effects are further present for both enrollment in 4-year institutions (at 6.9 percentage points, or 26 percent) and 4-year selective institutions (at 5.1 percentage points, or 32 percent). This suggests that most of the college enrollment effects are concentrated in 4-year, rather than 2-year institutions, a finding we have confirmed in unreported results.

We find evidence that eligibility for the EdChoice voucher not only increases the likelihood of college attendance, but also the likelihood of college graduation within six years. We estimate that voucher eligibility in high school increases the rate of college graduation by 3.9 percentage points (25 percent), four-year institution graduation by 2.9 percentage points (28 percent), and selective institution graduation by 1.7 percentage points (23 percent). Again, we find zero evidence of voucher eligibility effects for elementary/middle grades across any college outcome measure.

3.2 Robustness

We seek to confirm that our main results are robust to a wide range of alternative methodologies, control-group construction approaches, and samples. We begin with Tables A3 and A4

[10] Although modestly smaller in almost all specifications, the estimates with and without controls are comparable. This is consistent with balance of covariates documented in Table 1.

where we test quadratic specification (column 2), excluding imputed test scores (column 3), dropping baseline test score controls given that we impute them for about 9% of the elementary/middle school sample and about 13% of the high school sample (column 4), and excluding majority Black schools given the imbalance issue discussed in Section 2.3 (column 5). Our results are unaffected by these permutations. We then present results from non-parametric estimation proposed by Calonico et al. (2014) in columns 6 and 7. These results tend to be similar to our local-linear estimates in terms of both magnitudes and statistical significance, though for college graduation, they are more noisy. We also verify that our stacking across multiple grades does not affect our results. To do so, in column 8, we choose a random observation for each student so that we observe them only once in the data. The results are largely unaffected. Finally, although our college graduation sample is based only on data from before the EdChoice Expansion program, the same is not true for the college attendance sample. Thus, in column 9 of Table A3 we restrict estimation to years prior to EdChoice Expansion introduction. The results remain similar.

We also tested the stability of our estimates to different bandwidth choices (between $77 \leq$ PI score ≤ 83 and $56 \leq$ PI score ≤ 104) as well as donut regression (excluding observations close to the cutoff with $79.5 \leq$ PI score ≤ 80.5 to $77.5 \leq$ PI score ≤ 82.5). The former ensures that our results are not driven by an arbitrary (albeit consistent with bandwidths suggested by Calonico et al. (2014)) choice of sample around the policy cutoff while the latter deals with any potential manipulation around the cutoff which we might be worried about given the estimates for Black students presented in Table 1. Figure A3 tests the different bandwidths and Figure A4 tests the different donuts. Except for very small bandwidths our pooled and high school estimates are always positive and statistically distinct from zero at conventional levels. Furthermore, the effect sizes are very stable in the pooled sample but decline for the high school sample, especially for the college attendance outcomes. Nonetheless, ever at very large bandwidths they imply about 10% effects. For elementary and middle school sample the estimates become positive and statistically at large

bandwidths but we treat these findings with caution given that they are not consistently positive or statistically significant at smaller bandwidths. Turning to Figure A4 the results are very consistent: no estimate is meaningfully affected by excluding observations close to the cutoff. This gives us more confidence that any manipulations detected in Table 1 are inconsequential. Thus, we conclude that our results are robust to alternative bandwidth selections and donut samples.

In our main specification we clustered our standard errors at school level since this is our treatment unit. At the same time, it is unclear how to adjust standard errors in regression discontinuity designs, especially when the running variable is discrete and has masspoints (Kolesar and Rothe, 2018). Therefore, in Table A5 we test alternative ways for calculating standard errors. We in turn employ heteroskedasticity robust errors, cluster at running variable, cluster at student level (since we have repeated observations per student in our sample), cluster two-way at student and running variable, and cluster at school-by-year level since this is the level of our experiment. Irrespective of the approach, our main results remain statistically significant at conventional levels and they suggest that clustering at school level is almost always the most conservative approach to standard errors. We present two more computations in Table A5. First, we use the randomization inference propose by Young (2019). Second, given that our outcomes are correlated, we adjust for multiple hypotheses testing as advised by Romano and Wolf (2005). Neither approach is consequential for our findings. Thus, we conclude that our main results are robust to alternative forms of inference.

3.3 Heterogeneity

We next investigate whether the estimated effects of voucher eligibility vary depending upon the likelihood that a student will attend college, based on their covariates. We predict college attendance/graduation based upon student gender, race/ethnicity, the fraction of school years a student is in poverty, the fraction of school years a student is enrolled in an English language learner program, the fraction of school years a student has a disability

flag, and average baseline reading and math scores across all grades. These variables explain 21 percent of the variation in college attendance and 26 percent of the variation in college graduation. We then split the sample into quintiles based on predicted college outcomes.

Figure 1 presents heterogeneity results by quintile of predicted college outcomes. We observe that, for the college attendance outcomes, estimated effects of voucher eligibility are generally pretty similar across groups; or perhaps somewhat smaller for those most likely to attend college. On the other hand, for college graduation, the estimated heterogeneous effects of voucher eligibility are more pronounced: The estimated positive effects of voucher eligibility are concentrated among students relatively unlikely to graduate from college, and the estimated effects are less positive for groups predicted to be more likely to graduate from college. Among students predicted to be most likely to graduate from college, the estimated effects of voucher eligibility in high school on college graduation outcomes are *negative* and statistically distinct from zero for four-year institution graduation outcomes.[11] Therefore, while the weight of the results suggest positive overall effects of voucher eligibility and positive – or at least non-negative – estimated effects across a wide range of subgroups, there is evidence that the highest-performing students, conditional on being in a relatively low-performing school, may be negatively affected by voucher eligibility. Understanding whether this finding is due to changing school behaviors (e.g., focusing on different groups of students) or changing student behaviors (e.g., choosing different types of colleges) is a question beyond the scope of the present study, but it does present an important cautionary note in an otherwise positive story about the overall effects of a targeted school voucher for students attending struggling schools.[12]

[11] Appendix Tables A4 (college attendance) and A5 (graduation) present heterogeneity by gender, race/ethnicity, poverty status, test score history, and disability status. In the case of high school voucher eligibility, we find across-the-board positive estimates that are generally statistically distinct from zero for the college attendance outcomes. Subgroup estimates are much less likely to be statistically distinct from zero in the case of college graduation, though the evidence indicates that the results are concentrated with students in poverty rather than those not in poverty, as well as students without a disability rather than those with a disability. We observe no non-trivial negative estimates in this sample split.

[12] There exists survey evidence from other settings that schools change their programs and practices in response to school voucher and accountability pressure (e.g., Rouse et al. (2013)).

4 Conclusions

Our findings indicate that Ohio's EdChoice school voucher program had meaningful and lasting positive effects on the primarily low-income and relatively low-performing students who were eligible for them – regardless of whether or not they used them (as the vast majority of eligible students continued to be educated in public schools). EdChoice-eligible students were more likely to attend and graduate from college, including four-year and selective institutions, than were students attending schools that just barely missed the eligibility threshold, and these results are stronger when the student is eligible for a voucher in high school versus elementary/middle school. The results are concentrated in the eligible students who were least likely to attend college, as predicted based on a robust set of observables. Indeed, the negative college graduation results for the students most likely to attend and succeed in college are an important counterpoint to this finding.

An important limitation of any study that examines long-run outcomes is that the results necessarily reflect the conditions of the voucher program at the time, rather than the voucher program as it currently exists. In this case, the most recent EdChoice participants included in the analysis joined the program more than six years ago when it was still largely targeted to students at low-performing public schools or low-income students. As a result, the findings may not accurately forecast the impacts of newer programs that are open to all students and include expenses other than private school tuition (e.g., education savings accounts).

Understanding who participates in private school choice programs today and how students eligible for such programs fare in school and beyond will be important to evaluate and improve these programs. But they will be harder to study for several reasons. Most importantly, in universal voucher programs the identification of voucher treatment effects is necessarily less clean than the regression discontinuity approach that we can bring to bear in this study. A shift-share type instrument that identifies voucher effects off of pre-program characteristics (a la Figlio et al. (2023)) may help to identify differential effects of the voucher program, if (as in their Florida exercise) pre-policy geographic variation was orthogonal to

many potential confounds. Second, the new universal programs won't have baseline test scores and other key factors for many students who never enroll in public school that would allow researchers to study heterogeneous effects of vouchers – something we observe to be important in the present context. And, of course, it will take many years before students who became eligible for universal vouchers – some of which have not yet even begun enrollment – will have had sufficient time to attend and graduate from college. Therefore, it will take quite some time before we have solid evidence on the types of universal school voucher programs being implemented in the USA today.

In conclusion, we see reasons for both optimism and caution. The case for optimism is that our results indicate that Ohio's targeted school voucher, based on school performance and with eligibility strongly skewed toward low-income and low-performing students, improved important long-run outcomes for students eligible for the voucher, regardless of the sector that students attended. That said, the significant differences between the targeted program that produced this encouraging evidence and the universal programs currently being scaled up across the country mean that more evidence is needed to verify that these strong results will continue, or change dramatically as the scale and scope of vouchers and their potential consequences for public and private schools alike changes. Furthermore, the potentially negative estimated effects of the voucher for the set of eligible students most likely to be successful in college suggests that more work needs to be done to understand how schools and students change their behaviors as a consequence of these types of school vouchers.

References

Abdulkadiroglu, Atila, Parag Pathak, and Christopher Walters (2018) "Free to choose: Can school choice reduce student achievement?" *American Economic Journal: Applied Economics*, Vol. 10, No. 1, pp. 175–206.

Adger, Chandon and Brianna Felegi (2025) "Supply-side responses in school choice," Unpublished.

Anderson, Theodore and Herman Rubin (1949) "Estimation of the parameters of a single equation in a complete system of stochastic equations," *Annals of Mathematical Statistics*, Vol. 20, No. 1, pp. 46–63.

Böhlmark, Anders and Mikael Lindahl (2015) "Independent Schools and Long-run Educational Outcomes: Evidence from Sweden's Large-scale Voucher Reform," *Economica*, Vol. 82, No. 327, pp. 508–551.

Bravo, David, Sankar Mukhopadhyay, and Petra Todd (2010) "Effects of school reform on education and labor market performance: Evidence from Chile's universal voucher system," *Quantitative Economics*, Vol. 1, No. 1, pp. 47–95.

Calonico, Sebastian, Matias D. Cattaneo, and Rocio Titiunik (2014) "Robust data-driven inference in the regression-discontinuity design," *The Stata Journal*, Vol. 14, No. 4, pp. 909–946.

Chingos, Matthew, Daniel Kuehn, Tomas Monarrez, Patrick Wolf, John Witte, and Brian Kisida (2019) "The effects of means-tested private school choice programs on college enrollment and graduation," *Urban Institute report.*

Clark, Melissa, Philip Gleason, Christina Tuttle, and Marsha Silverber (2015) "Do charter schools improve student achievement?" *Educational Evaluation and Policy Analysis*, Vol. 37, No. 4, pp. 419–436.

Correa, Juan, Francisco Parro, and Rafael Sanchez (2021) "The effect of school voucher spending on initial earnings," *IZA discussion paper 14552.*

Edmark, Karin, Iftikhar Hussain, and Carla Haelermans (2024)) "Unpacking the impact of voucher schools: Evidence from Sweden," *IFAU working paper 2024:17.*

Egalite, Anna and Andrew Catt (2025)) "Effects of the Indiana Choice Scholarship Program on public school students' achievement and graduation rates," *AERA Open*, Vol. 11, No. 2.

Erickson, Heidi, Jonathan Mills, and Patrick Wolf (2021) "The effects of the Louisiana Scholarship Program on student achievement and college entrance," *Journal of Research on Educational Effectiveness*, Vol. 14, No. 4, pp. 861–899.

Figlio, David and Cassandra Hart (2014) "Competitive effects of means-tested school vouchers," *American Economic Journal: Applied Economics*, Vol. 6, No. 1, pp. 133–156.

Figlio, David, Cassandra Hart, and Krzysztof Karbownik (2023) "Effects of Maturing Private School Choice Programs on Public School Students," *American Economic Journal: Economic Policy*, Vol. 15, No. 4, pp. 255–294.

Figlio, David and Krzysztof Karbownik (2016) "Evaluation of Ohio's EdChoice Scholarship Program: Selection, Competition, and Performance Effects," Fordham Institute Report.

Kolesar, Michal and Christoph Rothe (2018) "Inference in Regression Discontinuity Designs with a Discrete Running Variable," *American Economic Review*, Vol. 108, No. 8, pp. 2277–2304.

Mills, Jonathan and Patrick Wolf (2017) "Vouchers in the Bayou: The effects of the Louisiana Scholarship Program on student achievement after 2 years," *Educational Evaluation and Policy Analysis*, Vol. 39, No. 3, pp. 464–484.

Romano, Joseph and Michael Wolf (2005) "Exact and approximate stepdown methods for multiple hypothesis testing," *Journal of the American Statistical Association*, Vol. 100, No. 469, pp. 94–108.

Rouse, Cecilia, 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*, Vol. 5, No. 2, pp. 251–281.

Waddington, Joseph and Mark Berends (2018) "Impact of the Indiana Choice Scholarship Program: Achievement effects for students in upper elementary and middle schools," *Journal of Policy Analysis and Management*, Vol. 37, No. 4, pp. 783–808.

Young, Alwyn (2019) "Channeling Fisher: Randomization Tests and the Statistical Insignificance of Seemingly Significant Experimental Results," *Quarterly Journal of Economics*, Vol. 134, No. 2, pp. 557–598.

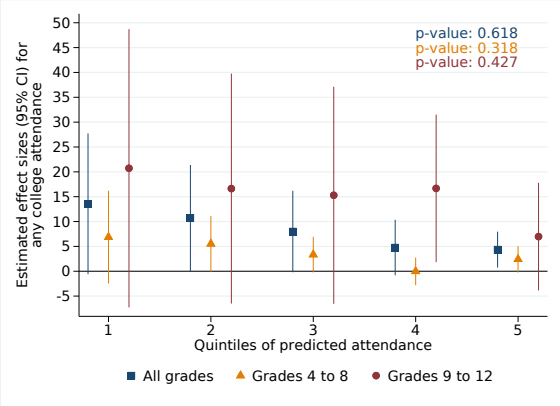
5 Figures

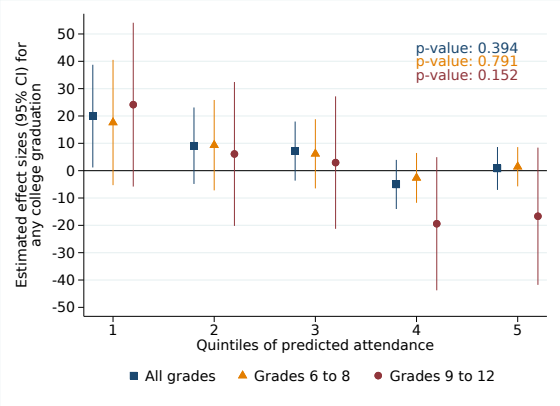
Figure 1: Heterogeneity by deciles of predicted attendance

I. College attendance

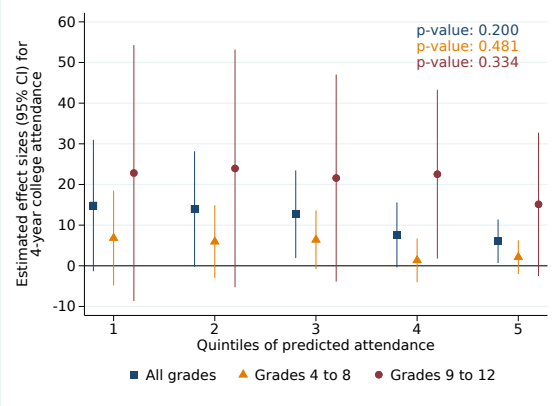
II. College graduation

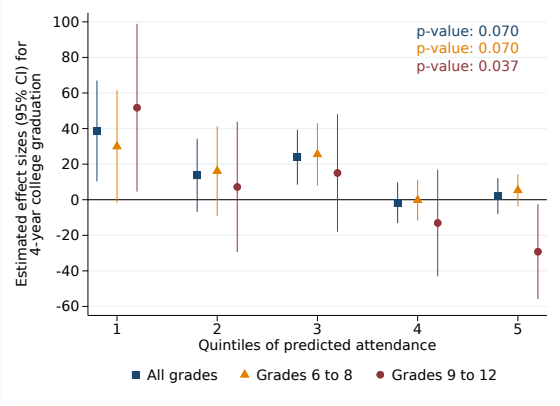
Panel A. Any



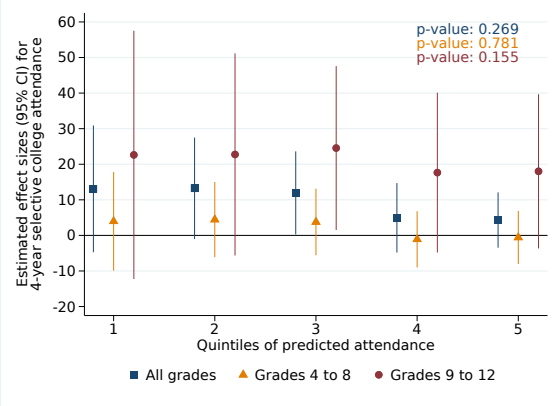


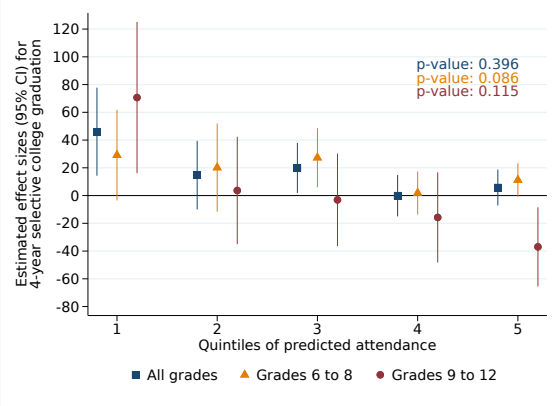
Panel B. 4-year institution





Panel C. Selective 4-year institution





Notes: These figures present results of the heterogeneity analyses based on estimating Equation 1 separately for quintiles of predicted college attendance (column I) and college graduation (column II). Point estimates and 95% confidence intervals are presented. We obtain these predicted probability based on logit regression of college attendance or college graduation on indicators for being female, Black, Hispanic, other racial/ethnic category (White is the omitted category), fraction of years being in poverty, fraction of years not being classified as ELL, fraction of years not being disabled, average test scores in reading and average test scores in mathematics across all observed grades. There is a single observation per student in these logit regression and R^2 from these regressions are 0.21 and 0.26 for attendance and graduation, respectively. Panel A presents results for any college attendance/graduation, panel B for 4-year college attendance/graduation, and panel C for 4-year selective college attendance/graduation as defined by Barron's Selectivity Index of 1 to 3. Navy scatters present results for all grades (4 to 12 for college attendance and 6 to 12 for college graduation samples), orange scatters results for grades 4 to 8 (college attendance sample) and 6 to 8 (college graduation sample), and maroon scatters results for grades 9 to 12. P-values reported in the graphs are based on fully saturated model that tests if the displayed coefficients are jointly equal to zero. Standard errors clustered at school level.

6 Tables

Table 1: Balance test

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Female	Minorities			In poverty	Non ELL	Not disabled	Baseline scores		Predicted outcome
		Black	Hispanic	Other				Reading	Math	
Panel A. Full sample attendance outcomes										
PI score < 80	-0.012	0.124**	-0.002	-0.003	0.014	0.006	-0.014*	-0.004	-0.010	0.009
	(0.008)	(0.049)	(0.008)	(0.008)	(0.024)	(0.013)	(0.008)	(0.025)	(0.020)	(0.006)
Mean of Y	0.495	0.226	0.039	0.061	0.624	0.975	0.829	-0.248	-0.285	0.483
Observations	1,301,077	1,301,077	1,301,077	1,301,077	1,301,077	1,301,077	1,301,077	1,301,077	1,301,077	1,301,077
Panel B. Grades 4 to 8 attendance outcomes										
PI score < 80	-0.007	0.094**	0.004	0.000	0.004	-0.006	-0.007	-0.028	-0.022	0.005
	(0.004)	(0.045)	(0.010)	(0.007)	(0.016)	(0.009)	(0.007)	(0.022)	(0.023)	(0.007)
Mean of Y	0.490	0.162	0.032	0.057	0.566	0.980	0.837	-0.154	-0.176	0.503
Observations	864,416	864,416	864,416	864,416	864,416	864,416	864,416	864,416	864,416	864,416
Panel C. Grades 9 to 12 attendance outcomes										
PI score < 80	-0.021	0.168	-0.022*	-0.015	0.013	0.035	-0.029	0.054	0.018	0.018*
	(0.018)	(0.109)	(0.013)	(0.017)	(0.053)	(0.032)	(0.018)	(0.056)	(0.036)	(0.009)
Mean of Y	0.505	0.348	0.052	0.070	0.733	0.966	0.813	-0.426	-0.492	0.444
Observations	436,661	436,661	436,661	436,661	436,661	436,661	436,661	436,661	436,661	436,661
Panel D. Full sample graduation outcomes										
PI score < 80	-0.004	0.050	-0.004	-0.002	-0.018	0.003	-0.011	0.001	-0.010	0.002
	(0.008)	(0.058)	(0.008)	(0.007)	(0.025)	(0.010)	(0.008)	(0.030)	(0.028)	(0.005)
Mean of Y	0.497	0.217	0.032	0.052	0.557	0.979	0.831	-0.207	-0.249	0.205
Observations	344,420	344,420	344,420	344,420	344,420	344,420	344,420	344,420	344,420	344,420
Panel E. Grades 6 to 8 graduation outcomes										
PI score < 80	-0.001	0.097	0.006	0.009	0.009	-0.012	-0.006	-0.042	-0.020	-0.002
	(0.008)	(0.060)	(0.008)	(0.008)	(0.029)	(0.011)	(0.010)	(0.030)	(0.031)	(0.006)
Mean of Y	0.488	0.119	0.024	0.043	0.465	0.987	0.840	-0.090	-0.114	0.237
Observations	197,853	197,853	197,853	197,853	197,853	197,853	197,853	197,853	197,853	197,853
Panel E. Grades 9 to 12 graduation outcomes										
PI score < 80	-0.003	-0.018	-0.022	-0.020	-0.058	0.027*	-0.013	0.067	0.001	0.009
	(0.015)	(0.113)	(0.015)	(0.014)	(0.045)	(0.016)	(0.016)	(0.058)	(0.052)	(0.009)
Mean of Y	0.509	0.342	0.042	0.062	0.674	0.968	0.819	-0.357	-0.422	0.166
Observations	146,567	146,567	146,567	146,567	146,567	146,567	146,567	146,567	146,567	146,567

Note: This table presents balancing checks based on estimating Equation 1 for outcomes listed in the column headers. Panels A to C use college attendance outcomes sample based on school years 2007/08 to 2017/18. Panel D to F use college graduation outcomes sample based on school years 2007/08 to 2013/14. Panel A is for students attending grades 4 to 12, panel B for students attending grades 4 to 8, panel C for students attending grades 9 to 12, panel D for students attending grades 6 to 12, panel E for students attending grades 6 to 8, and panel F for students attending grades 9 to 12. Outcomes in columns 1 to 7 are basic sociodemographic variables that are either fixed over time (columns 1 to 4) or varying at annual level (columns 5 to 7). Outcomes in columns 8 and 9 are baseline test scores measured in a year before treatment assignment for grades 4 to 9 and in grade 8 for grades 10 to 12. Outcomes is column 10 are based on predicted college attendance (panels A to C) or predicted graduation (panels E to F). We obtain these predicted values by regressing attendance or graduation using logit on variables in column 1 to 9 along with controls for imputed test scores. R^2 from these regressions are 0.14 and 0.20 for attendance and graduation, respectively. Control mean of Y is defined as mean of the outcome to the right of the cutoff in control group. Standard errors clustered at school level. ***, ** and * mark estimates statistically different from zero at the 99, 95 and 90 percent confidence level.

Table 2: Main results

	(1)	(2)	(3)	(4)	(5)	(6)
	Pooled grades		Elementar and middle school grades		High school grades	
			Panel A. Any college attendance			
PI score < 80	0.046**	0.034**	0.010	0.004	0.113***	0.088***
	(0.019)	(0.016)	(0.012)	(0.008)	(0.039)	(0.031)
Control mean of Y	0.487		0.515		0.432	
			Panel B. Attended 4-year institution			
PI score < 80	0.036**	0.027**	0.008	0.004	0.089***	0.069***
	(0.014)	(0.013)	(0.010)	(0.008)	(0.027)	(0.025)
Control mean of Y	0.324		0.355		0.265	
			Panel C. Attended 4-year selective institution			
PI score < 80	0.017*	0.013	-0.006	-0.008	0.061***	0.051***
	(0.009)	(0.009)	(0.008)	(0.007)	(0.014)	(0.014)
Control mean of Y	0.206		0.230		0.160	
Observations	1,301,077	1,301,077	864,416	864,416	436,661	436,661
			Panel D. Any college graduation			
PI score < 80	0.019**	0.018**	-0.000	0.002	0.046***	0.039***
	(0.010)	(0.008)	(0.009)	(0.007)	(0.016)	(0.014)
Control mean of Y	0.208		0.248		0.158	
			Panel E. Graduated 4-year institution			
PI score < 80	0.020***	0.018***	0.008	0.009	0.035***	0.029***
	(0.007)	(0.006)	(0.008)	(0.006)	(0.012)	(0.011)
Control mean of Y	0.145		0.178		0.103	
			Panel E. Graduated 4-year selective institution			
PI score < 80	0.010*	0.008*	0.001	0.001	0.021***	0.017**
	(0.006)	(0.005)	(0.007)	(0.005)	(0.008)	(0.007)
Control mean of Y	0.104		0.129		0.074	
Observations	344,420	344,420	197,853	197,853	146,567	146,567
Controls	No	Yes	No	Yes	No	Yes

Note: This table presents main results based on estimating Equation 1 for college attendance (panels A to C) and college graduation (panel D to F) outcomes. Panels A and D consider any college attendance/graduation, panels B and E consider attendance/graduation for 4-year institution, and panels C and F consider attendance/graduation for 4-year selective institution as defined by Barron's Selectivity Index of 1 to 3. Columns 1 and 2 pool all grades together (4 to 12 for college attendance and 6 to 12 for college graduation), columns 3 and 4 present results for grades 4 to 8 (college attendance) and grades 6 to 8 (college graduation), and columns 5 and 6 present results for grades 9 to 12. Odd numbered columns present estimates w/o any additional controls while event numbered columns present estimates with demographic controls. These include gender, racial/ethnic, poverty, ELL and disability indicators as well as baseline test scores and indicators for imputed test scores. Control mean of Y is defined as mean of the outcome to the right of the cutoff in control group. Standard errors clustered at school level. ***, ** and * mark estimates statistically different from zero at the 99, 95 and 90 percent confidence level.

Appendix: For Online Publication

Overall Effects of Targeted School Vouchers on College Enrollment and Graduation: Evidence from Ohio

Matthew Chingos, David Figlio, and Krzysztof Karbownik

A Appendix Tables

Table A1: Descriptive statistics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	College attendance sample				College graduation sample			
	Full sample	Running variable			Full sample	Running variable		
		<-12	(-12; 12)	> 12		<-12	(-12; 12)	> 12
	Panel A. Background characteristics							
Female	49.02	45.44	49.37	48.97	49.26	46.81	49.71	49.15
White	78.55	24.17	60.81	85.31	80.39	26.73	64.59	86.42
Black	12.90	65.39	28.81	6.77	12.36	63.88	26.90	6.77
Hispanic	2.80	4.96	4.12	2.33	2.25	4.76	3.28	1.87
Other race/ethnicity	5.75	5.48	6.26	5.59	5.00	4.64	5.23	4.93
In-poverty	40.13	86.09	66.19	30.73	35.70	80.00	59.38	27.30
Not ELL	98.18	95.28	97.17	98.56	98.51	95.54	97.78	98.80
Not disabled	85.79	69.83	82.09	87.27	86.15	72.46	82.34	87.62
Reading test scores	0.07	-0.98	-0.32	0.22	0.10	-0.97	-0.28	0.23
Math test scores	0.07	-0.99	-0.36	0.23	0.10	-1.00	-0.32	0.26
	Panel B. Outcomes							
Any attendance/graduation	62.63	29.35	46.94	68.38	36.34	5.21	19.12	42.43
4-year institution	48.31	14.43	30.61	54.74	29.28	2.91	13.25	34.90
4-year selective institution	34.95	6.65	18.99	40.70	22.90	1.71	9.45	27.61
Observation	5,277,556	62,498	1,301,077	3,913,981	1,432,433	18,724	344,420	1,069,289
Unique students	816,629	38,310	412,557	712,265	275,781	13,130	137,455	234,255

Note: This table presents descriptive statistics for our two analyses samples: college attendance (columns 1 to 4) and college graduation (columns 5 to 8). Columns 1 and 5 are based on all schools with with valid second-best PI score; columns 2 and 6 are for schools with running variable less than -12 (PI score < 68); columns 3 and 7 are for schools in our preferred running variable bandwidth ($68 \leq$ PI score ≤ 92); and columns 4 and 8 are for schools with running variable greater than 12 (PI score > 92). College attendance sample is based on school years 2007/08 to 2017/18 for grades 4 to 12 while college graduation sample is based on school years 2007/08 to 2013/14 for grades 6 to 12. Panel A presents means of control variables while panel B means of outcomes. All indicator variables are multiples by 100. Test scores are standardized in the full population of public school students.

Table A2: Main results: Instrumental variables

	(1)	(2)	(3)	(4)	(5)	(6)
	College attendance			College graduation		
	Any	4-year	Selective 4-year	Any	4-year	Selective 4-year
			Panel A. Pooled grades			
Voucher eligible	0.066**	0.053*	0.025	0.038*	0.038**	0.018
	(0.034)	(0.027)	(0.018)	(0.022)	(0.018)	(0.012)
First-stage	0.516***	0.516***	0.516***	0.460***	0.460***	0.460***
	(0.056)	(0.056)	(0.056)	(0.093)	(0.093)	(0.093)
Effective F-statistic	4.6	4.5	2.2	4.8	8.1	2.8
AR p-value	0.053	0.054	0.165	0.062	0.019	0.135
Control mean of Y	0.517	0.354	0.232	0.235	0.168	0.123
Observations	1,984,532	1,984,532	1,984,532	528,423	528,423	528,423
			Panel B. Elementary and middle school grades			
Voucher eligible	0.007	0.008	-0.015	0.004	0.018	0.002
	(0.015)	(0.015)	(0.012)	(0.013)	(0.011)	(0.010)
First-stage	0.540***	0.540***	0.540***	0.513***	0.513***	0.513***
	(0.049)	(0.049)	(0.049)	(0.083)	(0.083)	(0.083)
Effective F-statistic	0.2	0.3	1.5	0.1	2.7	0.1
AR p-value	0.666	0.601	0.229	0.748	0.110	0.836
Control mean of Y	0.542	0.383	0.255	0.273	0.201	0.148
Observations	1,321,168	1,321,168	1,321,168	297,103	297,103	297,103
			Panel C. High school grades			
Voucher eligible	0.189**	0.147**	0.108***	0.106	0.080	0.046
	(0.0839)	(0.0658)	(0.0406)	(0.068)	(0.051)	(0.033)
First-stage	0.469***	0.469***	0.469***	0.371**	0.371**	0.371**
	(0.108)	(0.108)	(0.108)	(0.173)	(0.173)	(0.173)
Effective F-statistic	8.2	7.8	12.4	8.4	7.6	5.5
AR p-value	0.033	0.033	0.011	0.055	0.065	0.097
Control mean of Y	0.467	0.298	0.185	0.187	0.127	0.091
Observations	663,364	663,364	663,364	231,320	231,320	231,320

Note: This table presents instrumental variables estimates where treatment is defined as voucher eligibility at school-by-year level and instrument is PI score < 80 used in a regression discontinuity framework based on Equation 1 and preferred bandwidth of +/- 12 points around the policy cutoff (PI score = 80). Columns 1 to 3 present results for college attendance outcomes while columns 4 to 6 for college graduation outcomes. Panel A is for pooled grades (4 to 12 in college attendance and 6 to 12 in college graduation samples), panel B is for elementary/middle school grades (4 to 8 in college attendance and 6 to 8 in college graduation samples), and panel C is for high school grades (9 to 12). The first row of each panel presents instrumental variables treatment effects while the second row presents the firs-stage estimates for a given sample. We then present Anderson-Rubin effective F-statistic and p-values that are robust to weak instruments (Anderson and Rubin, 1949). ***, ** and * mark estimates statistically different from zero at the 99, 95 and 90 percent confidence level.

Table A3: Robustness: Attendance

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
						Rd Robust			
	Baseline	Quadratic	Excluding imputed scores	No test scores controls	Excluding majority black schools	Conventional	Robust	Random observation per student	EdChoice sample only (pre-2015)
Panel A. Any college attendance: All grades									
PI score < 80	0.034**	0.058**	0.035**	0.040**	0.040***	0.044**	0.048**	0.048***	0.020*
	(0.016)	(0.026)	(0.016)	(0.017)	(0.014)	(0.021)	(0.022)	(0.017)	(0.011)
Observations	1,301,077	1,301,077	1,165,171	1,301,077	758,444	5,277,556	5,277,556	217,708	1,133,653
Panel B. Any college attendance: Elementary and middle school grades									
PI score < 80	0.004	0.014	0.005	0.004	0.011	0.014	0.016	0.015	
	(0.008)	(0.011)	(0.008)	(0.009)	(0.014)	(0.011)	(0.012)	(0.010)	N/A
Observations	864,416	864,416	785,236	864,416	569,443	2,641,541	2,641,541	134,868	
Panel C. Any college attendance: High school grades									
PI score < 80	0.088***	0.126***	0.090***	0.104***	0.089***	0.102***	0.109***	0.096***	0.073***
	(0.031)	(0.041)	(0.031)	(0.033)	(0.022)	(0.038)	(0.042)	(0.031)	(0.027)
Observations	436,661	436,661	379,935	436,661	189,001	2,636,015	2,636,015	82,840	269,237
Panel D. Attended 4-year institution: All grades									
PI score < 80	0.027**	0.037*	0.027**	0.033**	0.029**	0.032*	0.034*	0.032**	0.015
	(0.013)	(0.020)	(0.013)	(0.014)	(0.014)	(0.019)	(0.020)	(0.013)	(0.010)
Observations	1,301,077	1,301,077	1,165,171	1,301,077	758,444	5,277,556	5,277,556	217,708	1,133,653
Panel E. Attended 4-year institution: Elementary and middle school grades									
PI score < 80	0.004	0.013	0.005	0.005	0.001	0.013	0.014	0.009	
	(0.008)	(0.010)	(0.008)	(0.008)	(0.013)	(0.011)	(0.013)	(0.009)	N/A
Observations	864,416	864,416	785,236	864,416	569,443	2,641,541	2,641,541	134,868	
Panel F. Attended 4-year institution: High school grades									
PI score < 80	0.069***	0.073**	0.070***	0.084***	0.078***	0.068**	0.069**	0.065**	0.051**
	(0.025)	(0.034)	(0.026)	(0.026)	(0.023)	(0.032)	(0.034)	(0.025)	(0.024)
Observations	436,661	436,661	379,935	436,661	189,001	2,636,015	2,636,015	82,840	269,237
Panel G. Attended 4-year selective institution: All grades									
PI score < 80	0.013	0.019	0.013	0.018*	0.019*	0.014	0.014	0.017*	0.005
	(0.009)	(0.013)	(0.009)	(0.010)	(0.011)	(0.011)	(0.012)	(0.009)	(0.008)
Observations	1,301,077	1,301,077	1,165,171	1,301,077	758,444	5,277,556	5,277,556	217,708	1,133,653
Panel H. Attended 4-year selective institution: Elementary and middle school grades									
PI score < 80	-0.008	0.002	-0.008	-0.007	-0.002	0.004	0.006	-0.004	
	(0.007)	(0.008)	(0.007)	(0.007)	(0.010)	(0.008)	(0.010)	(0.007)	N/A
Observations	864,416	864,416	785,236	864,416	569,443	2,641,541	2,641,541	134,868	
Panel I. Attended 4-year selective institution: High school grades									
PI score < 80	0.051***	0.038**	0.052***	0.062***	0.057***	0.033**	0.030**	0.047***	0.044***
	(0.014)	(0.019)	(0.015)	(0.016)	(0.021)	(0.014)	(0.015)	(0.015)	(0.015)
Observations	436,661	436,661	379,935	436,661	189,001	2,636,015	2,636,015	82,840	269,237

Note: This table presents robustness analysis for estimates based on Equation 1 for college attendance sample. Panels A to C present results for any college attendance, panels D to F for 4-year institution college attendance, and panels G to I for selective 4-year institution college attendance as defined by Barron's Selectivity Index of 1 to 3. Column 1 replicates the results with controls from Table 2 for comparison, column 2 estimates quadratic rather than linear specification, column 3 excludes observations where we had to impute baseline test scores, column 4 uses full sample but drops test score controls, column 5 excludes schools in the top 2 deciles of share of Black students, columns 6 and 7 use nonparametric estimation based on rdrobust developed by Calonico et al. (2014) and present conventional (column 6) as well as robust (column 7) estimates, column 8 chooses randomly a single observation per student, and column 9 limits the estimation sample to school years 2007-8 to 2013-14. ***, ** and * mark estimates statistically different from zero at the 99, 95 and 90 percent confidence level.

Table A4: Robustness: Graduation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
						Rd Robust		
	Baseline	Quadratic	Excluding imputed scores	No test scores controls	Excluding majority black schools	Conventional	Robust	Random observation per student
				Panel A. Any college graduation: All grades				
PI score < 80	0.018**	0.025*	0.014*	0.019**	0.017	0.020	0.020	0.014
	(0.008)	(0.014)	(0.008)	(0.008)	(0.012)	(0.013)	(0.014)	(0.009)
Observations	344,420	344,420	310,262	344,420	210,819	1,432,433	1,432,433	73,733
				Panel B. Any college graduateion: Middle school grades				
PI score < 80	0.002	0.008	0.001	0.004	0.002	0.009	0.009	-0.000
	(0.007)	(0.009)	(0.007)	(0.008)	(0.011)	(0.010)	(0.012)	(0.010)
Observations	197,853	197,853	182,105	197,853	142,426	517,092	517,092	39,374
				Panel C. Any college graduation: High school grades				
PI score < 80	0.039***	0.039*	0.034**	0.040***	0.042*	0.024	0.021	0.030**
	(0.014)	(0.020)	(0.014)	(0.015)	(0.023)	(0.021)	(0.022)	(0.014)
Observations	146,567	146,567	128,157	146,567	68,393	915,341	915,341	34,359
				Panel D. Graduated 4-year institution: All grades				
PI score < 80	0.018***	0.019*	0.015**	0.019***	0.018**	0.017*	0.017	0.013*
	(0.006)	(0.011)	(0.006)	(0.007)	(0.009)	(0.010)	(0.011)	(0.007)
Observations	344,420	344,420	310,262	344,420	210,819	1,432,433	1,432,433	73,733
				Panel E. Graduated 4-year institution: Middle school grades				
PI score < 80	0.009	0.006	0.008	0.011	0.007	0.006	0.004	0.005
	(0.006)	(0.007)	(0.006)	(0.007)	(0.008)	(0.007)	(0.009)	(0.008)
Observations	197,853	197,853	182,105	197,853	142,426	517,092	517,092	39,374
				Panel F. Graduated 4-year institution: High school grades				
PI score < 80	0.029***	0.032**	0.027**	0.030***	0.033**	0.026	0.025	0.023**
	(0.011)	(0.016)	(0.011)	(0.012)	(0.016)	(0.017)	(0.018)	(0.011)
Observations	146,567	146,567	128,157	146,567	68,393	915,341	915,341	34,359
				Panel G. Graduated 4-year selective institution: All grades				
PI score < 80	0.008*	0.013	0.007	0.009*	0.007	0.010	0.010	0.003
	(0.005)	(0.008)	(0.005)	(0.005)	(0.007)	(0.007)	(0.008)	(0.006)
Observations	344,420	344,420	310,262	344,420	210,819	1,432,433	1,432,433	73,733
				Panel H. Graduated 4-year selective institution: Middle school grades				
PI score < 80	0.001	0.005	0.000	0.002	-0.002	0.005	0.005	-0.004
	(0.005)	(0.007)	(0.005)	(0.007)	(0.008)	(0.007)	(0.008)	(0.007)
Observations	197,853	197,853	182,105	197,853	142,426	517,092	517,092	39,374
				Panel I. Graduated 4-year selective institution: High school grades				
PI score < 80	0.017**	0.018	0.016**	0.018**	0.021*	0.009	0.007	0.009
	(0.007)	(0.011)	(0.008)	(0.008)	(0.011)	(0.010)	(0.010)	(0.008)
Observations	146,567	146,567	128,157	146,567	68,393	915,341	915,341	34,359

Note: This table presents robustness analysis for estimates based on Equation 1 for college graduation sample. Panels A to C present results for any college attendance, panels D to F for 4-year institution college attendance, and panels G to I for selective 4-year institution college attendance as defined by Barron's Selectivity Index of 1 to 3. Column 1 replicates the results with controls from Table 2 for comparison, column 2 estimates quadratic rather than linear specification, column 3 excludes observations where we had to impute baseline test scores, column 4 uses full sample but drops test score controls, column 5 excludes schools in the top 2 deciles of share of Black students, columns 6 and 7 use nonparametric estimation based on rdrobust developed by Calonico et al. (2014) and present conventional (column 6) as well as robust (column 7) estimates, and column 8 chooses randomly a single observation per student. ***, ** and * mark estimates statistically different from zero at the 99, 95 and 90 percent confidence level.

Table A5: Robustness: Alternative standard error computations

Outcomes:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Any attendance/graduation			4-year institution attendance/graduation			4-year selective institution attendance/graduation		
Grades:	All	G4/G6 to G8	G9 to G12	All	G4/G6 to G8	G9 to G12	All	G4/G6 to G8	G9 to G12
					Panel A. College attendance				
PI score < 80	0.034**	0.004	0.088***	0.027**	0.004	0.069***	0.013	-0.008	0.051***
Cluster at school level	0.031	0.666	0.004	0.034	0.601	0.006	0.140	0.224	p < 0.001
Heteroskedasticity robust	p < 0.001	0.150	p < 0.001	p < 0.001	0.062	p < 0.001	p < 0.001	p < 0.001	p < 0.001
Cluster at running variable (RV)	p < 0.001	0.447	p < 0.001	p < 0.001	0.352	p < 0.001	p < 0.001	0.033	p < 0.001
Cluster at school-by-year level	0.002	0.581	p < 0.001	0.002	0.485	p < 0.001	0.028	0.109	p < 0.001
Cluster at RV and student level	0.008	0.631	0.001	0.007	0.537	0.001	0.055	0.216	p < 0.001
Cluster at student level	p < 0.001	0.239	p < 0.001	p < 0.001	0.129	p < 0.001	p < 0.001	p < 0.001	p < 0.001
Randomization inference (Young 2019)	p < 0.001	0.970	p < 0.001	p < 0.001	0.64	p < 0.001	0.050	0.330	p < 0.001
Multiple hypotheses testing	0.010	0.654	0.010	0.010	0.535	0.010	0.040	0.089	0.010
					Panel B. College graduation				
PI score < 80	0.018**	0.002	0.039***	0.018***	0.009	0.029***	0.008*	0.001	0.017**
Cluster at school level	0.030	0.748	0.004	0.005	0.102	0.006	0.097	0.836	0.020
Heteroskedasticity robust	p < 0.001	0.564	p < 0.001	p < 0.001	0.006	p < 0.001	p < 0.001	0.701	p < 0.001
Cluster at running variable (RV)	0.004	0.669	0.002	p < 0.001	0.037	0.003	0.029	0.791	0.007
Cluster at school-by-year level	0.023	0.725	0.016	0.002	0.073	0.016	0.069	0.821	0.025
Cluster at RV and student level	0.038	0.717	0.029	0.005	0.083	0.028	0.081	0.833	0.040
Cluster at student level	p < 0.001	0.599	p < 0.001	p < 0.001	0.013	p < 0.001	p < 0.001	0.726	p < 0.001
Randomization inference (Young 2019)	p < 0.001	0.580	p < 0.001	p < 0.001	0.010	p < 0.001	0.010	0.780	p < 0.001
Multiple hypotheses testing	0.010	0.713	0.010	0.010	0.030	0.010	0.030	0.733	0.010

Note: This table presents estimates based on Equation 1 for our preferred bandwidth of +/- 12 points around the policy cutoff (PI score = 80). Panel A presents results for college attendance sample while panel B for college graduation sample. First row presents point estimates replicated from Table 2. Subsequent rows present p-values based on standard errors clustered at school level, heteroskedasticity robust standard errors, standard errors clustered at running variable (PI score − 80), standard errors clustered at school-by-year level, standard errors clustered two way at running variable and student level, and standard errors clustered at student level. The last two rows present p-values from randomization inference (Young, 2019) and p-values adjusted for multiple hypotheses testing (Clark et al., 2015). ***, ** and * mark estimates statistically different from zero at the 99, 95 and 90 percent confidence level.

Table A6: Heterogeneity: Attendance

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Gender		Race/Ethnicity		In-poverty		scores		Disabled	
	Female	Male	Black	Non-Black	Yes	No	Below median	Above median	Yes	No
Panel A. Any college attendance: All grades										
PI score < 80	0.078***	0.056**	0.034	0.055***	0.068***	0.035	0.062***	0.074***	0.048***	0.072***
	(0.026)	(0.024)	(0.022)	(0.019)	(0.025)	(0.022)	(0.024)	(0.025)	(0.018)	(0.026)
Mean of Y	0.464	0.353	0.454	0.384	0.371	0.521	0.274	0.526	0.216	0.454
Panel B. Attended 4-year institution: All grades										
PI score < 80	0.062***	0.039**	0.032*	0.038**	0.050***	0.027	0.037***	0.066***	0.025***	0.057***
	(0.021)	(0.018)	(0.018)	(0.015)	(0.019)	(0.021)	(0.014)	(0.023)	(0.009)	(0.021)
Mean of Y	0.293	0.206	0.273	0.236	0.212	0.361	0.13	0.353	0.095	0.285
Panel C. Attended 4-year selective institution: All grades										
PI score < 80	0.029**	0.018*	0.016	0.017	0.023**	0.006	0.018***	0.029*	0.007	0.028**
	(0.013)	(0.011)	(0.011)	(0.012)	(0.010)	(0.019)	(0.007)	(0.016)	(0.005)	(0.013)
Mean of Y	0.171	0.121	0.148	0.145	0.115	0.24	0.062	0.219	0.046	0.169
Observations	227,298	231,021	187,218	271,101	355,525	102,794	229,170	229,149	90,950	367,369
Panel D. Any college attendance: Elementary and middle school grades										
PI score < 80	0.031*	0.030*	0.022	0.024	0.031*	0.008	0.034**	0.028**	0.031*	0.030*
	(0.017)	(0.016)	(0.018)	(0.016)	(0.016)	(0.017)	(0.017)	(0.014)	(0.017)	(0.016)
Mean of Y	0.524	0.394	0.484	0.446	0.399	0.606	0.298	0.564	0.232	0.508
Panel E. Attended 4-year institution: Elementary and middle school grades										
PI score < 80	0.036**	0.023*	0.033**	0.016	0.029**	0.010	0.025**	0.035**	0.025***	0.031**
	(0.016)	(0.012)	(0.015)	(0.015)	(0.012)	(0.020)	(0.010)	(0.015)	(0.008)	(0.015)
Mean of Y	0.35	0.24	0.303	0.29	0.235	0.441	0.144	0.393	0.106	0.335
Panel F. Attended 4-year selective institution: Elementary and middle school grades										
PI score < 80	0.016	0.009	0.019*	0.003	0.013	-0.004	0.014**	0.011	0.008	0.014
	(0.014)	(0.009)	(0.011)	(0.013)	(0.008)	(0.020)	(0.006)	(0.014)	(0.006)	(0.012)
Mean of Y	0.217	0.148	0.174	0.186	0.133	0.302	0.072	0.255	0.055	0.21
Observations	137,845	143,417	108,012	173,250	211,315	69,947	125,349	155,913	53,076	228,186
Panel G. Any college attendance: High school grades										
PI score < 80	0.164***	0.105**	0.055	0.138***	0.132***	0.122***	0.100**	0.195***	0.077**	0.152***
	(0.049)	(0.049)	(0.043)	(0.033)	(0.050)	(0.045)	(0.043)	(0.060)	(0.032)	(0.052)
Mean of Y	0.385	0.295	0.421	0.288	0.338	0.353	0.252	0.455	0.198	0.377
Panel H. Attended 4-year institution: High school grades										
PI score < 80	0.109***	0.069*	0.033	0.095***	0.086**	0.081**	0.055**	0.147***	0.026	0.108**
	(0.038)	(0.039)	(0.036)	(0.027)	(0.039)	(0.036)	(0.028)	(0.055)	(0.018)	(0.043)
Mean of Y	0.219	0.157	0.241	0.154	0.185	0.203	0.117	0.279	0.082	0.215
Panel I. Attended 4-year selective institution: High school grades										
PI score < 80	0.052***	0.035*	0.015	0.053***	0.042**	0.043*	0.024*	0.078**	0.006	0.055**
	(0.019)	(0.021)	(0.020)	(0.018)	(0.020)	(0.025)	(0.013)	(0.031)	(0.008)	(0.023)
Mean of Y	0.111	0.082	0.12	0.082	0.091	0.119	0.053	0.153	0.036	0.112
Observations	89,453	87,604	79,206	97,851	144,210	32,847	103,821	73,236	37,874	139,183

Note: This table presents heterogeneity analyses for college attendance sample based on Equation 1 and our preferred bandwidth of +/- 12 points around the policy cutoff (PI score = 80). Panels A to C present results for all grades (4 to 12), panels D to F for elementary and middle school grades (4 to 8), and panel G to I for high school grades (9 to 12). Poverty and disability are measured contemporaneously every school year. Test scores are prior year scores in grades 4 to 9 and grade 8 scores for grades 10 to 12 divided into above vs. below median in the estimation sample. ***, ** and * mark estimates statistically different from zero at the 99, 95 and 90 percent confidence level.

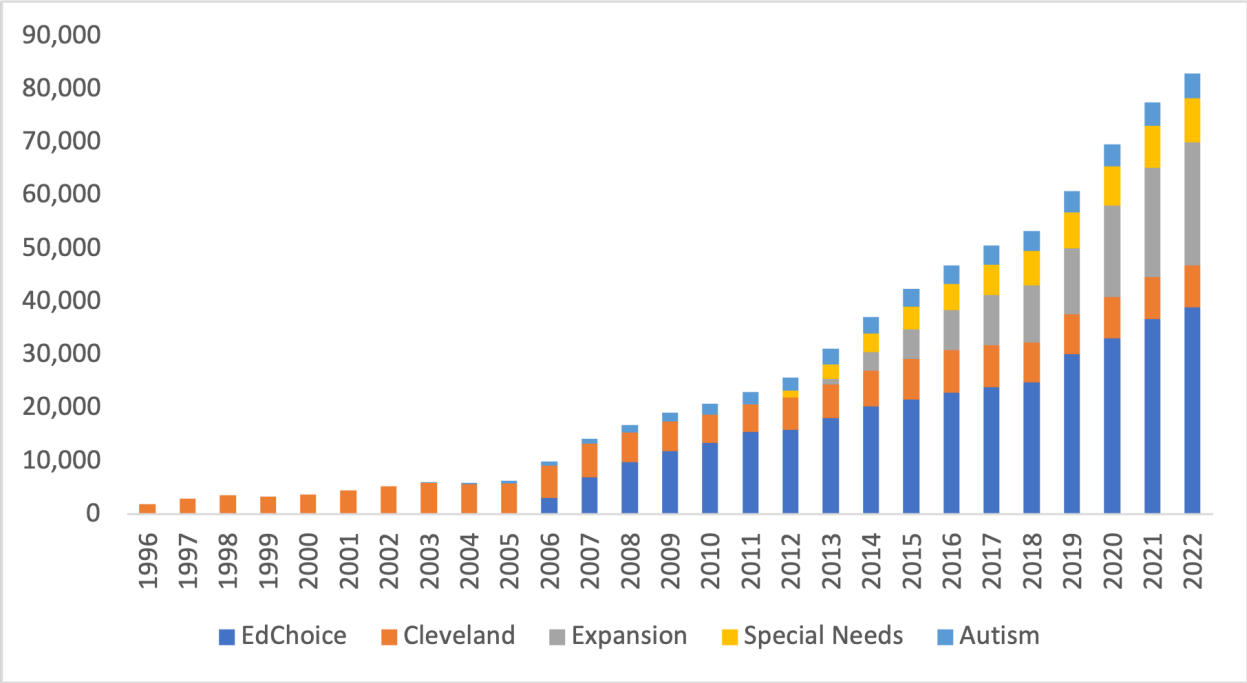
Table A7: Heterogeneity: Graduation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Gender		Race/Ethnicity		In-poverty		scores		Disabled	
	Female	Male	Black	Non-Black	Yes	No	Below median	Above median	Yes	No
Panel A. Any college graduation: All grades										
PI score < 80	0.024	0.015	0.027**	0.014	0.024**	-0.004	0.012*	0.028	-0.002	0.027**
	(0.015)	(0.011)	(0.013)	(0.014)	(0.010)	(0.019)	(0.007)	(0.018)	(0.006)	(0.014)
Mean of Y	0.172	0.111	0.111	0.158	0.095	0.245	0.052	0.22	0.039	0.165
Panel B. Graduated 4-year institution: All grades										
PI score < 80	0.022*	0.016*	0.025**	0.012	0.020**	0.007	0.012***	0.027*	0.000	0.025**
	(0.013)	(0.008)	(0.011)	(0.011)	(0.008)	(0.016)	(0.005)	(0.015)	(0.005)	(0.011)
Mean of Y	0.114	0.073	0.076	0.102	0.056	0.176	0.024	0.154	0.018	0.11
Panel C. Graduated 4-year selective institution: All grades										
PI score < 80	0.012	0.011*	0.014*	0.008	0.011*	0.004	0.006**	0.017	0.001	0.015*
	(0.010)	(0.006)	(0.008)	(0.009)	(0.006)	(0.014)	(0.003)	(0.012)	(0.003)	(0.009)
Mean of Y	0.081	0.053	0.053	0.074	0.039	0.13	0.014	0.113	0.012	0.079
Observations	58,589	58,521	45,674	71,436	83,904	33,206	58,961	58,149	22,839	94,271
Panel D. Any college graduation: Middle school grades										
PI score < 80	0.019	0.020	0.035*	0.012	0.020*	-0.001	0.015	0.018	-0.002	0.027*
	(0.018)	(0.013)	(0.019)	(0.015)	(0.011)	(0.020)	(0.010)	(0.016)	(0.009)	(0.015)
Mean of Y	0.23	0.142	0.142	0.201	0.112	0.3	0.062	0.264	0.051	0.213
Panel E. Graduated 4-year institution: Middle school grades										
PI score < 80	0.014	0.017*	0.032**	0.008	0.013	0.005	0.013**	0.015	0.002	0.021*
	(0.015)	(0.010)	(0.015)	(0.013)	(0.009)	(0.017)	(0.006)	(0.014)	(0.006)	(0.012)
Mean of Y	0.159	0.096	0.103	0.136	0.069	0.22	0.03	0.19	0.025	0.148
Panel F. Graduated 4-year selective institution: Middle school grades										
PI score < 80	0.009	0.016**	0.020*	0.010	0.009	0.008	0.008**	0.014	0.003	0.016
	(0.014)	(0.008)	(0.012)	(0.012)	(0.007)	(0.018)	(0.004)	(0.014)	(0.005)	(0.011)
Mean of Y	0.116	0.073	0.074	0.101	0.049	0.165	0.019	0.142	0.017	0.11
Observations	30,631	32,325	21,367	41,589	41,478	21,478	27,523	35,433	11,428	51,528
Panel G. Any college graduation: High school grades										
PI score < 80	0.032	0.012	0.026	0.015	0.028*	-0.004	0.011	0.037	-0.001	0.030
	(0.023)	(0.016)	(0.017)	(0.026)	(0.016)	(0.036)	(0.010)	(0.038)	(0.010)	(0.022)
Mean of Y	0.112	0.075	0.089	0.099	0.081	0.141	0.044	0.155	0.027	0.111
Panel H. Graduated 4-year institution: High school grades										
PI score < 80	0.032*	0.015	0.023	0.017	0.026**	0.008	0.012*	0.039	0.000	0.031*
	(0.018)	(0.012)	(0.014)	(0.018)	(0.013)	(0.028)	(0.007)	(0.030)	(0.007)	(0.017)
Mean of Y	0.066	0.045	0.057	0.055	0.045	0.094	0.019	0.1	0.011	0.067
Panel I. Graduated 4-year selective institution: High school grades										
PI score < 80	0.018	0.004	0.012	0.005	0.014	-0.002	0.006	0.017	0.001	0.015
	(0.013)	(0.009)	(0.009)	(0.013)	(0.009)	(0.018)	(0.004)	(0.021)	(0.004)	(0.012)
Mean of Y	0.044	0.03	0.038	0.037	0.03	0.064	0.011	0.069	0.007	0.045
Observations	27,958	26,196	24,307	29,847	42,426	11,728	31,438	22,716	11,411	42,743

Note: This table presents heterogeneity analyses for college graduation sample based on Equation 1 and our preferred bandwidth of +/- 12 points. Panels A to C present results for all grades (6 to 12), panels D to F for middle school grades (6 to 8), and panel G to I for high school grades (9 to 12). Poverty and disability are measured contemporaneously every school year. Test scores are prior year scores in grades 6 to 9 and grade 8 scores for grades 10 to 12 divided into above vs. below median in the estimation sample. ***, ** and * mark estimates statistically different from zero at the 99, 95 and 90 percent confidence level.

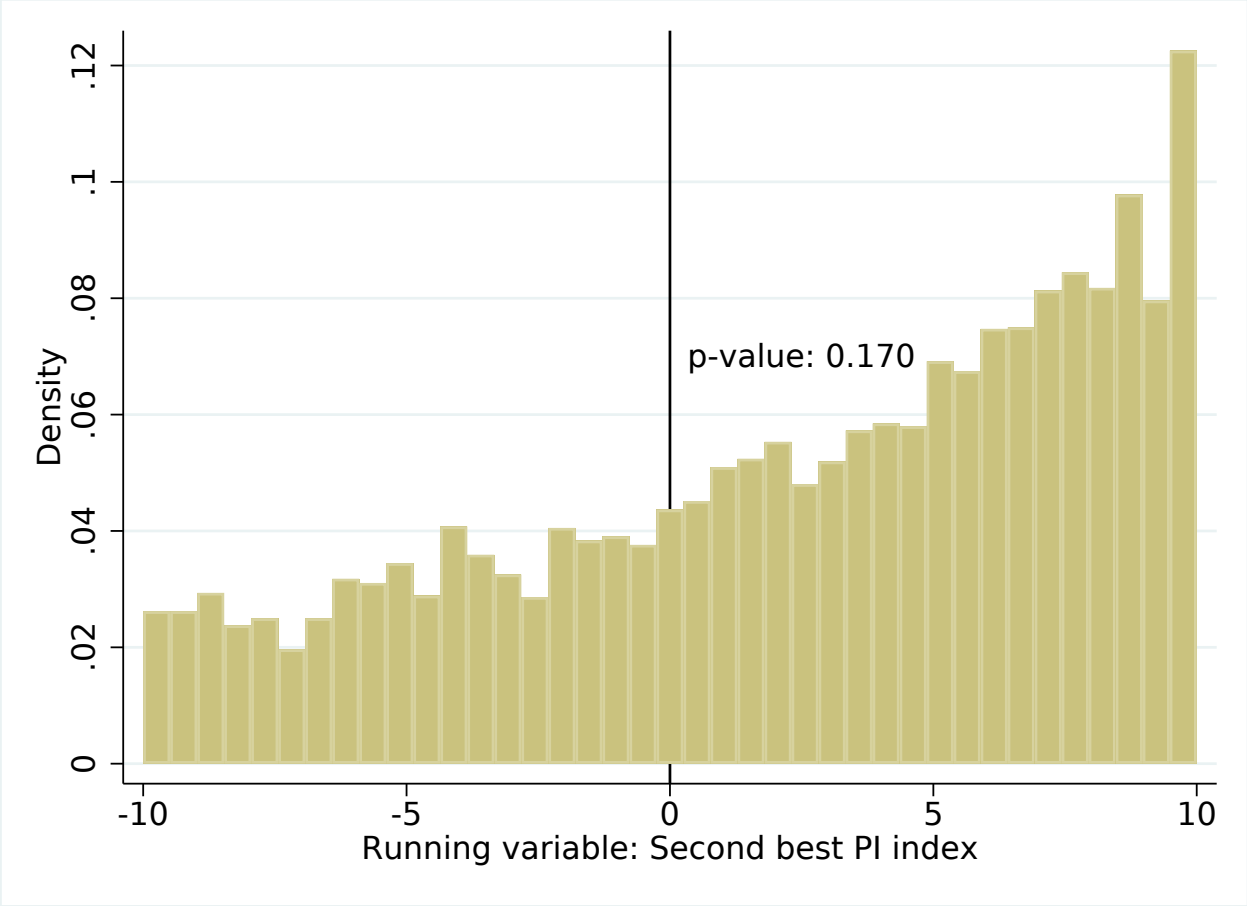
B Appendix Figures

Figure A1: Density test



Notes: This figure presents evolution of various private school choice programs in Ohio. Our focus is on the EdChoice (blue bars) and EdChoice Expansion (gray bars).

Figure A2: Density test



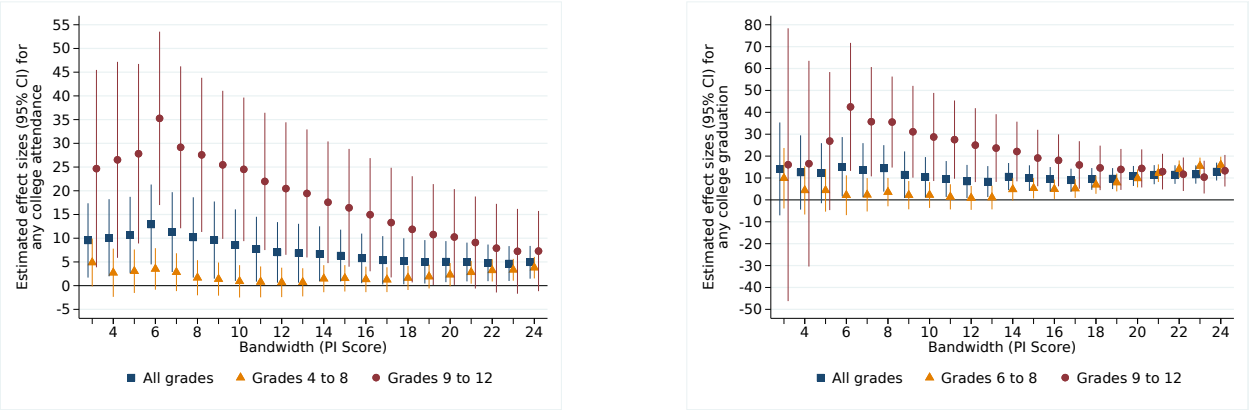
Notes: This figure presents density of schools around the policy cutoff of PI score = 80. The p-value is based on density test proposed by Calonico et al. (2014).

Figure A3: Robustness: Bandwidth selection

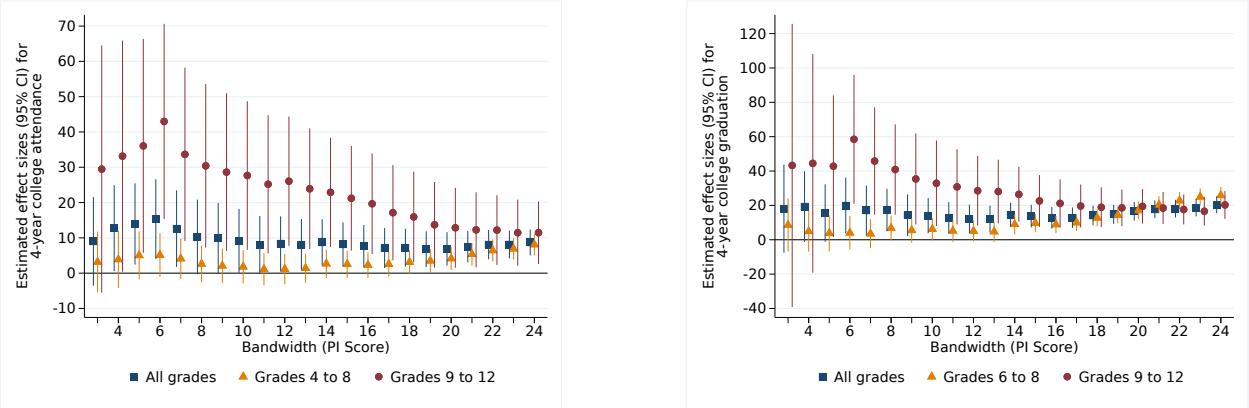
I. College attendance

II. College graduation

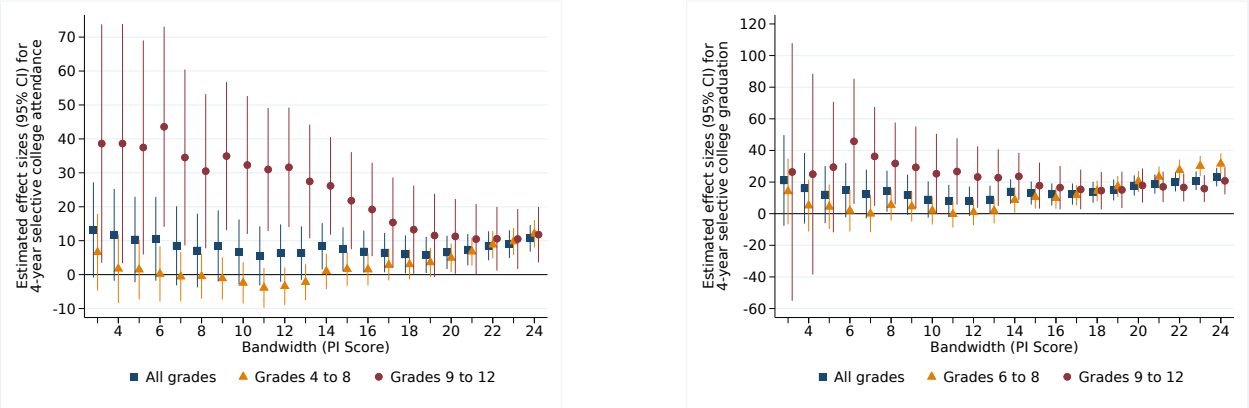
Panel A. Any



Panel B. 4-year institution

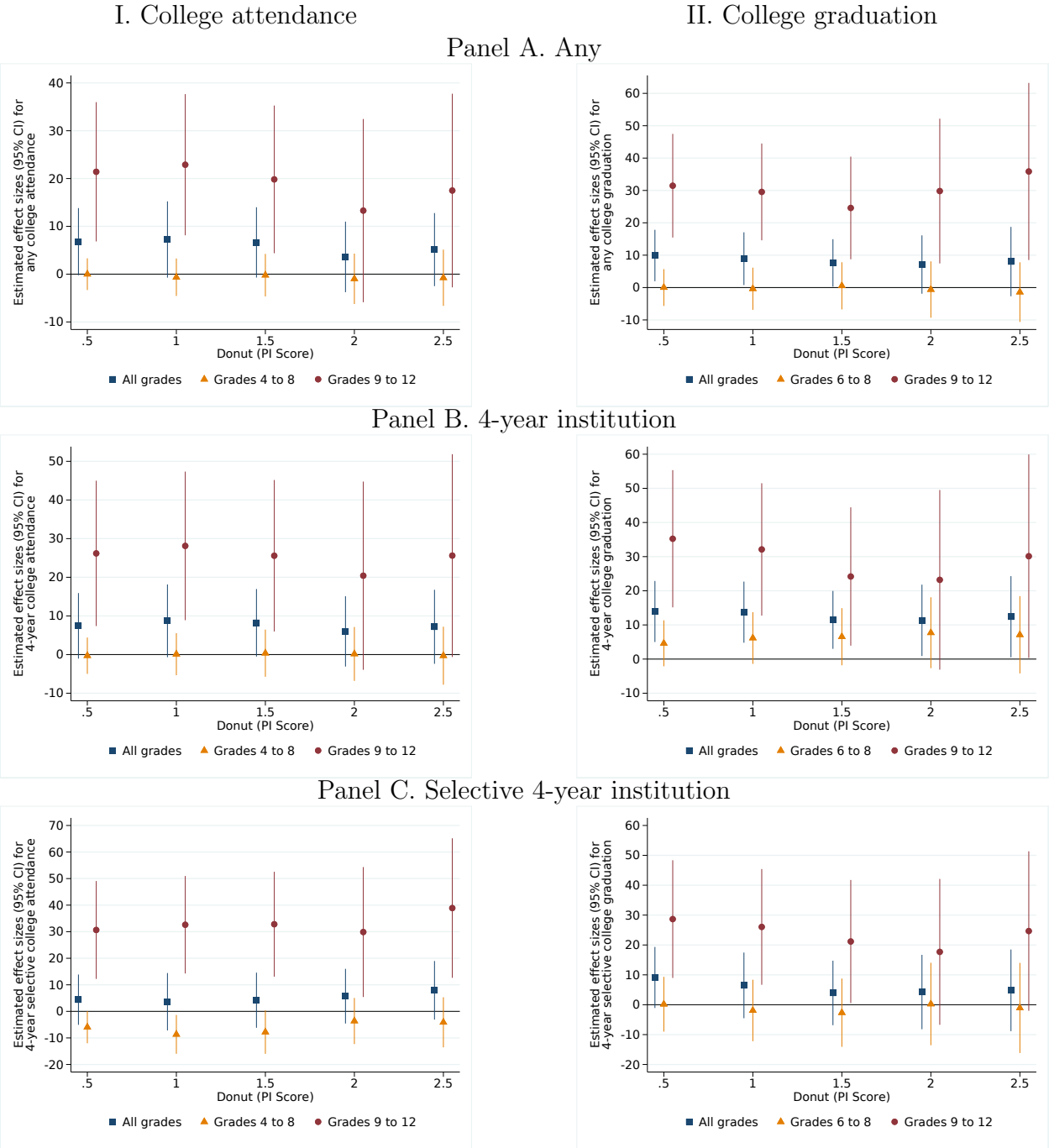


Panel C. Selective 4-year institution



Notes: Point estimates and 95% confidence intervals. This figure presents robustness to alternative bandwidths used to estimate Equation 1. We vary the bandwidth from +/- 3 points ($77 \leq$ PI score ≤ 83 to +/ 24 points ($56 \leq$ PI score ≤ 104). Column I presents results for college attendance sample while column II for college graduation sample. Panel A presents results for any college attendance/graduation, panel B for 4-year institution attendance/graduation, and panel C for 4-year selective institution attendance/graduation as defined by Barron's Selectivity Index of 1 to 3. Navy estimates are for the pooled sample (grades 4 to 12 for college attendance and 6 to 12 for college graduation), orange estimates for elementary/middle school sample (grade 4 to 8 for college attendance and 6 to 8 for college graduation), and maroon estimates are for high school sample (grades 9 to 12). Standard errors clustered at school level.

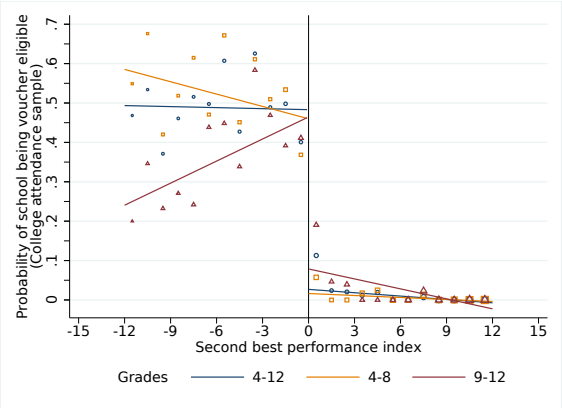
Figure A4: Robustness: Donut regressions



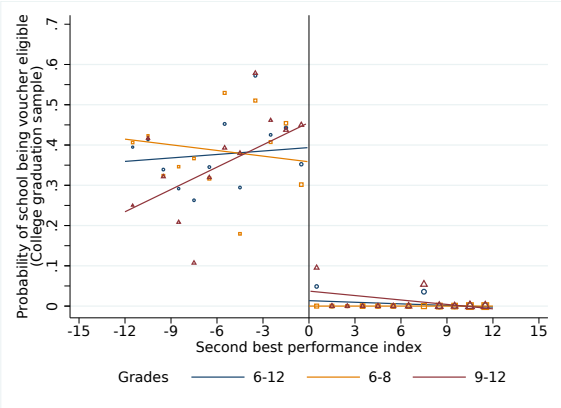
Notes: Point estimates and 95% confidence intervals. This figure presents robustness to donut regression samples used to estimate Equation 1 based on our preferred bandwidth of +/- 12 points ($68 \leq$ PI score ≤ 92). We exclude observations close to the cutoff from +/- 0.5 points to +/ 2.5 points. Column I presents results for college attendance sample while column II for college graduation sample. Panel A presents results for any college attendance/graduation, panel B for 4-year institution attendance/graduation, and panel C for 4-year selective institution attendance/graduation as defined by Barron's Selectivity Index of 1 to 3. Navy estimates are for the pooled sample (grades 4 to 12 for college attendance and 6 to 12 for college graduation), orange estimates for elementary/middle school sample (grade 4 to 8 for college attendance and 6 to 8 for college graduation), and maroon estimates are for high school sample (grades 9 to 12). Standard errors clustered at school level.

Figure A5: First-stage: Graphical evidence

I. College attendance



II. College graduation



Notes: This figure presents graphical evidence for first stage. It computes the probability of being eligible for school voucher based on lists posted by the Department of Education as a function of the second-best PI score (i.e., our running variable). We limit the sample to +/- 12 points (our preferred bandwidth) around the policy cutoff of PI score = 80. For display purposes PI score is rounded to integer values. Column I presents it for college attendance sample while column II for college graduation sample. Navy lines are for the pooled sample (grades 4 to 12 for college attendance and 6 to 12 for college graduation), orange lines for elementary/middle school sample (grade 4 to 8 for college attendance and 6 to 8 for college graduation), and maroon lines are for high school sample (grades 9 to 12).