ANNENBERG
BROWN UNIVERSITY

EdWorkingPaper No. 19-87

The Effects of Student Growth Data on School District Choice: Evidence from a Survey Experiment

David M. Houston
Harvard University

Jeffrey R. Henig
Columbia University

We conduct an online survey experiment in which participants are asked to imagine that they are parents moving to a new metropolitan area. They then choose between the five largest school districts in that area. All participants receive demographic data for each district. In addition, some participants are randomly assigned to receive average achievement and/or average growth data for each district. While there are strong relationships between student demographics and student achievement, the links between student demographics and student growth are much weaker. We find that, on average, the provision of growth data causes participants to choose less white and less affluent districts. Moreover, the provision of both achievement and growth data causes participants to choose less white and less affluent districts than the provision of achievement data alone.

VERSION: October 2019

Suggested citation: Houston, David M., and Jeffrey R. Henig. (2019). The Effects of Student Growth Data on School District Choice: Evidence from a Survey Experiment. (EdWorkingPaper: 19-87). Retrieved from Annenberg Institute at Brown University: http://www.edworkingpapers.com/ai19-87

WORKING PAPER

The Effects of Student Growth Data on School District Choice: Evidence from a Survey Experiment

David M. Houston
Harvard University, Kennedy School of Government

Jeffrey R. Henig
Teachers College, Columbia University

October 30, 2019

Abstract:

We conduct an online survey experiment in which participants are asked to imagine that they are parents moving to a new metropolitan area. They then choose between the five largest school districts in that area. All participants receive demographic data for each district. In addition, some participants are randomly assigned to receive average achievement and/or average growth data for each district. While there are strong relationships between student demographics and student achievement, the links between student demographics and student growth are much weaker. We find that, on average, the provision of growth data causes participants to choose less white and less affluent districts. Moreover, the provision of both achievement and growth data causes participants to choose less white and less affluent districts than the provision of achievement data alone.

Funding:

This work is supported by the Russell Sage Foundation and the William T. Grant Foundation. Any opinions, findings, conclusions, or recommendations expressed here are those of the authors and do not necessarily reflect the views of the Russell Sage Foundation or the William T. Grant Foundation.

Pre-Registration and Replication Files:

This experiment has been pre-registered on the American Economic Association's registry for randomized controlled trials (https://www.socialscienceregistry.org/trials/3401). The pre-analysis plan, survey instrument, data, and statistical code are available on the Open Science Framework (https://osf.io/3e8wv/).

The level and content of available information has been a subject of ongoing interest for those who study the workings of markets, politics, and public policy. We often assume that information is the grease that makes society's gears turn more fluidly. Informed consumers can make better choices among goods and services, driving markets toward more efficient and satisfying solutions. Informed citizens can make better choices among candidates, driving democracies toward more responsive and effective decisions. Informed beneficiaries of public policies can make those policies work more smoothly and effectively.

While social scientists have often found it useful to model markets, politics, and policies on the assumption of accurate, complete, and widely distributed information, contemporary research increasingly focuses on the ways in which reality falls short of this ideal. Even relatively minor differences in the content, amount, source, or presentation of information can generate substantial differences in behavior—sometimes desirable and sometimes not. The dissemination of information can have complex and unexpected consequences when it encounters individuals' prior beliefs, political loyalties, learned biases, and cognitive shortcuts (Bolsen, Druckman, & Cook, 2014; Clinton & Grissom, 2015; Druckman, Peterson, & Slothuus, 2013; Tversky & Kahneman, 1974). Naïve hopes for a straightforward link between better information and socially desirable outcomes need to be tempered, but the lesson that information is not all that matters does not mean that information does not matter at all. We need a better understanding of how the distribution of information can supplement or undermine social goals.

We consider an important arena for these issues: information about academic performance and its implications for families' school and housing decisions. This paper presents the results of an online survey experiment to assess whether the provision of different types of academic performance information affects participants' choices between a set of school districts.

As part of the survey, participants engage in a simulation in which they are asked to imagine themselves as parents moving to a new metropolitan area. Participants then indicate their preferred choice between the five largest school districts in that area. To guide this decision, all participants receive information on the demographic characteristics of the districts. In addition, some participants are randomly assigned to receive some form of academic performance information: either average student achievement, average student growth, both, or neither. This process is repeated for the metro areas of the nation's five largest cities (according to the US Census Bureau's 2017 estimates): New York, Los Angeles, Chicago, Houston, and Phoenix.

Average student growth (i.e., the rate of improvement in students' academic performance over time) offers two chief advantages over more traditional measures of average student achievement (i.e., student academic performance at a single point in time). First, student growth arguably provides a more accurate and useful—if still imperfect—indicator of schools' and districts' contributions to student learning. Second, compared to student achievement, student growth is less tied to the racial and socio-economic composition of the student body. Average student growth data allow educators and the public to revisit the conventional thinking that the "best" school districts and the most affluent school districts are one and the same. They make it possible to recognize high performing districts that serve a broad range of students, from the severely disadvantaged to the exceptionally privileged. Perhaps most importantly, a growth-based understanding of educational performance could potentially alter how families choose schools and districts for their children in ways that reduce the racial and socio-economic segregation that shapes American education systems.

Taking advantage of the comparatively weak relationship between average student growth and districts' racial and socio-economic compositions, we explore whether individuals who are given average growth data are more likely to choose less white and less affluent school districts than their peers who receive either no academic performance data or only average achievement data. On one hand, existing research indicates that various perceptual filters can make people impervious to new information. The effects of providing academic performance information—of any kind—may pale in comparison to the effects of providing information about the demographic composition of the student body. It could also be the case that people are simply more interested in student achievement than student growth when making choices between districts, rendering the provision of the latter ineffectual. On the other hand, there is also a growing literature on the ability to influence individuals' choices by altering the content or presentation of relevant information. In most of the metropolitan areas that we consider, there are relatively high growth districts that serve a disproportionate number of disadvantaged students. By emphasizing student growth rather than (or in addition to) student achievement, it may be possible to encourage some people to consider districts that they would otherwise rule out.

Our results suggest that providing student growth data can indeed cause individuals to choose less white and less affluent districts. We observe modest effects compared to individuals who receive no academic performance data and moderately large effects compared to individuals who receive only student achievement data. Moreover, the provision of both achievement and growth data causes individuals to choose less white and less affluent districts than the provision of achievement data alone. These patterns are conditional on the presence of one or more relatively high growth districts that serve a disproportionate number of disadvantaged students. Such districts exist in some metro areas but not others.

Following the passage of the Every Student Succeeds Act of 2015, 48 states and the District of Columbia have added measures of student growth to school and district report cards

(Data Quality Campaign, 2019). The results of our experiment can help researchers, policymakers, educators, and the public think through some of the possible consequences of this change in the way we measure and report academic performance. To the extent that school and district report cards—as well as secondary sources that draw on these data—influence families' educational and residential choices, our findings suggest that we may begin to see stronger preferences for schools and housing in communities where histories of relatively low student achievement have obscured relatively high student growth.

Literature Review

Choosing Homes and Choosing Schools

Early theorists of residential and school choice frequently borrowed standard microeconomic assumptions about perfect information and mobility. Tiebout (1956) famously applied this to household choices among local jurisdictions, suggesting that rational selection of competing packages of taxes and services would enable households to "vote with their feet" and maximize alignment between household preferences and public goods. Tiebout noted that the assumption of perfect information was problematic and called out for empirical probing, but he took comfort from a handful of studies suggesting "surprising awareness" about local revenues and expenditures (p. 423). He also noted that policies to promote citizens' knowledge could improve the applicability of the model.

An extensive literature documents the relationship between the housing market and measures of academic performance from nearby schools: Homeowners pay a premium for houses in higher performing districts and near higher performing schools (Bayer, Ferreira, & McMillan, 2007; Black, 1999; Kane, Riegg, & Staiger, 2006). Moreover, when new information on academic performance is released, there often appears to be a subsequent effect on housing prices. In Florida, Figlio and Lucas (2004) find large but short-lived effects of the release of school report cards on property values. Fiva and Kirkebøen (2011) observe a similar pattern in Norway. On the other hand, following the release of teacher and school value-added data in Los Angeles, housing prices did not appear to respond systematically in the subsequent year (Imberman & Lovenheim, 2016). The authors speculate that the absence of an effect may have been due to the contentious debate following the release of the value-added data in the *Los Angeles Times*, undermining the information's credibility.

The early literature on school choice, stemming from Friedman's (1962) call for school vouchers, similarly began with economic models assuming widely distributed information and projections that increased choice would also promote integration. The norm in most school districts was and, albeit to a lesser extent, still is to assign children to schools based on residential attendance zones. Proponents of vouchers and, beginning in the 1990s, charter schools, suggested that segregation in the public schools reflected patterns of racial, ethnic, and economic segregation in housing. Promoting school choice, they argued, would free households to shop for schools based on student needs and interests, resulting in more integrated school enrollments. This hypothesis rested either on a "racial neutrality" presumption that preferences regarding the demographic composition of schools are of minor importance or a "managed choice" presumption that public officials could strategically design incentives that would induce households to make choices supportive of integration (Henig, 1990). Critics questioned the compatibility of choice and integration, arguing that informational disparities would give more affluent families structural advantages over lower-income and minority families who, due to less education, less time, less access to accurate and reliable information, and fewer resources to devote to gathering information, would be less likely to exercise choice or to do so effectively

(Schneider, Teske, & Marschall, 2000).

As it became apparent that many families lacked information about their educational options, some proponents of choice responded by increasing attention to structured mechanisms for providing it. Policymakers and researchers placed a new emphasis on efforts to maximize information in both the housing and school choice arenas. Studies of housing mobility programs like Section 8 of the Housing Act of 1937, Moving to Opportunity for Fair Housing, and Chicago's Gautreaux Assisted Housing Program concluded that counseling and informational mechanisms were necessary to combat the tendency for households to relocate in neighborhoods that replicated the segregation that such programs were intended to combat (de Souza Briggs, Popkin, & Goering, 2010; Goering & Feins, 2003; Hartung & Henig, 1997). Responding to a perceived need for better information, districts offering public school choice disseminated brochures, organized parent fairs, and created online databases with information on school enrollment, programs, and performance (Dougherty et al., 2013; Schneider & Buckley, 2002).

Disappointing experiences with various information interventions led to disillusionment with the belief that more and better information would inevitably promote integration. Despite advocates' hopes and parents' claims that choice would be driven by considerations of the quality and nature of academic offerings, multiple studies of actual search and choice behavior revealed operant preferences relating to the racial and socio-economic composition of schools, reliance on heuristic shortcuts that correlate with student demographics, or measures of academic performance that better reflect peer characteristics than school effectiveness (Abdulkadiroglu et al., 2019; Dougherty et al., 2013; Rothstein, 2006; Schneider & Buckley, 2002). In districts with centralized school application and enrollment systems where families rank their preferred schools (such as Charlotte-Mecklenburg, New Orleans, and Washington, DC), analyses of application data revealed that families placed a high value on school proximity, a demographic composition that resembled their own backgrounds, and, to a lesser extent, academic performance (Hastings, Kane, & Staiger, 2006; Harris & Larsen, 2015; Glazerman & Dotter, 2017). Moreover, preferences for academic performance (insofar as these preferences were reflected in the ranked choices submitted to the district) were heterogeneous: demand for higher academic performance increased along with family income and prior student achievement (Hastings, Kane, & Staiger, 2006).

However, the empirical literature also points to the potential for steering individuals' attitudes and behaviors in meaningful ways. Multiple survey experiments have suggested that the dissemination of student achievement data and other indicators of educational performance can influence attitudes towards educational institutions (Barrows, Henderson, Peterson, & West, 2016; Clinton & Grissom, 2015; Jacobsen, Snyder, and Saultz, 2014; Schneider, Jacobsen, White, & Gehlbach, 2018). Moreover, variations in the amount, format, and sequence of available information can shape school preferences when choosing between options in a hypothetical school district (Glazerman, Nichols-Barrer, Valant, & Burnett, 2018). These findings are not limited to survey experiments: the distribution of pamphlets containing academic performance information for nearby schools (average test scores, graduation rates, etc.) to low-income students and families can induce some recipients to enroll in higher performing educational options (Corcoran, Jennings, Cohodes, & Sattin-Bajaj, 2018; Hastings & Weinstein, 2008). It appears that some individuals are sensitive to relatively small interventions that shape the accessibility and presentation of data.

The distribution of academic performance information can also have unexpected and undesirable consequences. Recent work suggests that the proliferation of school information websites can shape residential decisions in ways that exacerbate rather than ameliorate inequality. For example, the rollout of GreatSchools.org—a website that seeks to provide users with information about schools and school systems—across districts throughout the US has coincided with a divergence in housing prices, income levels, educational attainment, and racial composition between zip codes in a given community (Hasan & Kumar, 2018). This pattern is consistent with the concern that better-off families would use school information more effectively to secure additional educational advantages for their children. The timeline of this study, 2006-2015, covered the period in which GreatSchools.org's academic performance indicators focused primarily on student achievement (in subsequent years, the website has included information on student growth when and where data are available).

There is a renewed demand for research on how to make school information available to families. The federal Every Student Succeeds Act of 2015 (ESSA) requires states and districts to provide public report cards that include information on state, district, and school performance in an "easily accessible and user-friendly" manner (Burnette, 2017). However, there is relatively little federal guidance on the content and design of these report cards, leaving states and districts with considerable discretion over what to include, what to emphasize, and how to present it. School districts such as Camden, Denver, Indianapolis, Newark, New Orleans, and Washington, DC, have experimented with unified enrollment systems in which families can apply to any public school—including alternative, magnet, and charter schools—via common websites that provide data about each educational option (Hesla, 2018). The presentation of academic performance data on these enrollment websites could have considerable influence on families' choices. Moreover, non-governmental organizations, such as GreatSchools.org, continue to play a large role in the dissemination of school information to the public. We seek to inform these efforts by exploring some of the potential consequences of how districts, states, and other entities measure and report educational quality.

Student Achievement Versus Student Growth

Educational quality is a multidimensional concept that includes, but is not limited to, adequate resources, effective practices, and satisfactory student academic performance (Ladd & Loeb, 2013). In our experiment, we focus on those aspects of educational quality that are captured by student academic performance as measured by state assessments. The results of these tests offer important—albeit imperfect and incomplete—indicators of school and school district effectiveness. The choices about how to measure and report academic performance have meaningful consequences. Teachers, administrators, and policymakers make extensive use of student test scores to guide decision-making at all levels of the education system. Many families also rely on student test scores when making school enrollment and housing decisions.

We consider two forms of student academic performance data: 1) the average achievement on state assessments of students in grades 3-8 and 2) the average growth of students as they progress through grades 3-8. The research literature uses a range of terms to distinguish between these two concepts (i.e., referring to achievement level or achievement status in contrast to growth). For simplicity, we prefer the basic dichotomy: achievement and growth. Achievement attempts to measure the status of students' skills and knowledge in the tested domains at a given point in time. Examples of achievement include the average score of a group of students or the percentage of students who meet an established proficiency target. Growth, on the other hand, attempts to measure the rate at which the same students' achievement improves over time (Auty et al., 2008; Goldschmidt et al., 2005). There are multiple ways to measure growth. For example, gain-based growth models simply calculate the difference between a student’s test scores from one year to the next. Conditional and multivariate growth models measure the extent to which a student performs relative to his or her peers with similar prior achievement and, in some cases, the same demographic characteristics (Castellano & Ho, 2013; Goldschmidt, Choi, & Beaudoin, 2012).

Average growth is not a perfect measure of educational quality. As it is typically constructed, growth only measures the rate of improvement in reading and math. Although these subjects are important elements of schooling, educational institutions also have other academic and non-academic objectives. Moreover, growth models do not necessarily isolate the causal effects of schools or districts on students’ reading and math performance. There are also other non-school factors that influence the rate at which students learn (Reardon, 2018). Some measures of growth—including the one we employ in this study—can also be partially confounded by the changing composition of students in a community (Reardon & Hinze-Pifer, 2017). Lastly, the average academic performance of a child’s peers is also a potentially important element of educational quality (and one that may be relevant to families making school and district enrollment decisions).

However, compared to average achievement, average growth offers more insight into schools’ and districts’ contributions to student learning (Reardon, 2018). Average achievement bears a strong relationship to the demographic composition of the student body (Hegedus, 2018; Reardon, 2016; Sirin, 2005; van Ewijk & Sleegers, 2010). Schools and districts that serve a larger proportion of disadvantaged students tend to have lower average achievement for reasons unrelated to the effectiveness of the educational institutions. Such students face more out-of-school obstacles than their more advantaged peers. Some (if not most) of the differences in average achievement between schools and between districts are due to differences in the backgrounds of the students they serve—not differences in the quality of schooling. While there is also a link between student demographics and average growth, this relationship is relatively weak and there are far more exceptions to the rule. Measuring growth makes it possible to identify schools and districts where students progress faster than average, regardless of their starting point. When viewed through this lens, there are many examples of schools and districts that serve low income and minority students remarkably well. Our experiment seeks to shed light on the effects of distributing this information.

Methods

Research Questions

1. Compared to the provision of no academic performance data, does the provision of average achievement data and/or average growth data lead participants to choose school districts with different racial and socio-economic compositions?
2. Compared to the provision of average achievement data, does the provision of average growth data—on its own or in addition to achievement data—lead participants to choose school districts with different racial and socio-economic compositions?
3. Do these effects vary by:
    a. The presence of children age 0-17 in the household?
    b. The presence of public school students age 0-17 in the household?
    c. Race (white or persons of color)?
    d. Family income (under or over $75,000)?

In addition to these research questions, we also explore whether the provision of average achievement and/or average growth data leads participants to choose higher performing districts on those measures. This analysis does not appear in our pre-analysis plan, but it provides important context for understanding the primary findings.

Achievement, Growth, and Demographic Data

For measures of district-level average achievement, average growth, median household income, free and reduced-price lunch eligibility, and racial composition, we use the Stanford Education Data Archive v2.1 (SEDA). SEDA contains data from state standardized tests in reading and math in grades 3-8 administered from 2009-2015 for almost every school district in the US. SEDA defines school districts in geographic terms: the dataset contains academic performance data for all public schools located in the geographic boundaries of the district, including charter schools. For each district, SEDA contains average achievement and growth in reading and math as well as the average across both subjects (we employ these combined values in our experiment). The student test score data have been converted to a common scale that allows district-to-district comparisons across the country (Fahle et al., 2018; Reardon, Kalogrides, & Ho, 2018). We use the empirical Bayes grade cohort scale (GCS) estimates for the measures of achievement and growth. Achievement is measured such that a score of six represents a school district where the average student scores at about the same level as the average sixth grader in the national reference cohort (students who entered fourth grade in 2009 and eighth grade in 2013). Growth is measured such that a score of 1.2 represents a school district in which the average student's test scores improve about 1.2 grade level equivalents in one year. These metrics are likely unfamiliar to the participants in the survey experiment, and they may be difficult to understand. To aid in the interpretability of these values for participants, we also report achievement and growth scores in terms of national percentiles. The use of percentiles facilitates the comparison of districts' relative performance on a given metric.

SEDA provides us with the unique opportunity to conduct our experiment using real academic performance and demographic data that map directly to real districts that families could conceivably choose. Using SEDA data—rather than, for example, asking participants to choose between hypothetical districts—allows us to provide participants with options that accurately reflect the joint distribution of academic and demographic characteristics. Furthermore, using real data gives our experiment an essential element of verisimilitude. We expect that some participants will enter the experiment with prior knowledge of the available districts, making the decision-making process more realistic.

To assess the representativeness of our subject pool, we collect nationwide measures of age, gender, race/ethnicity, income, political party identification, political ideology, the presence of children in the household, and the presence of public school students in the household from the 2016 American National Election Study.

Experimental Design

We used Amazon's Mechanical Turk (MTurk) to recruit adults (age 18+) living in the US for the survey experiment. MTurk is an online marketplace in which users complete short tasks (such as online surveys) for a small fee. Although individuals on MTurk are not representative of the US population, social and behavioral scientists often use the service to recruit subject pools that are more diverse and more representative than samples recruited from their immediate environments (Berinsky, Huber, & Lenz, 2012; Follmer, Sperlin, & Suen, 2017). In the context of survey experiments exploring information, priming, and framing effects, researchers tend to find similar effect sizes in both MTurk samples and nationally representative samples (Coppock, 2018; Mullinix, Leeper, Druckman, & Freese, 2015).

We conducted our experiment three times: a 50 participant pilot study administered in October 2018, a 2,500 participant study administered later in the same month, and a 2,500 participant replication study administered in February 2019. The research design, data collection, and analysis were consistent across all three administrations.

Our experiment consists of a simulation embedded in an online survey. Participants are first asked to imagine that they are parents moving to the New York City metropolitan area. They are instructed that, when deciding where to live, one of their top priorities is to choose a school district for their elementary school-age child. The survey then provides basic demographic information for the five largest school districts in the metro area (median household income, the percentage of students eligible for free and reduced-price lunch, and the racial composition of the study body). In addition to the demographic information, participants are randomly assigned to receive either 1) average achievement data, 2) average growth data, 3) both average achievement data and average growth data, or 4) neither (the control group). Based on these data, participants choose their preferred school district. Figure 1 displays the information visible to participants assigned to receive both achievement and growth data. After selecting a school district in the New York area, participants repeat this process for the four remaining metro areas: Los Angeles, Chicago, Houston, and Phoenix.

[Figure 1]

Within each metro area, the sequence of school district options is randomized in order to prevent ordering effects (e.g., participants always selecting the first option). However, the sequence of academic and demographic characteristics within each district remains constant. This allows participants to compare districts on any given characteristic with relative ease, but it does not allow us to test whether the order of information influences participants' choices. Moreover, the order of metro areas (first New York, then Los Angeles, etc.) is constant for all participants. If participant choice behavior varies over the course of the experiment (e.g., making more careful choices in the first task than the last task), we would be unable to distinguish those effects from the effects of the metro area context.

Before completing the survey, participants answer a small battery of demographic questions about their age, gender, race/ethnicity, family income, political party identification, political ideology, the presence of children in the household, and the presence of public school students in the household.

Our experimental design allows us to observe how the provision of achievement and/or growth data shapes participants' district choices with respect to racial and socio-economic composition. We do not examine participants' comprehension of these metrics, nor do we shed light on the relative weight that participants place on various pieces of information in their decision-making. There are also a broad range of additional real world factors that affect the choice environment that we do not include in our survey experiment: housing prices, commute times, crime rates, etc. We limit the information set to measures of academic performance and demographics in order not to overwhelm participants with data as they choose between districts. The inclusion of more descriptive characteristics for each district could plausibly reduce the magnitude of our estimates of the impact of academic performance information, but it is unlikely that such additions would change the direction of the effect. By focusing on how the distribution of information that better captures educational effectiveness—instead of the characteristics of the student body—affects participants' district choices, we zero in on a factor that can be manipulated directly and readily by policy-makers and so has substantial potential to inform immediate policy choices.

Analytic Approach

To check for balance between experimental groups, we compare the demographic composition of the control group with the demographic compositions of each of the other randomly assigned groups. To accomplish this, we use a series of OLS regressions:

$$X_i = a + bA_i + cG_i + dB_i + u_i,$$

where X_i is one of the available demographic covariates (derived from the battery of demographic questions asked at the end of the survey); A_i, G_i, and B_i are indicators of experimental group status (the achievement group, the growth group, and the both group); and u_i is the error term for individual i.

To assess the representativeness of the subject pool, we compare the demographic composition of the whole sample to the demographic composition of the 2016 American National Election Study. We provide this comparison to inform judgements about how to generalize our findings. As previously noted, the MTurk sample is unlikely to be demographically representative of the US adult population. Accordingly, we forgo statistical testing.

To convey the descriptive relationships between district demographics, average achievement, average growth, and participants' choices, we present our data in three ways. First, we present descriptive statistics for the 25 districts featured in our experiment. Second, we graph the relationships between district-level demographic variables and either average achievement or average growth (for the 25 districts in our experiment as well as for all districts in the country). Third, we graph the distribution of control group participants' district choices in each metro area.

To examine whether the provision of achievement and/or growth data causes participants to choose higher achieving and/or higher growth districts, we regress each outcome (the average achievement percentile or the average growth percentile of participants' chosen districts) on indicators of experimental group status. To examine whether the provision of achievement and/or growth data causes participants to choose districts with different racial and socio-economic compositions, we regress each outcome (the median household income, the percentage of students eligible for free and reduced-price lunch, or the percentage of white students in participants' chosen districts) on indicators of experimental group status. For these analyses, we use the following general model:

$$Y_i = a + bA_i + cG_i + dB_i + \boldsymbol{eX_i} + u_i,$$

where Y_i is the outcome and $\boldsymbol{X_i}$ is a vector of individual-level demographic characteristics. The default comparison is to the control group, but we are also interested in calculating average treatment effects with the achievement group considered as the comparison group. We run all regressions twice: with and without $\boldsymbol{X_i}$. In the appendices, we report the results from both analyses. In the text, we focus on the unadjusted estimates. There were no substantive differences between these two sets of results.

When calculating heterogeneous treatment effects, we use the following general model:

$$Y_i = a + bA_i + cG_i + dB_i + eZ_i + f(A_iZ_i) + g(G_iZ_i) + h(B_iZ_i) + u_i,$$

where Z_i is one of the following demographic variables:

1. The presence of children age 0-17 in the household
2. The presence of public school students age 0-17 in the household
3. Race (white or person of color)
4. Family income (under or over $75,000)

Because the heterogeneity analyses require a notable increase in the number of statistical tests and therefore an increase in the likelihood of false positives, the results should be viewed as exploratory.

To estimate treatment effects across all five metro areas, we reorganize the data into a long-form dataset. Each participant appears in the dataset five times: once for each school district selection in a metro area. When analyzing these data, we cluster standard errors at the individual level.

Findings

Representativeness and Balance

We begin by presenting the results of the initial 2,500 participant experiment conducted in October 2018. Table 1 displays the frequency of participants' demographic characteristics by experimental condition. It also displays the frequency of the same characteristics in the 2016 American National Election Study (ANES), a nationally representative survey of political attitudes. Compared to the US as a whole, the subject pool is younger, better educated, more liberal, more likely to have a school-age child, and more likely to have a child enrolled in a public school. On other dimensions—such as sex, race/ethnicity, family income, and political party identification—the subject pool roughly approximates the characteristics of the population. We consider the implications of the representativeness of the subject pool as well as other potential threats to external validity in a later section.

[Table 1]

Random assignment effectively established experimental conditions with similar demographic compositions. There are only two instances in which the demographic profile of an experimental condition is statistically different from the control group: the growth group is modestly more likely to have earned a college degree and modestly less likely to have only completed some college. Given the number of statistical tests (21 demographic variables and four experimental conditions), we expect a handful of comparisons to be statistically significant by chance alone.

There are only a few instances of missing covariate data: about one or two percent of cases in each experimental condition. In all subsequent analyses, we impute an arbitrary value for the missing information and control for an indicator of missingness. There are no instances of missing outcome data. All participants were required to make a school district selection in each metro area before continuing with the survey.

Descriptive Analysis of Demographics, Achievement, Growth, and Choice

Table 2 displays descriptive statistics for the five largest school districts by enrollment in the metro areas of New York, Los Angeles, Chicago, Houston, and Phoenix. These values focus on students in grades 3-8 in each district. Note that the degree of variation in achievement, growth, and demographic composition differ from metro area to metro area. While there are dramatic differences in academic performance between the five largest New York area districts, there are smaller differences between the Los Angeles area districts. Similarly, while there is considerable demographic consistency from one district to another in the Phoenix area, there is greater demographic variation between districts in the Chicago area. As we demonstrate later, this variation in district-level characteristics—or the lack thereof—has meaningful consequences for the magnitude of the effects of receiving growth data on participants' school district choices.

[Table 2]

In some metro areas, there are clear examples of relatively high growth districts that serve a large proportion of disadvantaged students (e.g., Jersey City Public Schools in the New York area and Chicago Public Schools in the Chicago area). The presence of these high growth districts that serve predominantly low-income and minority students is key to the research questions that guide this study. Where there are high performing educational options that enroll proportionally fewer white and affluent students, we are interested to examine whether the provision of information alerting participants to these options induces some participants to alter their district choices.

Figure 2 presents the relationships between the district-level demographic variables (median household income, percent free and reduced-price lunch, and percent white) and either average achievement or average growth. We see that all three demographic variables are strongly related to average achievement—both for the 25 districts in our experiment as well as for all districts in the country. By contrast, the relationships between district demographics and average growth are much weaker, and there are many exceptions to the rule. In all districts nationwide and in the subset of districts we explore in our experiment, there are relatively high growth districts that serve a disproportionate number of disadvantaged students.

[Figure 2]

Figure 3 displays the percentage of control group participants that chose each district in each metro area. Consistent with conventional wisdom, the most popular districts—Toms River in the New York area; Long Beach in the Los Angeles area; Indian Prairie and Schaumburg in the Chicago area; Katy in the Houston area; Chandler, Deer Valley, and Mesa in the Phoenix area—tend to serve whiter and more affluent student bodies than other districts in their metro areas. It is also noteworthy that participants tend to select districts that serve populations outside of the central city. The survey does not indicate whether districts are primarily urban or suburban, suggesting that participants' choices may be shaped not only by the data we present but also by their prior beliefs about the school districts.

[Figure 3]

These overall patterns indicate that participants make deliberate, non-random choices between districts and that their preferences are consistent with our expectations based on previous research and conventional wisdom. In other words, it is not particularly surprising that participants tend to choose whiter and more affluent districts. In fact, these patterns give us greater confidence that the subject pool is a relatively typical subset of the population. Moreover, although we acknowledge that some responses may be conditioned by social desirability bias (i.e., that some participants feel pressure to select more racially and socio-economically diverse districts than they would pursue in reality), the distribution of choices suggests that such bias is relatively minor. Relatedly, due to our experimental design, the extent to which social desirability bias influences participants' choices ought to be equally distributed across all experimental conditions, leaving our estimates of the average treatment effects intact.

Experimental Results

First, we examine whether the provision of achievement and/or growth data causes participants to choose higher performing districts on those measures. The results of this analysis are displayed in Figure 4 (to view the results in tabular form, please see Table A1 in Appendix A). Compared to participants who receive no academic performance data, participants who receive only achievement data tend to choose districts that score about three percentile points higher in terms of achievement. Similarly, compared to the control group, participants who receive only growth data tend to choose districts that score about five percentile points higher in terms of growth. Interestingly, participants who receive both achievement and growth data tend to choose districts with similar achievement but higher growth by about two and a half percentile points compared to the districts chosen by their counterparts in the control group. The provision of growth data—even in conjunction with achievement data—appears to be particularly salient, leading participants towards higher growth districts. This also suggests that participants can make distinctions between the two measures of academic performance when provided both.

While this analysis is not in our pre-analysis plan, the results help elucidate the mechanisms behind the findings that follow. Specifically, they show that the provision of achievement and/or growth data leads participants to choose higher achieving and/or higher growth districts. The different underlying relationships between average achievement, average growth, and student demographics produce the pattern of results we describe below.

[Figure 4]

Figure 5 displays the primary results of this study (see also Table A2). Compared to participants who receive no academic performance data, participants who receive average achievement data tend to choose even whiter and more affluent districts. This is consistent with the strong relationship between student demographic composition and average achievement. The relationship between student demographic composition and average growth, on the other hand, is more diffuse. Participants who receive only average growth data tend to choose notably less white and less affluent districts. Participants who receive both average achievement data and average growth data tend to choose districts with similar demographic compositions as their peers in the control group. However, these participants tend to choose districts that are less white and less affluent than their peers who only receive average achievement data. In short, the provision of growth data induces participants to choose less privileged districts on average. This is particularly true for participants who only receive growth data, but the provision of both forms of academic performance data also shifts choices relative to the provision of achievement data alone.

[Figure 5]

We calculate standardized effect sizes (ES) by dividing each estimate of the average treatment effect by the standard deviation of the same outcome in the control group. Relative to their peers in the control group, participants who receive only achievement data tend to choose districts with about \$2,300 higher median household incomes (ES = 0.10), about two percentage points fewer students eligible for free and reduced-price lunch (ES = -0.09), and about two percentage points more white students (ES = 0.08). Alternatively, participants who receive only growth data tend to choose districts with about \$2,900 lower median household incomes (ES = -0.13), about two percentage points more students eligible for free and reduced-price lunch (ES = 0.09), and about four percentage points fewer white students (ES = -0.17). The choices of participants who receive both achievement and growth data are indistinguishable from the choices of participants in the control group across these three dimensions.

Relative to their peers in the achievement group, participants who receive only growth data tend to choose districts with about \$5,100 lower median household incomes (ES = -0.24), about four percentage points more students eligible for free and reduced-price lunch (ES = 0.18), and about six percentage points fewer white students (ES = -0.25). Moreover, participants who receive both achievement and growth data tend to choose districts with about \$2,800 lower median household incomes (ES = -0.13), about two percentage points more students eligible for free and reduced-price lunch (ES = 0.08), and about three percentage points fewer white students (ES = -0.10).

Figure 6 displays the primary results disaggregated by metro area (see also Table A3 and A4). The same basic pattern—in which the growth group tends to choose less white and less affluent districts than the control and achievement groups—occurs in the New York, Los Angeles, Chicago, and Houston metro areas. However, the pattern reverses in the Phoenix metro area. The reason for this is straightforward. Among the five largest districts in the New York, Los Angeles, Chicago, and Houston metro areas, there is at least one relatively high growth district that serves a relatively low-income and/or non-white student body. By contrast, the highest growth district in the Phoenix area serves a relatively high-income and white student body. The racial and socio-economic implications of the distribution of growth data are conditional on the demographic compositions of higher growth districts in a given area.

[Figure 6]

Moreover, the magnitudes of the average effect sizes are conditional on the level of demographic variation between districts. We tend to see the largest effects in the New York area where the highest achieving district and the highest growth district have a 69 percentage point difference in the proportion of white students, a 51 percentage point difference in the proportion of students eligible for free and reduced-price lunch, and a $41,000 difference in the median household income. We see smaller but same-signed effects in the Los Angeles, Chicago, and Houston areas, where there is also less overall demographic variation between districts.

Although we think that Figure 6 provides the most straightforward presentation of metro-level results, this figure makes it difficult to visualize the size of the effects relative to the demographic context in each location. Figure 7 displays the same results with outcomes standardized within each metro area (i.e., we divide each outcome by the standard deviation of the relevant demographic characteristic in each metro area) and centered around the control group mean. In New York, Los Angeles, Chicago, and Houston, the average differences between the control group and the growth group tend to be between one-tenth and four-tenths of a standard deviation, and the average differences between the achievement group and the growth group tend to be between one-tenth and six-tenths of a standard deviation. While the standardized effect sizes also vary by metro area, they fall within a narrower and more intuitive range than their analogous unstandardized counterparts.

[Figure 7]

Lastly, this pattern of results gives us confidence that the sequence of the metro areas did not unduly influence participant behavior. In both the first task (New York) and the last task (Phoenix), participants who receive achievement and/or growth information tend to choose higher performing districts on those measures, with the expected implications for district demographics. However, it is also possible that the larger magnitudes in the New York metro area may be partially attributable to participants' greater conscientiousness with the first task.

Treatment Effect Heterogeneity

Figure 8 (see also Table A5) explores the possibility of heterogeneous treatment effects by the presence of children in the household, the presence of children in the household who attend a public school, race (white or person of color), and family income (greater or less than $75,000 a year). The demographic characteristics that structure our investigation of heterogeneity are laid out in our pre-analysis plan in order to place an intentional limit on the number of comparisons we explore and the likelihood that we mistakenly report false positives. Nevertheless, given the proliferation of statistical tests that accompany this investigation, the reader should view these findings as exploratory.

[Figure 8]

Previous research suggests that parents—particularly those of children in public school—tend to respond differently to survey questions about public education than their peers without children (Bali, 2016; Chingos, Henderson, & West, 2012; Loveless, 1997; Peterson, Henderson, & West, 2014). Parents tend to be more sanguine about the quality of local schools, and they are more likely to hold beliefs about academic performance that align with official measures of student achievement. They may also have taken the survey more seriously than non-parents, given their direct connection to the subject matter. However, we find no evidence of treatment effect heterogeneity by the presence of children in the household (in general or specific to children who attend a public school).

We also find no evidence of heterogeneous treatment effects by race. Earlier research suggests that both families of color and white families tend to prefer educational environments where their child would not be part of a small racial minority (Glazerman & Dotter, 2017; Henig, 1990). Given our overall findings that the provision of growth data causes participants to choose districts with a larger proportion of non-white students, we might expect a different pattern of responses among individuals of color. However, this is not the case in our study.

Alternatively, we do find tentative evidence of treatment effect heterogeneity by income. The effects of providing academic performance data of any kind tend to be larger for participants with family incomes greater than $75,000. Upon receiving achievement and/or growth data, these participants tend to make larger overall shifts towards less white and less affluent districts than their counterparts with family incomes less than $75,000. This analysis does not shed light on why we see larger effects among more affluent participants, but it is consistent with Hastings, Kane, and Staiger's (2006) finding that higher income individuals are more responsive to academic performance data when selecting schools.

Replication Results

Four months after the initial 2,500 participant study, we replicated the experiment with 2,500 different participants using an identical research design. Any given experiment is subject to the possibility that one or more relevant but unobservable background characteristics remain unbalanced between experimental conditions. Moreover, it is possible that the effects we observe are specific to a certain subject pool, a certain point in time, or an idiosyncratic feature of the survey administration. We replicated our experiment to probe the robustness of our findings to these potential threats. The results of the replication are almost entirely consistent with those of the original experiment. On average, the provision of growth data continues to cause participants to choose less white and less affluent districts. Moreover, the provision of both achievement and growth data continues to cause participants to choose less white and less affluent districts than the provision of achievement data alone. There are, however, two findings from the initial study that did not replicate. First, in the replication experiment, participants in the achievement group still choose whiter and more affluent districts than their counterparts in the control group, but the difference is not statistically significant. Second, we no longer observe evidence of treatment effect heterogeneity by family income. Given the number of statistical tests we conduct in our heterogeneity analyses, the likelihood of a false positive in either the initial experiment or the replication experiment is nontrivial. For figures and tables documenting all replication experiment results, please see Appendix B.

External Validity

Our research design offers some unique advantages and disadvantages regarding external validity. First, our subject pool is not a probability sample drawn from the entire US adult population. However, MTurk participants are remarkably diverse, coming from a wide variety of racial, ethnic, economic, educational, political, and regional backgrounds. Our participants are also disproportionately likely to have children in their households, allowing us to over-represent the most relevant subset of the population with respect to school district enrollment decisions. Less desirably, our subject pool contains relatively few individuals with a family income below $25,000 or without a high school diploma, making it difficult to draw inferences about the choices of low-income communities. Moreover, our survey is only available in English, limiting our understanding of how these dynamics apply to speakers of other languages. Given the non-probability nature of the sample, there are presumably a range of other unobservable differences between our subject pool and the population as a whole. To the extent that the effects we identify vary systematically by these observable and unobservable differences, we should expect our findings to differ from those of an analogous experiment employing a nationally representative sample. However, the absence of consistent evidence of treatment effect heterogeneity along observable dimensions offers some confidence that the average effects identified here are generalizable to other populations.

It is also clear that choosing a school district out of a list of five options in a survey is not substantively equivalent to choosing an actual school district for one's child. Our online survey experiment exists in an abstract environment with intentionally limited information and no direct consequences for the participants. However, our experimental context grants us considerable control over the factors under consideration, allowing us to identify the effects of providing achievement and/or growth data on district choices without the complications of distributing information in a more naturalistic setting with preoccupied recipients, conflicting claims from other sources, and highly variable timetables for making educational choices. Furthermore, our study contains important elements of verisimilitude. Participants are asked to make choices between real districts using actual educational and demographic data. They are also likely to come into the experiment with prior beliefs about the quality of the district options, contributing to the realism of the decision-making process.

Conclusions

Our study is motivated by two basic observations. First, average student growth is arguably a better way to measure schools' and districts' contributions to student learning than average student achievement at a single point in time. Second, the relationship between student demographics and growth is much weaker than the relationship between student demographics and achievement. These two observations prompt our primary research question: would the distribution of academic performance data based on student growth influence individuals' school district preferences in ways that run counter to the conventional wisdom that the "best" districts are almost always the whitest and most affluent districts?

There are plenty of reasons to think that this would not be the case. An extensive research literature persuasively documents that families' decisions about their kids' educational options often rely heavily on information about the racial and socio-economic composition of the student body. Small distinctions in the measurement of academic performance may be irrelevant in the face of demographic preferences. Alternatively, it may be the case that families are uninterested in or unconvinced by data on the rate of student progress, preferring to base educational decisions on the overall level of academic performance of their child's potential peers. Moreover, informational interventions designed to promote racial and socio-economic integration in the context of school choice have produced a relatively lackluster track record.

However, our results suggest some reason for optimism for an approach that could steer families towards high quality educational options that serve a wider range of students. We find that the provision of average student growth data can cause individuals to choose less white and less affluent school districts from a set of real options in a series of metropolitan areas. Furthermore, the provision of both average achievement data and average growth data can cause individuals to choose less white and less affluent districts than the provision of achievement data alone. In our initial experiment, the largest effects tend to occur among the most affluent individuals in our subject pool, which is consistent with previous research suggesting that academic performance data plays a particularly large role in the decision-making of high-income families. However, this effect size difference by income was smaller and non-significant in the replication experiment.

These results appear to be dependent on one important condition: the presence of a relatively high growth district that serves a relatively disadvantaged student body. In the absence of such a district (e.g., among the five largest districts in the Phoenix area), we do not observe an effect of the provision of growth data on the racial and socio-economic compositions of participants' district choices. Moreover, the magnitude of the effect appears to be dependent on the overall level of demographic variation between districts. Among the five largest districts in the New York area, the highest achievement district is overwhelmingly white while the highest growth district educates mostly students of color, resulting in a large average difference in the racial compositions of the district choices made by participants in the achievement and growth groups. By contrast, there is less district-level demographic variation in Houston, resulting in the same general pattern of results but much smaller average effect sizes.

There are some obvious limitations to our study. Chiefly, we examine school district choice in the context of an online survey experiment. The act of choosing a district for one's child is both more constrained (by resources, mobility, employment, discrimination, etc.) and more complex (featuring a far greater variety of information and much higher stakes) than presented in our highly stylized experimental environment. We do not expect that our findings will translate seamlessly to the lived experiences of families making difficult educational choices. However, we do think our study provides useful insight to researchers examining the effects of distributing student growth data in more realistic settings as well as educational leaders making decisions about how to measure and report information on academic performance. As districts and states respond to federal requirements to create and distribute public report cards on state, district, and school performance and progress, it will be useful to consider how choices of measurement and emphasis can influence families' behavior. The same is true for districts developing and employing universal enrollment systems in which families use a common application to rank school preferences. We also believe that our study could inform the work of non-governmental organizations, such as GreatSchools.org, that provide data about academic performance to the community.

This experiment also contributes to our broader understanding of the role of information in residential and school choice. No single metric can capture all relevant aspects of a complex, multi-dimensional concept like educational quality. When making choices about which measurements to employ, which results to distribute, and which elements of those results to emphasize, educational institutions need to be aware of how these metrics can be shaped not only by variation in the relevant construct at the institutional level—schools' and districts' effects on student outcomes—but also by systematic variation in the advantages and disadvantages experienced by the individuals who comprise those institutions. When we inaccurately attribute differences in educational quality to school districts because of the students they serve rather than their effectiveness in serving those students, we shortchange both district and student. When that misattribution maps closely to racial and socio-economic lines, we exacerbate long-standing inequalities.

References

Abdulkadiroglu, A., Pathak, P. A., Schellenberg, J., & Walters, C. R. (2019). Do parents value school effectiveness? NBER Working Paper No. 23912.

Auty, W., Bielawski, P., Deeter, T., Hirata, G., Hovanetz-Lassila, C., Rheim, J., Goldschmidt, P., O'Malley, K., Blank, R., & Williams, A. (2008). Implementer's guide to growth models. Washington, DC: Council of Chief State School Officers.

Bali, V. A. (2016). Evolving trends in public opinion on the quality of local schools. *Educational Policy, 30*(5), 688-720.

Barrows, S., Henderson, M., Peterson, P. E., & West, M. R. (2016). Relative performance information and perceptions of public service quality: Evidence from American school districts. *Journal of Public Administration Research and Theory, 26*(3), 571-583.

Bayer, P., Ferreira, F., & McMillan, R. (2007). A unified framework for measuring preferences for schools and neighborhoods. *Journal of Political Economy, 115*(4), 588-638.

Berinksy, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating online labor markets for experimental research: Amazon.com's Mechanical Turk. *Political Analysis, 20*(3), 351-368.

Black, S. E. (1999). Do better schools matter? Parental valuation of elementary education. *Quarterly Journal of Economics, 114*(2), 577-599.

Bolsen, T., Druckman, J. N., & Cook, F. L. (2014). The influence of partisan motivated reasoning on public opinion. *Political Behavior, 36*(2), 235-262.

Burnett, D. (2017). ESSA brings user-friendly makeover of state report cards. *Education Week*.

Castellano, K. E., & Ho, A. D. (2013). A practitioner's guide to growth models. Washington, DC: Council of Chief State School Officers.

Chingos, M. M., Henderson, M. B., & West, M. R. (2012). Perceptions of government service quality: Citizen evidence from public schools. *Quarterly Journal of Political Science, 7*(4), 411-445.

Clinton, J. D., & Grissom, J. A. (2015). Public information, public learning, and public opinion: Democratic accountability in education policy. *Journal of Public Policy, 35*(3), 355-385.

Coppock, A. (2018). Generalizing from survey experiments conducted on Mechanical Turk: A replication approach. *Political Science Research and Methods.* Retrieved from https://doi-org.ezproxy.cul.columbia.edu/10.1017/psrm.2018.10

Corcoran, S. P., Jennings, J. L., Cohodes S. R., & Sattin-Bajaj, C. (2018). Leveling the playing field for high school choice: Results from a field experiment of informational interventions. NBER Working Paper.

Data Quality Campaign. (2019). Growth data: It matters, and it's complicated. Retrieved from https://dataqualitycampaign.org/resource/growth-data-it-matters-and-its-complicated/

de Souza Briggs, X., Popkin, S. J., & Goering, J. M. (2010). *Moving to opportunity: The story of an American experiment to fight ghetto poverty.* New York, NY: Oxford University Press.

Dougherty, J., Zannoni, D., Chowhan, M., Coyne, C., Benjamin, D., Guruge, T., & Nukic, B. (2013). School information, parental decisions, and the digital divide: The SmartChoices Project in Hartford, Connecticut. In G. Orfield & E. Frankenberg (Eds.), *Educational delusions? Why choice can deepen inequality and how to make schools fair*. Berkeley, CA: University of California Press.

Druckman, J. N., Peterson, E., & Slothuus, R. (2013). How elite partisan polarization affects public opinion formation. *American Political Science Review, 107*(1), 57-79.

Fahle, E. M., Shear B. R., Kalogrides, D., Reardon, S. R., DiSalvo, R., & Ho, A. D. (2018). Stanford education data archive: Technical documentation, version 2.1. Stanford University Center for Education Policy Analysis.

Figlio, D. N., & Lucas, M. E. (2004). What's in a grade? School report cards and the housing market. *American Economic Review, 94*(3), 591-604.

Fiva, J. H., & Kirkebøen, L. J. (2011). Information shocks and the dynamics of the housing market. *Scandinavian Journal of Economics, 113*(3), 525-552.

Follmer, D. J., Sperlin, R. A., & Suen, H. K. (2017). The role of MTurk in education research: Advantages, issues, and future directions. *Educational Researcher, 46*(6), 329-334.

Friedman, M. (1962). *Capitalism and freedom*. Chicago, IL: University of Chicago Press.

Glazerman, S., & Dotter, D. (2017). Market signals: Evidence on the determinants and consequences of school choice from a citywide lottery. *Educational Evaluation and Policy Analysis*, *39*(4), 593-619.

Glazerman, S., Nichols-Barrer, I., Valant, J., & Burnett, A. (2018) Presenting school choice information to parents: An evidence-based guide. National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences.

Goering, J. M., & Feins, J. (Eds.). (2003). *Choosing a better Life? Evaluating the Moving to Opportunity social experiment*. Washington DC: Urban Institute.

Goldschmidt, P., Roschewski, P., Choi, K., Auty, W., Hebbler, S., Blank, R., & Williams, A. (2005). Policymakers' guide to growth models for school accountability: How do accountability models differ? Washington, DC: Council of Chief State School Officers.

Goldschmidt, P., Choi, K., & Beaudoin, J. P. (2012). Growth model comparison study: Practical implications of alternative models for evaluating school performance. Washington, DC:

Council of Chief State School Officers.

Harris, D. N., & Larsen, M. F. (2015). What do families want (and why)? New Orleans families and their school choices before and after Katrina. Education Research Alliance for New Orleans.

Hartung, J. M., & Henig, J. R. (1997). Housing vouchers and certificates as a vehicle for deconcentrating the poor: Evidence from the Washington, D.C., metropolitan area. *Urban Affairs Review, 32*(3), 403-419.

Hasan, S., & Kumar, A. (2018). Digitization and divergence: Online school ratings and segregation in America. Retrieved from https://ssrn.com/abstract=3265316.

Hastings, J. S., Kane, T. K., & Staiger, D. O. (2006). Parental preferences and school competition: Evidence from a public school choice program. NBER Working Paper 11805.

Hastings, J. S., & Weinstein, J. M. (2008). Information, school choice, and academic achievement: Evidence from two experiments. *Quarterly Journal of Economics, 123*(4), 1373-1414.

Hegedus, A. (2018). Evaluating the relationships between poverty and school performance. Northwest Evaluation Association.

Henig, J. (1990). Choice in public schools: An analysis of transfer requests among magnet schools. *Social Science Quarterly, 71*(1), 69-82.

Hesla, K. (2018). Unified enrollment: Lessons learned from across the country. National Alliance for Public Charter Schools.

Imberman, S. A., & Lovenheim, M. F. (2016). Does the market value value-added? Evidence from housing prices after the public release of school and teacher value-added. *Journal of*

Urban Economics, 91(1), 104-121.

Jacobsen, R., Snyder, J. W., & Saultz, A. (2014). Informing or shaping public opinion: The influence of school accountability data format on public perceptions of school quality. *American Journal of Education, 121*(1), 1-27.

Kane, T. J., Riegg, S. K., Staiger, D. O. (2006). School quality, neighborhoods, and housing prices. *American Law and Economic Review, 8*(2), 183-212.

Ladd, H., & Loeb, S. (2013). The challenges of measuring school quality: Implications for educational equity. In R. Reich & D. Allen (Eds.), *Education, Justice, and Democracy*. Chicago, IL: Chicago University Press.

Loveless, T. (1997). The structure of public confidence in education. *American Journal of Education, 105*(2), 127-159.

Mullinix, K. J., Leeper, T. J., Druckman, J. N., & Freese, J. (2015). The generalizability of survey experiments. *Journal of Experimental Political Science, 2*(2), 109-138.

Peterson, P. E., Henderson, M., & West M. R. (2014). *Teachers versus the public: What Americans think about schools and how to fix them.* Washington, DC: Brookings Institution Press.

Reardon, S. F. (2016). School district socioeconomic status, race, and academic achievement. Stanford University Center for Education Policy Analysis. Retrieved from https://cepa.stanford.edu/content/school-district-socioeconomic-status-race-and-academic-achievement

Reardon, S. F. (2018). Educational opportunity in early and middle childhood: Variation by place and age. Stanford University Center for Education Policy Analysis. Retrieved from https://cepa.stanford.edu/sites/default/files/wp17-12-v201803.pdf

Reardon, S. F., & Hinze-Pifer, R. (2017). Test score growth among Chicago Public School students, 2009-2014. Stanford University Center for Education Policy Analysis. Retrieved from https://cepa.stanford.edu/sites/default/files/chicago%20public%20school%20test%20scores%202009-2014.pdf

Reardon, S. F., Kalogrides, D., & Ho, A. D. (2018). Validation methods for aggregate-level test scale linking: A case study mapping school district test score distributions to a common scale. Stanford University Center for Education Policy Analysis. Retrieved from https://cepa.stanford.edu/sites/default/files/wp16-09-v201807.pdf

Rothstein, J. M. (2006). Good principals or good peers? Parental valuation of school characteristics, Tiebout equilibrium, and the incentive effects of competition among jurisdictions. *The American Economic Review*, *94*(4), 1333-1350.

Schneider, J., Jacobsen, R., White, R. S., Gehlbach, H. (2018). The (mis)measure of schools: How data affect stakeholder knowledge and perceptions of quality. *Teachers College Record, 120*(6), 1-40.

Schneider, M., & Buckley, J. (2002). What do parents want from schools? Evidence from the Internet. *Educational Evaluation and Policy Analysis, 24*(2).

Schneider, M., Teske, P., & Marschall, M. (2000). *Choosing schools: Consumer choice and the quality of American schools*. Princeton, NJ: Princeton University Press.

Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research*, *75*(3), 417-453.

Tiebout, C. M. (1956). A pure theory of local expenditures. *The Journal of Political Economy, 64*(5), 416-424.

Tversky, A., & Kahneman, D. T. (1974). Heuristics and biases: Judgement under uncertainty.

Science, 185(4157), 1124-1131.

van Ewijk, R., & Sleegers, P. (2010). The effect of peer socioeconomic status on student achievement: A meta-analysis. *Educational Research Review, 5*(2), 134-150.

Figures

Figure 1. Excerpt from Survey (New York: "Both Achievement and Growth" Group)

Please imagine that you are a parent who is about to move to the **New York City, NY** metropolitan area. When deciding where to live, one of your top priorities is to choose a school district for your elementary school-age child. You have narrowed your search to five options. Which would you choose?

Please click the button to the LEFT of your choice.

Source: Stanford University Education Data Archive

Yonkers City School District

- Average achievement in grade 3-8: The average student scores at about the 4th grade level (10th percentile nationwide)
- Average growth in grade 3-8: The average student grows about 0.69 grade levels in one school year (2nd percentile nationwide)
- Median household income: $48,095
- Eligible for free/reduced price lunch: 70%
- White students: 18%
- Black students: 23%
- Hispanic students: 54%
- Asian students: 5%

New York City Public Schools

- Average achievement in grade 3-8: The average student scores at about the 5th grade level (36th percentile nationwide)
- Average growth in grade 3-8: The average student grows about 0.92 grade levels in one school year (35th percentile nationwide)
- Median household income: $41,888
- Eligible for free/reduced price lunch: 76%
- White students: 15%
- Black students: 29%
- Hispanic students: 40%
- Asian students: 15%

Paterson School District

- Average achievement in grade 3-8: The average student scores at about the 4th grade level (9th percentile nationwide)
- Average growth in grade 3-8: The average student grows about 1.01 grade levels in one school year (63rd percentile nationwide)
- Median household income: $31,892
- Eligible for free/reduced price lunch: 79%
- White students: 5%
- Black students: 28%
- Hispanic students: 62%
- Asian students: 4%

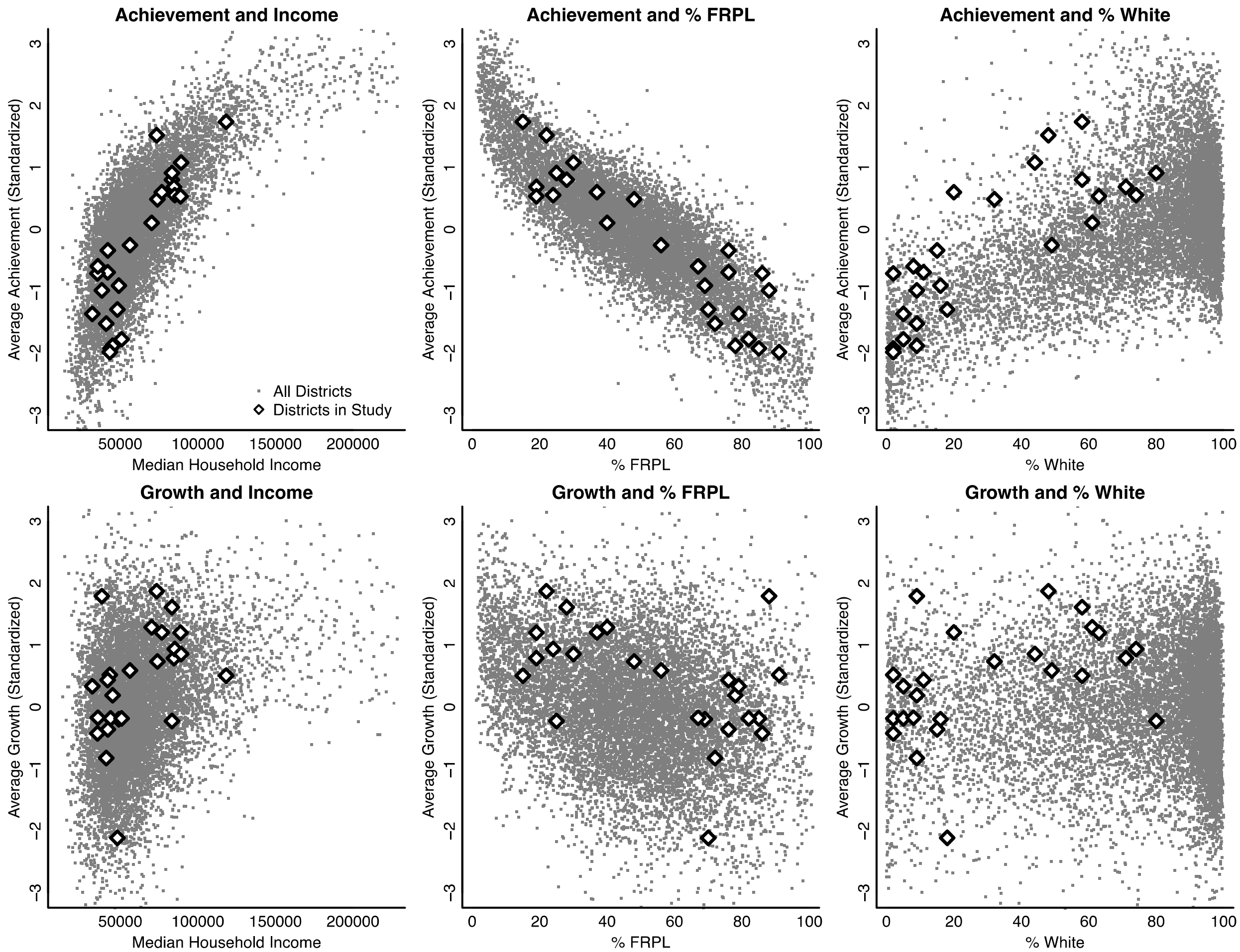
Jersey City Public Schools

- Average achievement in grade 3-8: The average student scores at about the 5th grade level (23rd percentile nationwide)
- Average growth in grade 3-8: The average student grows about 1.02 grade levels in one school year (67th percentile nationwide)
- Median household income: $41,773
- Eligible for free/reduced price lunch: 76%
- White students: 11%
- Black students: 35%
- Hispanic students: 38%
- Asian students: 15%

Toms River Regional School District

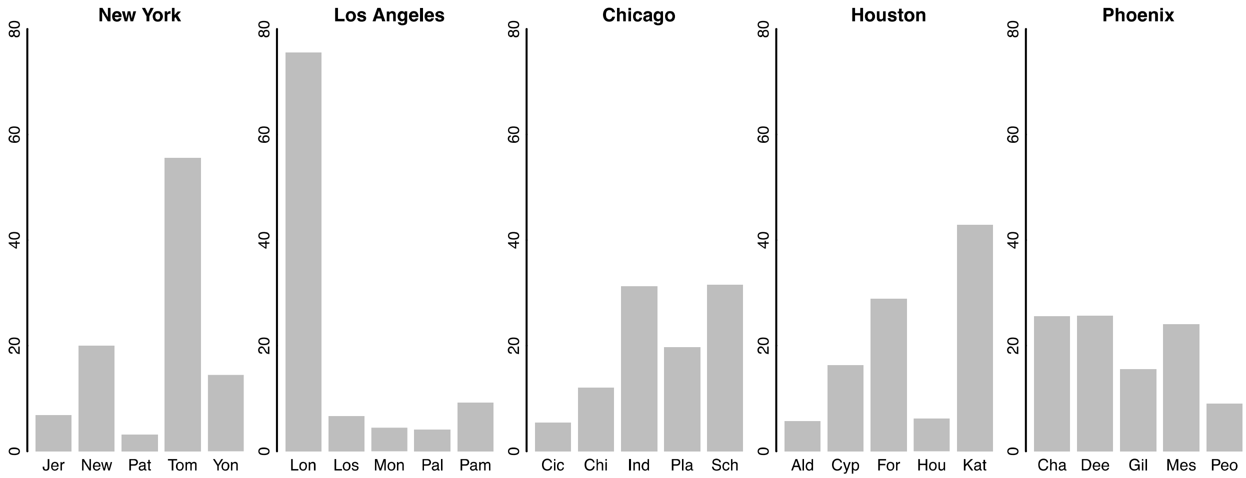
- Average achievement in grade 3-8: The average student scores at about the 7th grade level (84th percentile nationwide)
- Average growth in grade 3-8: The average student grows about 0.93 grade levels in one school year (40th percentile nationwide)
- Median household income: $83,256
- Eligible for free/reduced price lunch: 25%
- White students: 80%
- Black students: 5%
- Hispanic students: 11%
- Asian students: 4%

Figure 2. Achievement, Growth, and Student Demographics in American School Districts (Districts = 10,334)



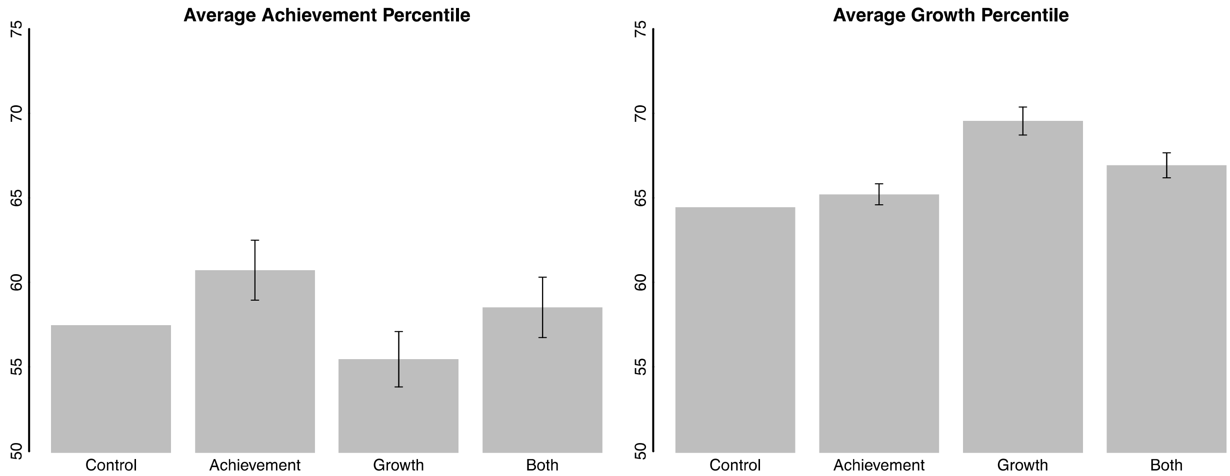
Notes. Points represent geographic school districts (pooled from 2009-2015); diamonds represent the 25 districts featured in the survey experiment. *Source.* Stanford Education Data Archive v2.1

Figure 3. Distribution of District Choices in Control Group (Individuals = 630)



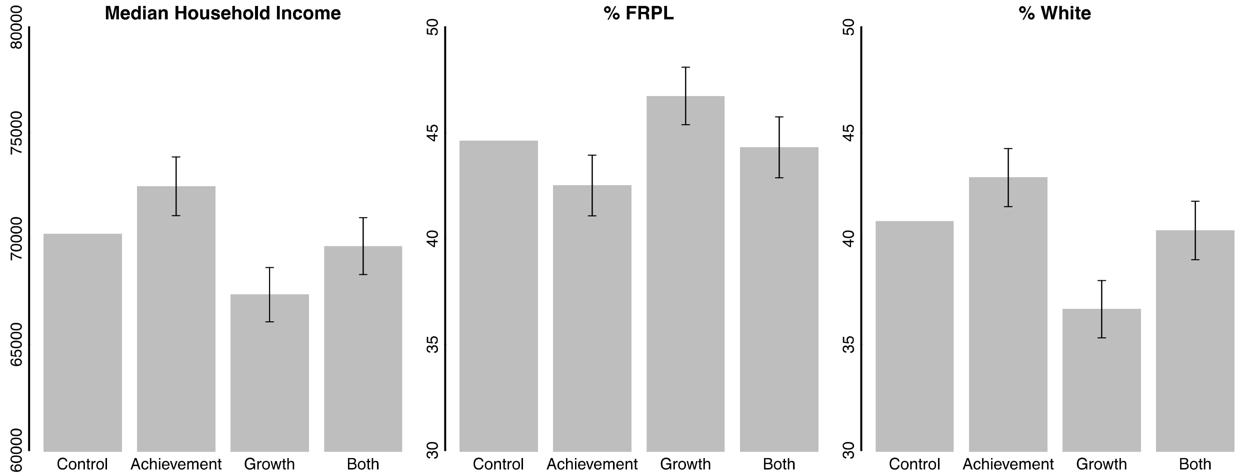
Notes. Y-axis displays percentage choosing each district; district names abbreviated to first three letters

Figure 4. Average Performance of Chosen Districts (Individuals = 2,500; Observations = 12,500)



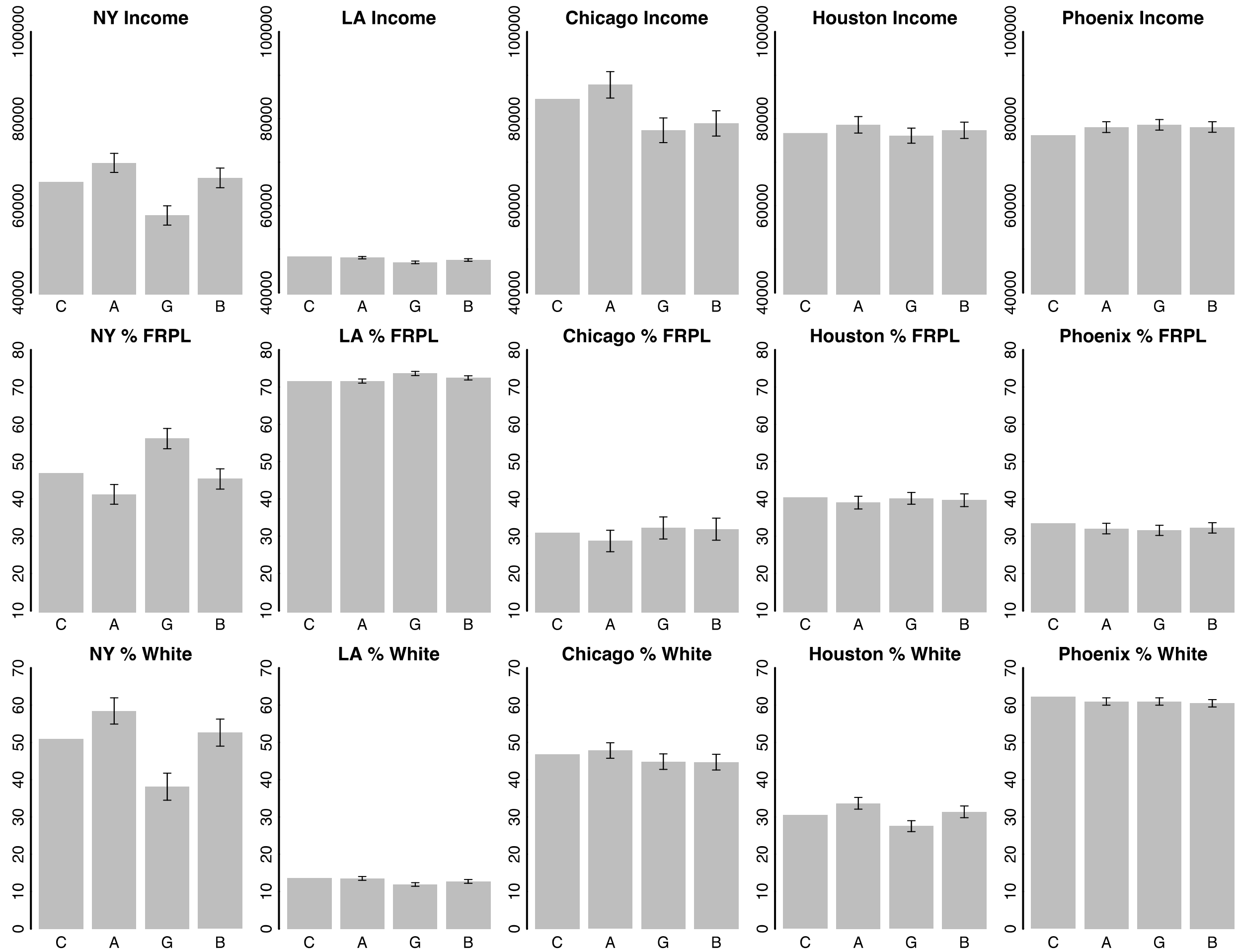
Notes. Error bars represent 95% confidence intervals (robust standard errors clustered at the individual level)

Figure 5. Average Demographics of Chosen Districts (Individuals = 2,500; Observations = 12,500)



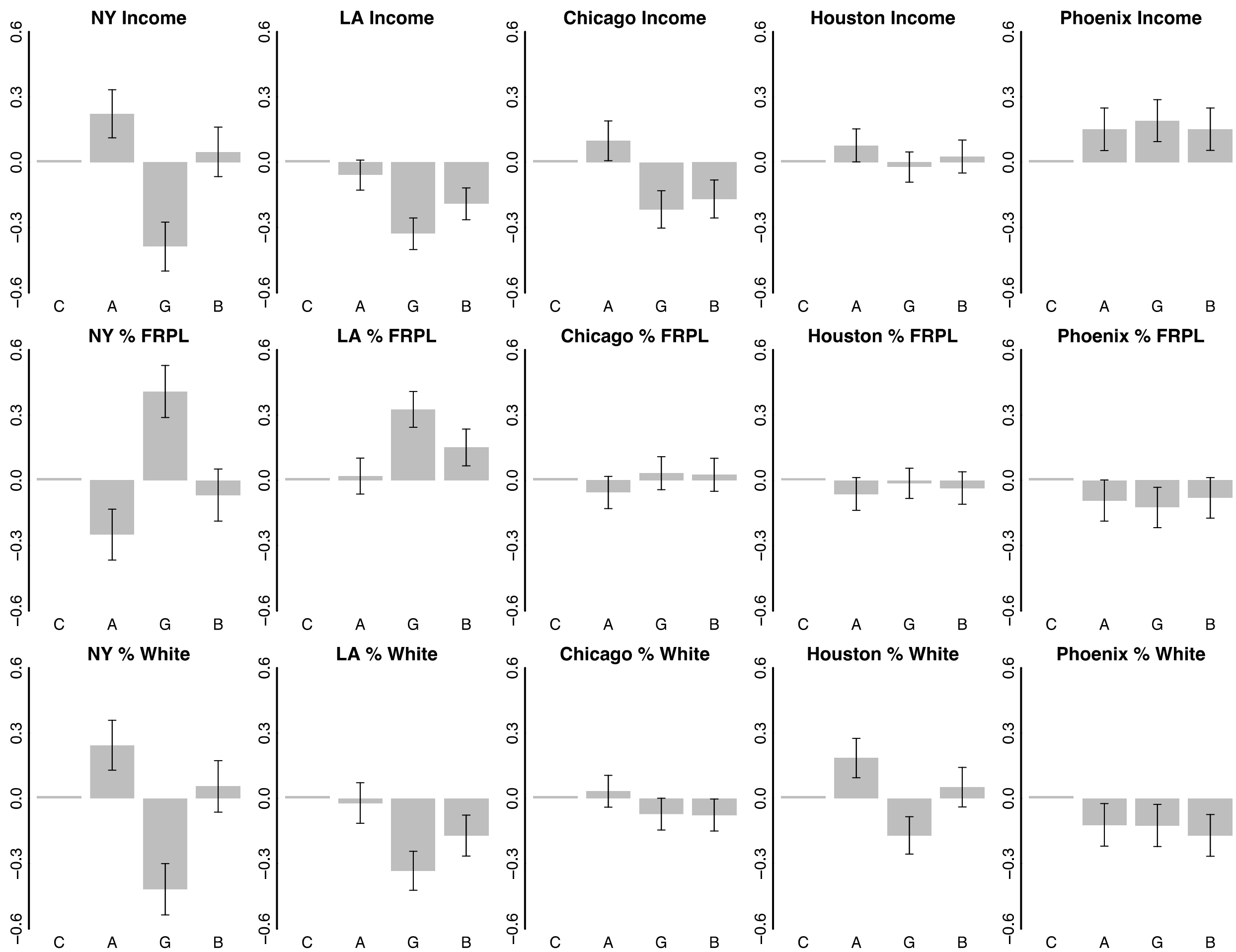
Notes. Error bars represent 95% confidence intervals (robust standard errors clustered at the individual level)

Figure 6. Experimental Results by Metro Area (Individuals = 2,500)



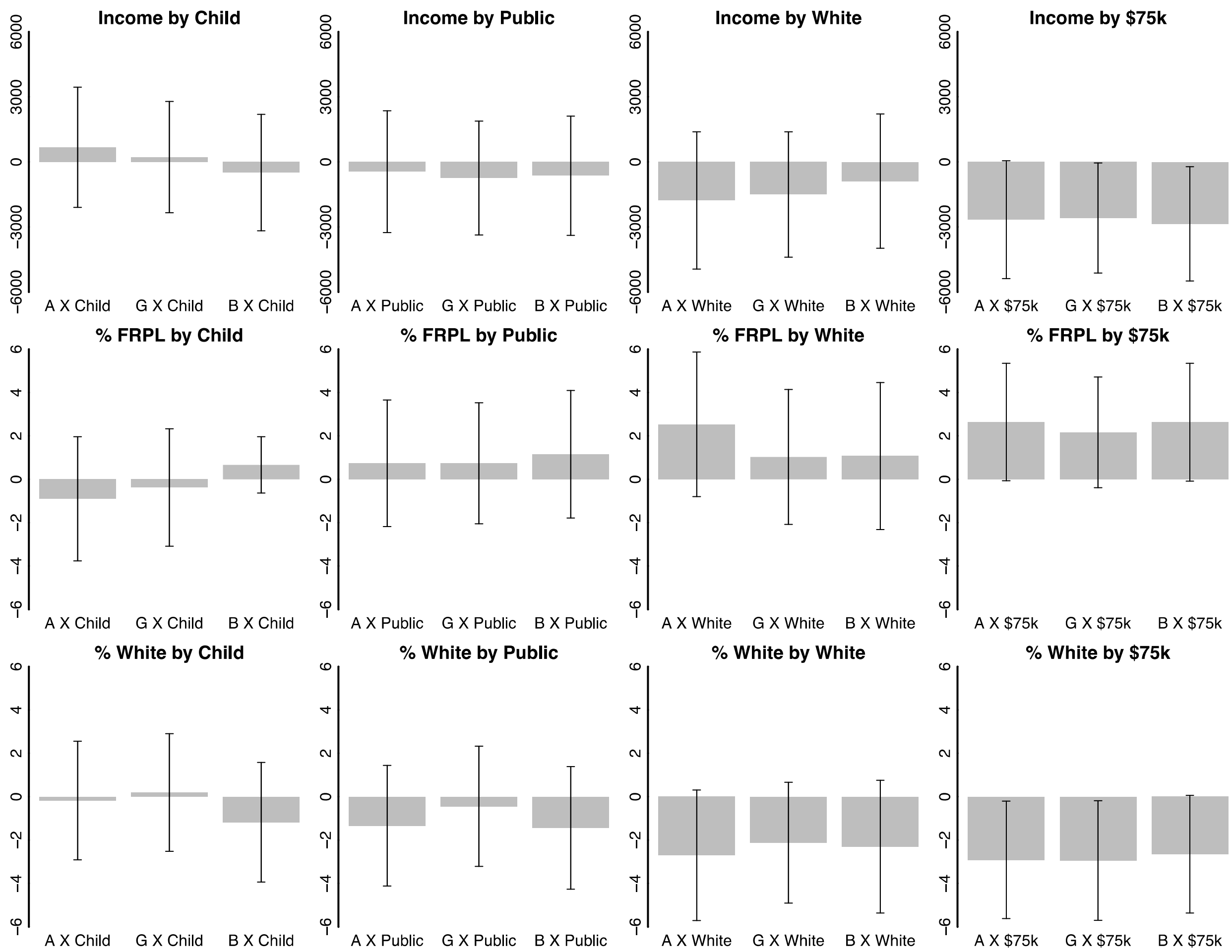
Notes. Error bars represent 95% confidence intervals (robust standard errors)

Figure 7. Experimental Results by Metropolitan Area with Standardized Outcomes (Individuals = 2,500)



Notes. Outcomes standardized within metro area and centered around control group mean; error bars represent 95% confidence intervals (robust standard errors)

Figure 8. Heterogeneous Treatment Effects (Individuals = 2,500; Observations = 12,500)



Notes. Values are OLS coefficients of treatment-by-demographic interaction terms; error bars represent 95% confidence intervals (robust standard errors clustered at the individual level)

Tables

Table 1. Experimental Balance and Representativeness (Individuals = 2,500)

%	Control (*n* = 630)	Achievement (*n* = 625)	Growth (*n* = 620)	Both (*n* = 625)	ANES 2016 (*n* = 4,271)
School-Age Child	53.17	51.84	47.66	51.84	33.73
School-Age Child in Public School	44.04	42.72	39.19	40.06	24.50
Female	51.36	47.43	49.43	51.84	52.90
White	71.70	73.44	74.35	75.00	71.68
Black	11.13	12.00	10.65	10.90	9.39
Asian	6.20	4.96	5.65	5.77	3.49
Native-American	1.11	0.96	0.65	0.96	0.64
Hispanic	7.63	6.24	6.61	5.93	10.62
Other Race	2.23	2.40	2.10	1.44	4.18
Less Than High School	0.80	0.48	0.65	0.16	6.66
High School	11.25	11.95	11.20	11.08	19.26
Some College	34.89	36.03	28.73*	32.10	35.44
College	53.05	51.53	59.42*	56.66	38.63
Family Income < $25,000	14.97	14.90	12.62	13.83	22.63
Family Income $25,000-$50,000	28.82	29.65	31.55	31.67	21.82
Family Income $50,000-$75,000	27.87	28.69	28.16	25.24	18.16
Family Income $75,000-$100,000	15.45	12.50	14.40	15.76	13.10
Family Income > $100,000	12.90	14.26	13.27	13.50	24.28
Party ID (1-7)	3.51	3.43	3.44	3.53	3.86
Ideology (1-7)	3.46	3.47	3.49	3.51	4.18
Age (Years)	36.97	37.28	36.17	36.33	49.58
Missing Covariates	2.22	1.92	1.29	1.12	

Notes. Achievement, Growth, and Both compared to Control; * $p < 0.05$

Table 2. School District Characteristics for Grades 3-8 (2009-2015)

District	Metro Area	Students	Average Achievement	Average Achievement %ile	Average Growth	Average Growth %ile	Median Household Income	% FRPL	% White
New York City	New York	430,463	5	36	0.92	35	41,888	76	15
Jersey City	New York	13,140	5	23	1.02	67	41,773	76	11
Paterson	New York	12,037	4	9	1.01	63	31,892	79	5
Yonkers	New York	11,265	4	10	0.69	2	48,095	70	18
Toms River	New York	7,181	7	84	0.93	40	83,256	25	80
Los Angeles	Los Angeles	297,419	4	7	0.86	20	40,928	72	9
Long Beach	Los Angeles	37,536	5	17	0.94	41	49,042	69	16
Palmdale	Los Angeles	14,107	4	4	0.99	57	45,040	78	9
Montebello	Los Angeles	14,071	4	3	0.94	42	43,669	85	2
Pomona	Los Angeles	12,607	4	4	0.94	42	50,825	82	5
Chicago	Chicago	175,036	5	15	1.19	96	38,036	88	9
Indian Prairie	Chicago	13,476	7	96	1.03	70	118,171	15	58
Plainfield	Chicago	13,392	6	73	1.12	89	88,818	19	63
Schaumburg	Chicago	9,225	7	94	1.20	97	73,412	22	48
Cicero	Chicago	8,523	4	3	1.03	70	43,246	91	2
Houston	Houston	98,228	5	26	0.94	42	35,505	67	8
Cypress-Fairbanks	Houston	49,757	6	71	1.06	77	73,884	48	32
Fort Bend	Houston	32,402	6	75	1.12	89	76,737	37	20
Katy	Houston	29,267	7	88	1.07	80	89,183	30	44
Aldine	Houston	28,988	5	22	0.91	32	35,289	86	2
Mesa	Phoenix	34,108	5	39	1.04	72	56,160	56	49
Gilbert	Phoenix	20,538	6	78	1.06	78	84,586	19	71
Chandler	Phoenix	20,129	6	81	1.17	95	83,199	28	58
Deer Valley	Phoenix	19,470	6	74	1.08	83	85,003	24	74
Peoria	Phoenix	18,198	6	55	1.13	91	70,220	40	61

Source. Stanford Education Data Archive v2.1

Appendix A: Experimental Results in Tables

Table A1. Average Performance of Chosen Districts (Individuals = 2,500; Observations = 12,500)

	Average Achievement Percentile		Average Growth Percentile	
Intercept	57.46*	59.70*	64.45*	64.79*
	(0.63)	(0.66)	(0.22)	(0.25)
Achievement	3.26*	3.37*	0.76*	0.78*
	(0.90)	(0.87)	(0.32)	(0.32)
Growth	-2.01*	-1.94*	5.10*	5.11*
	(0.83)	(0.79)	(0.42)	(0.42)
Both	1.06	1.05	2.48*	2.48*
	(0.91)	(0.87)	(0.38)	(0.37)
Covariates		Yes		Yes

Notes. Values are OLS coefficients (robust standard errors clustered at the individual level); * $p < 0.05$

Table A2. Average Demographics of Chosen Districts (Individuals = 2,500; Observations = 12,500)

	Comparison to Control Group					
	Median Household Income		% FRPL		% White	
Intercept	70,222*	71,920*	44.63*	42.84*	40.83*	42.54*
	(507)	(541)	(0.53)	(0.55)	(0.51)	(0.54)
Achievement	2,250*	2,320*	-2.12*	-2.19*	2.05*	2.12*
	(708)	(682)	(0.73)	(0.71)	(0.70)	(0.67)
Growth	-2,875*	-2,838*	2.11*	2.05*	-4.17*	-4.12*
	(655)	(626)	(0.69)	(0.66)	(0.69)	(0.66)
Both	-576	-586	-0.32	-0.31	-0.45	-0.45
	(687)	(660)	(0.73)	(0.70)	(0.71)	(0.68)
Covariates		Yes		Yes		Yes
	Comparison to Achievement Group					
	Median Household Income		% FRPL		% White	
Intercept	72,472*	74,240*	42.51*	40.65*	42.88*	44.66*
	(494)	(540)	(0.51)	(0.56)	(0.48)	(0.53)
Control	-2,250*	-2,320*	2.12*	2.19*	-2.05*	-2.12*
	(708)	(682)	(0.73)	(0.71)	(0.70)	(0.67)
Growth	-5,124*	-5,158*	4.23*	4.25*	-6.22*	-6.23*
	(645)	(627)	(0.68)	(0.66)	(0.67)	(0.65)
Both	-2,826*	-2,907*	1.80*	1.88*	-2.50*	-2.57*
	(678)	(662)	(0.72)	(0.70)	(0.68)	(0.66)
Covariates		Yes		Yes		Yes

Notes. Values are OLS coefficients (robust standard errors clustered at the individual level); * $p < 0.05$

Table A3. Metropolitan Area Results, Comparison to Control Group (Individuals = 2,500)

	New York					
	Median Household Income		% FRPL		% White	
Intercept	65,442*	68,063*	46.90*	43.71*	50.95*	55.22*
	(802)	(879)	(0.98)	(1.07)	(1.30)	(1.41)
Achievement	4,409*	4,479*	-5.69*	-5.77*	7.49*	7.62*
	(1,120)	(1,080)	(1.36)	(1.31)	(1.80)	(1.73)
Growth	-7,693*	-7,623*	9.28*	9.20*	-12.83*	-12.74*
	(1,139)	(1,101)	(1.39)	(1.34)	(1.85)	(1.79)
Both	941	968	-1.54	-1.57	1.68	1.71
	(1,150)	(1,116)	(1.39)	(1.35)	(1.85)	(1.80)
	Los Angeles					
	Median Household Income		% FRPL		% White	
Intercept	48,261*	48,323*	71.48*	71.50*	13.61*	13.63*
	(98)	(110)	(0.20)	(0.22)	(0.18)	(0.19)
Achievement	-238	-240	0.13	0.16	-0.11	-0.14
	(141)	(140)	(0.28)	(0.28)	(0.25)	(0.25)
Growth	-1,323*	-1,328*	2.18*	2.19*	-1.76*	-1.77*
	(150)	(148)	(0.28)	(0.28)	(0.24)	(0.24)
Both	-769*	-758*	1.01*	1.03*	-0.90*	-0.92*
	(150)	(147)	(0.29)	(0.29)	(0.25)	(0.25)
	Chicago					
	Median Household Income		% FRPL		% White	
Intercept	84,545*	86,986*	30.91*	28.44*	46.89*	48.58*
	(1,091)	(1,189)	(1.07)	(1.14)	(0.77)	(0.81)
Achievement	3,247*	3,349*	-2.20	-2.34	0.92	1.01
	(1,548)	(1,517)	(1.47)	(1.44)	(1.06)	(1.04)
Growth	-7,200*	-7,148*	1.27	1.09	-2.06	-1.96
	(1,453)	(1,415)	(1.51)	(1.45)	(1.07)	(1.04)
Both	-5,615*	-5,706*	0.96	0.98	-2.18*	-2.22*
	(1,475)	(1,443)	(1.51)	(1.47)	(1.07)	(1.05)
	Houston					
	Median Household Income		% FRPL		% White	
Intercept	76,684*	78,748*	40.46*	38.67*	30.48*	32.17*
	(657)	(717)	(0.59)	(0.65)	(0.55)	(0.60)
Achievement	1,927*	2,007*	-1.45	-1.52	3.18*	3.25*
	(971)	(951)	(0.88)	(0.87)	(0.80)	(0.78)
Growth	-562	-525	-0.32	-0.34	-2.94*	-2.87*
	(890)	(868)	(0.81)	(0.79)	(0.75)	(0.74)
Both	655	695	-0.82	-0.84	0.88	0.91
	(966)	(942)	(0.87)	(0.86)	(0.80)	(0.78)
	Phoenix					
	Median Household Income		% FRPL		% White	
Intercept	76,181*	77,478*	33.41*	31.89*	62.24*	63.10*
	(478)	(504)	(0.55)	(0.58)	(0.39)	(0.42)
Achievement	1,903*	2,006*	-1.38	-1.50*	-1.23*	-1.17*
	(630)	(618)	(0.71)	(0.70)	(0.51)	(0.50)
Growth	2,404*	2,436*	-1.86*	-1.89*	-1.26*	-1.25*
	(621)	(612)	(0.70)	(0.69)	(0.50)	(0.50)
Both	1,908*	1,869*	-1.19	-1.15	-1.73*	-1.74*
	(625)	(613)	(0.71)	(0.69)	(0.49)	(0.49)

Notes. Values are OLS coefficients (robust standard errors); for each outcome, estimates in the second column control for all available covariates; * $p < 0.05$

Table A4. Metropolitan Area Results, Comparison to Achievement Group (Individuals = 2,500)

	Median Household Income		% FRPL		% White	
New York						
Intercept	69,850*	72,542*	41.21*	37.94*	58.45*	62.84*
	(782)	(867)	(0.94)	(1.04)	(1.25)	(1.38)
Control	-4,409*	-4,479*	5.69*	5.77*	-7.49*	-7.62*
	(1,120)	(1,080)	(1.36)	(1.31)	(1.80)	(1.73)
Growth	-12,101*	-12,102*	14.97*	14.98*	-20.33*	-20.36*
	(1,125)	(1,100)	(1.36)	(1.33)	(1.82)	(1.78)
Both	-3,467*	-3,511*	4.14*	4.20*	-5.81*	-5.91*
	(1,136)	(1,116)	(1.36)	(1.34)	(1.82)	(1.79)
Los Angeles						
Intercept	48,024*	48,083*	71.61*	71.66*	13.50*	13.49*
	(102)	(115)	(0.20)	(0.23)	(0.18)	(0.20)
Control	238	240	-0.13	-0.16	0.11	0.14
	(141)	(140)	(0.28)	(0.28)	(0.25)	(0.25)
Growth	-1,085*	-1,088*	2.05*	2.03*	-1.64*	-1.63*
	(153)	(152)	(0.28)	(0.28)	(0.24)	(0.24)
Both	-531*	-518*	0.88*	0.87*	-0.79*	-0.77*
	(153)	(152)	(0.29)	(0.29)	(0.25)	(0.26)
Chicago						
Intercept	87,791*	90,335*	28.70*	26.10*	47.82*	49.59*
	(1,097)	(1,212)	(1.01)	(1.12)	(0.72)	(0.81)
Control	-3,247*	-3,349*	2.20	2.34	-0.92	-1.01
	(1,548)	(1,517)	(1.47)	(1.44)	(1.06)	(1.04)
Growth	-10,447*	-10,497*	3.48*	3.44*	-2.99*	-2.97*
	(1,457)	(1,441)	(1.46)	(1.44)	(1.03)	(1.02)
Both	-8,862*	-9,055*	3.17*	3.32*	-3.11*	-3.23*
	(1,479)	(1,471)	(1.46)	(1.45)	(1.04)	(1.03)
Houston						
Intercept	78,611*	80,755*	39.01*	37.15*	33.66*	35.43*
	(715)	(783)	(0.65)	(0.72)	(0.58)	(0.63)
Control	-1,927*	-2,007*	1.45	1.52	-3.18*	-3.25*
	(971)	(951)	(0.88)	(0.87)	(0.80)	(0.78)
Growth	-2,490*	-2,533*	1.13	1.18	-6.12*	-6.13*
	(934)	(915)	(0.85)	(0.84)	(0.78)	(0.76)
Both	-1,273	-1,313	0.63	0.68	-2.30*	-2.34*
	(1,006)	(987)	(0.91)	(0.90)	(0.82)	(0.80)
Phoenix						
Intercept	78,084*	79,484*	32.03*	30.39*	61.00*	61.93*
	(411)	(445)	(0.46)	(0.50)	(0.32)	(0.36)
Control	-1,903*	-2,006*	1.38	1.50*	1.23*	1.17*
	(630)	(618)	(0.71)	(0.70)	(0.51)	(0.50)
Growth	501	430	-0.47	-0.39	-0.03	-0.08
	(571)	(565)	(0.64)	(0.63)	(0.45)	(0.45)
Both	5	-137	0.19	0.35	-0.49	-0.57
	(575)	(566)	(0.64)	(0.63)	(0.45)	(0.44)

Notes. Values are OLS coefficients (robust standard errors); for each outcome, estimates in the second column control for all available covariates; * $p < 0.05$

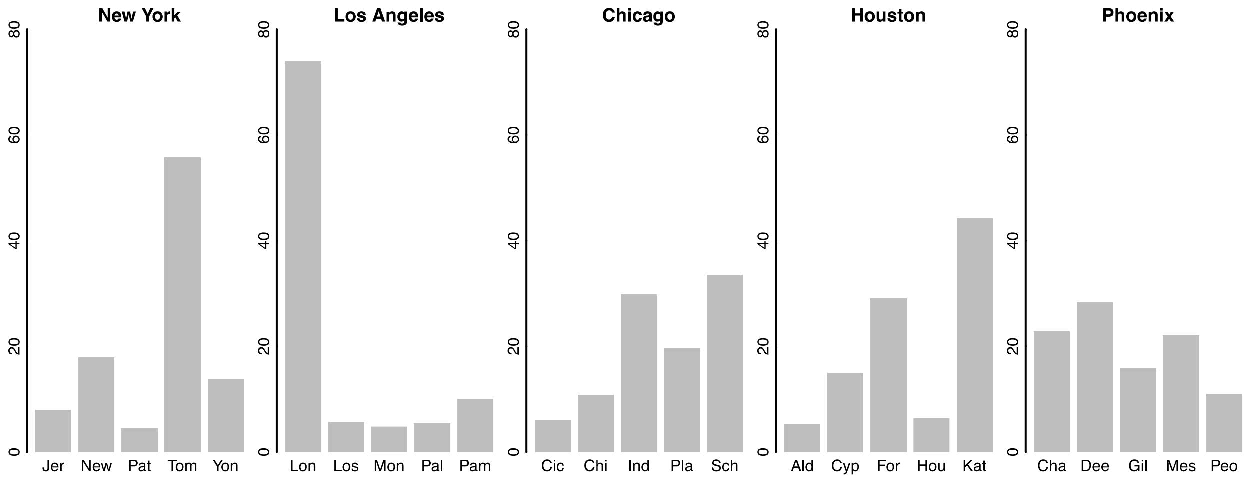
Table A5. Heterogeneous Treatment Effects (Individuals = 2,500; Observations = 12,500)

	Median Household Income				% FRPL				% White			
Intercept	70,906*	70,605*	64,563*	68,056*	43.73*	44.14*	51.26*	46.85*	41.14*	40.93*	33.51*	38.72*
	(714)	(671)	(983)	(618)	(0.73)	(0.69)	(1.06)	(0.66)	(0.72)	(0.67)	(0.92)	(0.63)
Achieve	1,886	2,399*	3,447*	3,142*	-1.63	-2.39*	-3.84*	-3.01*	2.14*	2.59*	3.88*	3.00*
	(1,012)	(927)	(1,413)	(859)	(1.03)	(0.95)	(1.51)	(0.90)	(0.99)	(0.91)	(1.33)	(0.85)
Growth	-3,051*	-2,646*	-1,935	-2,112*	2.38*	1.90*	1.57	1.45	-4.28*	-4.03*	-2.84*	-3.31*
	(904)	(840)	(1,284)	(790)	(0.94)	(0.88)	(1.40)	(0.85)	(0.96)	(0.89)	(1.19)	(0.83)
Both	-324	-358	-150	215	-0.65	-0.73	-0.83	-1.03	0.16	0.12	0.95	0.27
	(954)	(884)	(1,388)	(841)	(1.00)	(0.94)	(1.54)	(0.92)	(0.98)	(0.90)	(1.36)	(0.87)
Child	-1,337				1.73				-0.61			
	(1,010)				(1.05)				(1.02)			
A×Child	675				-0.90				-0.18			
	(1,415)				(1.46)				(1.40)			
G×Child	222				-0.38				0.19			
	(1,309)				(1.38)				(1.38)			
B×Child	-494				0.66				-1.18			
	(1,368)				(1.46)				(1.41)			
Public		-887				1.13				-0.23		
		(1,022)				(1.06)				(1.04)		
A×Public		-452				0.74				-1.34		
		(1,433)				(1.49)				(1.42)		
G×Public		-742				0.74				-0.44		
		(1,340)				(1.42)				(1.41)		
B×Public		-638				1.15				-1.44		
		(1,404)				(1.50)				(1.44)		
White			7,834*				-9.20*				10.17*	
			(1,130)				(1.20)				(1.08)	
A×White			-1,783				2.53				-2.70	
			(1,614)				(1.70)				(1.53)	
G×White			-1,503				1.03				-2.12	
			(1,473)				(1.59)				(1.42)	
B×White			-888				1.07				-2.30	
			(1,578)				(1.73)				(1.56)	
$75k				7,501*				-7.69*				7.36*
				(972)				(0.96)				(0.98)
A×$75k				-2,666				2.64				-2.91*
				(1,387)				(1.38)				(1.38)
G×$75k				-2,589*				2.16				-2.95*
				(1,294)				(1.30)				(1.41)
B×$75k				-2,857*				2.63				-2.65
				(1,345)				(1.39)				(1.38)

Notes. Values are OLS coefficients (robust standard errors clustered at the individual level); * $p < 0.05$

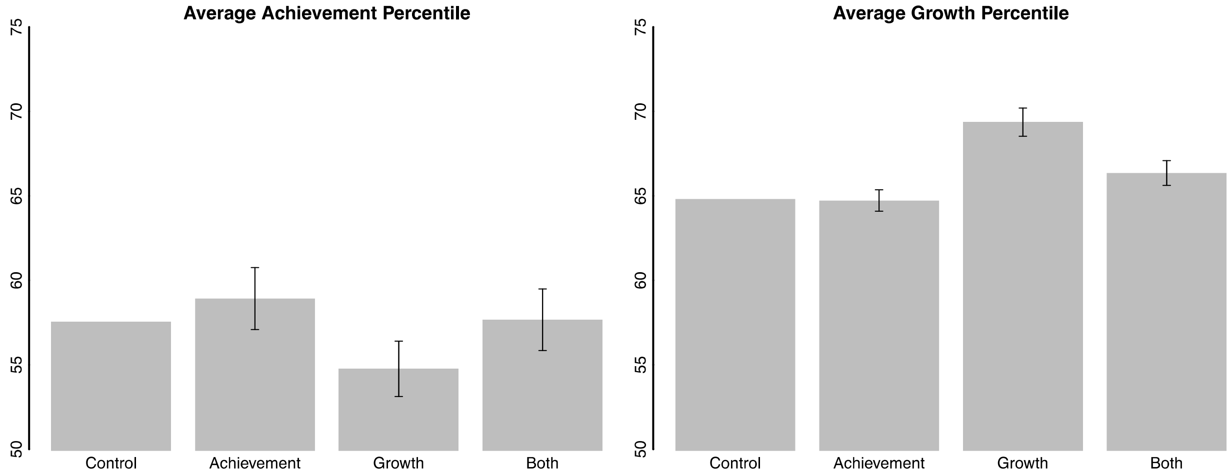
Appendix B: Replication Results

Figure B1. Replication: Distribution of District Choices in Control Group (Individuals = 626)



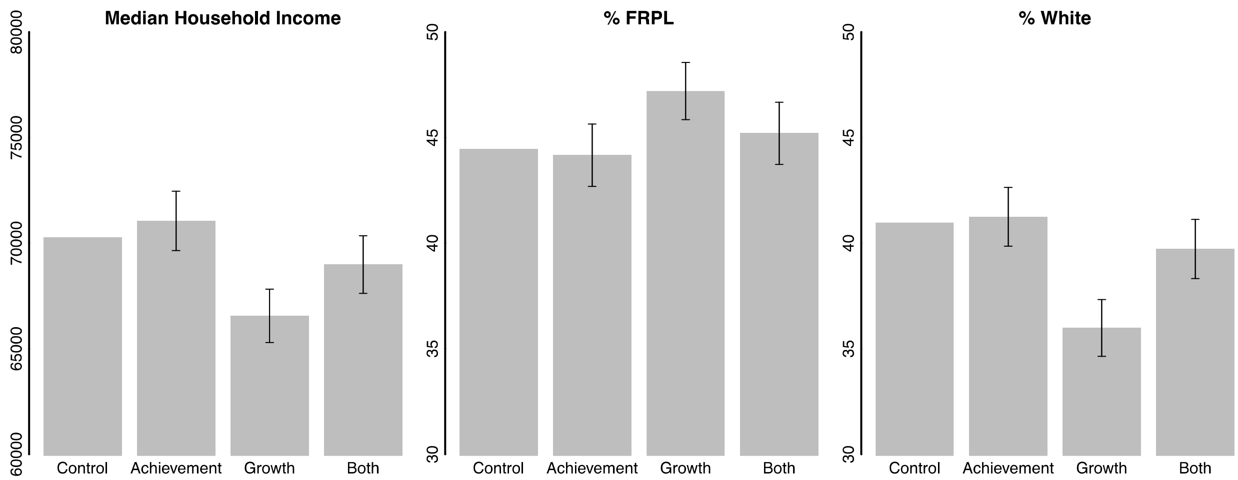
Notes. Y-axis displays percentage choosing each district; district names abbreviated to first three letters

Figure B2. Replication: Average Performance of Chosen Districts (Individuals = 2,500; Observations = 12,500)



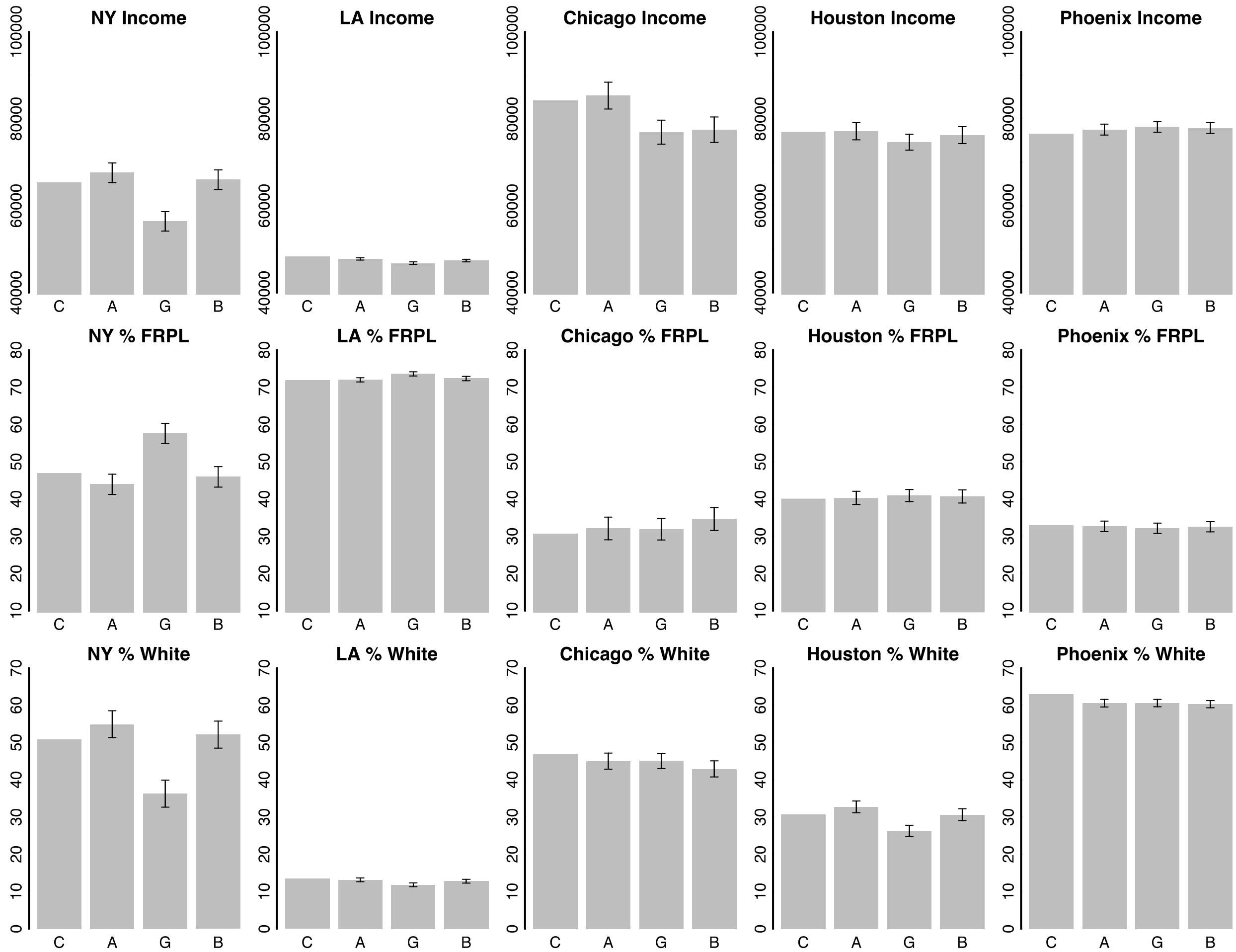
Notes. Error bars represent 95% confidence intervals (robust standard errors clustered at the individual level)

Figure B3. Replication: Average Demographics of Chosen Districts (Individuals = 2,500; Observations = 12,500)



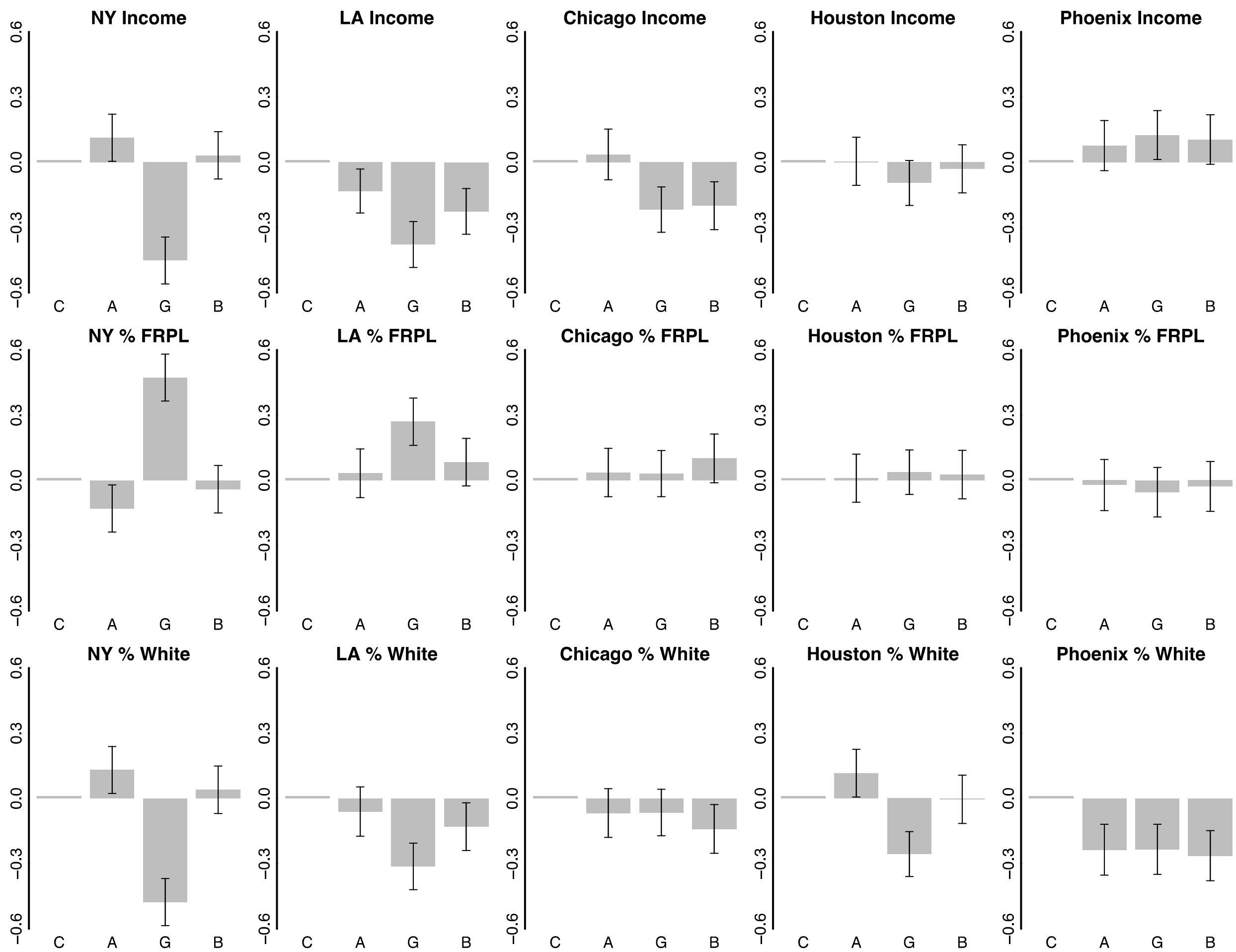
Notes. Error bars represent 95% confidence intervals (robust standard errors clustered at the individual level)

Figure B4. Replication: Results by Metropolitan Area (Individuals = 2,500)



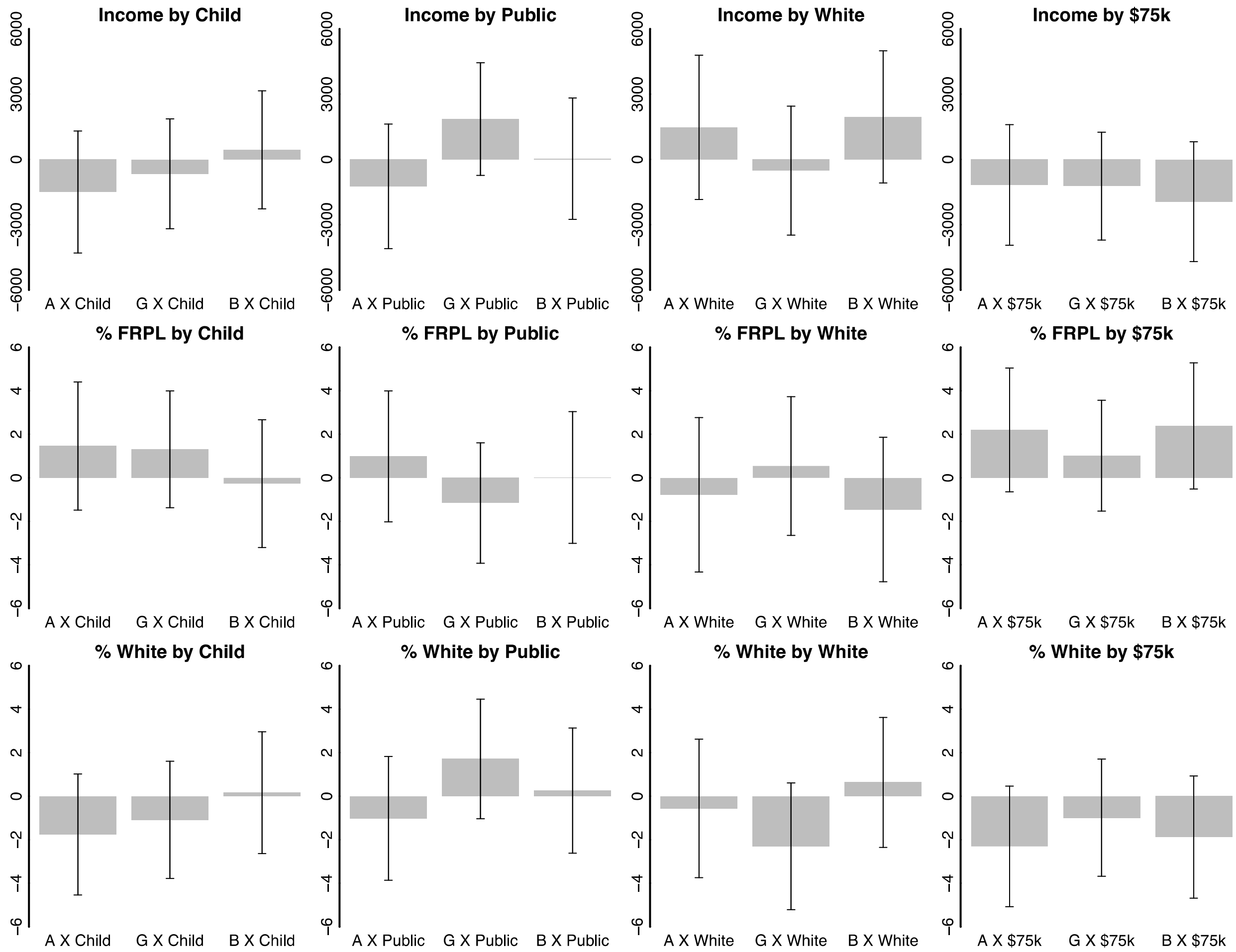
Notes. Error bars represent 95% confidence intervals (robust standard errors)

Figure B5. Replication: Results by Metropolitan Area with Standardized Outcomes (Individuals = 2,500)



Notes. Outcomes standardized within metro area and centered around control group mean; error bars represent 95% confidence intervals (robust standard errors)

Figure B6. Replication: Heterogeneous Treatment Effects (Individuals = 2,500; Observations = 12,500)



Notes. Values are OLS coefficients of treatment-by-demographic interaction terms; error bars represent 95% confidence intervals (robust standard errors clustered at the individual level)

Table B1. Replication: Balance and Representativeness (Individuals = 2,500)

%	Control (n = 626)	Achievement (n = 624)	Growth (n = 626)	Both (n = 624)	ANES 2016 (n = 4,271)
School-Age Child	50.08	47.67	49.92	53.45	33.73
School-Age Child in Public School	40.00	40.87	41.37	40.42	24.50
Female	48.24	49.28	52.08	47.67	52.90
White	73.08	75.16	77.85	72.44	71.68
Black	9.46	7.85	7.06	10.42	9.39
Asian	6.57	5.93	5.94	6.57	3.49
Native-American	1.44	2.08	1.77	2.24	0.64
Hispanic	7.85	7.37	6.10	5.77	10.62
Other Race	1.60	1.60	1.28	2.56	4.18
Less Than High School	0.32	0.48	0.80	0.64	6.66
High School	10.79	10.11	11.41	8.68	19.26
Some College	35.43	31.14	33.60	33.60	35.44
College	53.46	58.27	54.18	57.07	38.63
Family Income < $25,000	14.42	13.02	12.30	13.50	22.63
Family Income $25,000-$50,000	33.81	30.71	30.67	28.46*	21.82
Family Income $50,000-$75,000	22.60	24.92	25.40	26.21	18.16
Family Income $75,000-$100,000	14.10	16.72	17.41	16.72	13.10
Family Income > $100,000	15.06	14.63	14.22	15.11	24.28
Party ID (1-7)	3.65	3.55	3.57	3.59	3.86
Ideology (1-7)	3.72	3.55	3.67	3.64	4.18
Age (Years)	36.25	36.53	36.90	36.61	49.58
Missing Covariates	2.08	1.12	1.12	1.76	

Notes. Achievement, Growth, and Both compared to Control; * $p < 0.05$

Table B2. Replication: Average Performance of Chosen Districts (Individuals = 2,500; Observations = 12,500)

	Average Achievement Percentile		Average Growth Percentile	
Intercept	57.53*	59.16*	64.81*	64.77*
	(0.62)	(0.65)	(0.22)	(0.26)
Achievement	1.39	1.17	-0.09	-0.07
	(0.93)	(0.89)	(0.32)	(0.32)
Growth	-2.77*	-3.16*	4.56*	4.53*
	(0.83)	(0.79)	(0.42)	(0.42)
Both	0.13	-0.28	1.54*	1.46*
	(0.93)	(0.89)	(0.37)	(0.37)
Covariates		Yes		Yes

Notes. Values are OLS coefficients (robust standard errors clustered at the individual level); * $p < 0.05$

Table B3. Replication: Average Demographics of Chosen Districts (Individuals = 2,500; Observations = 12,500)

	Comparison to Control Group					
	Median Household Income		% FRPL		% White	
Intercept	70,261*	71,462*	44.46*	43.21*	40.97*	42.15*
	(492)	(524)	(0.52)	(0.54)	(0.51)	(0.54)
Achievement	788	640	-0.29	-0.13	0.27	0.09
	(717)	(686)	(0.75)	(0.72)	(0.71)	(0.68)
Growth	-3,707*	-3,974*	2.75*	3.06*	-4.99*	-5.31*
	(646)	(618)	(0.69)	(0.66)	(0.69)	(0.66)
Both	-1,277	-1,568*	0.74	1.06	-1.25	-1.52*
	(695)	(667)	(0.75)	(0.72)	(0.72)	(0.69)
Covariates		Yes		Yes		Yes
	Comparison to Achievement Group					
	Median Household Income		% FRPL		% White	
Intercept	71,049*	72,102*	44.17*	43.08*	41.24*	42.24*
	(522)	(543)	(0.55)	(0.57)	(0.50)	(0.52)
Control	-788	-640	0.29	0.13	-0.27	-0.09
	(717)	(686)	(0.75)	(0.72)	(0.71)	(0.68)
Growth	-4,495*	-4,614*	3.04*	3.19*	-5.26*	-5.40*
	(669)	(642)	(0.71)	(0.68)	(0.68)	(0.66)
Both	-2,065*	-2,208*	1.03	1.18	-1.52*	-1.61*
	(717)	(689)	(0.77)	(0.75)	(0.71)	(0.68)
Covariates		Yes		Yes		Yes

Notes. Values are OLS coefficients (robust standard errors clustered at the individual level); * $p < 0.05$

Table B4. Replication: Metropolitan Area Results, Comparison to Control Group (Individuals = 2,500)

	New York					
	Median Household Income		% FRPL		% White	
Intercept	65,357*	67,383*	46.87*	44.42*	50.89*	54.20*
	(813)	(886)	(0.99)	(1.07)	(1.31)	(1.43)
Achievement	2,221	1,886	-2.95*	-2.55	4.00*	3.46
	(1,151)	(1,118)	(1.39)	(1.35)	(1.84)	(1.79)
Growth	-8,959*	-9,419*	10.73*	11.26*	-14.65*	-15.38*
	(1,145)	(1,115)	(1.38)	(1.35)	(1.85)	(1.80)
Both	625	216	-0.94	-0.47	1.21	0.58
	(1,158)	(1,126)	(1.40)	(1.36)	(1.86)	(1.81)
	Los Angeles					
	Median Household Income		% FRPL		% White	
Intercept	48,280*	48,274*	71.74*	71.70*	13.44*	13.47*
	(96)	(109)	(0.21)	(0.22)	(0.18)	(0.19)
Achievement	-534*	-525*	0.21	0.20	-0.32	-0.31
	(147)	(142)	(0.30)	(0.29)	(0.26)	(0.26)
Growth	-1,523*	-1,565*	1.80*	1.83*	-1.65*	-1.69*
	(153)	(151)	(0.29)	(0.29)	(0.25)	(0.25)
Both	-910*	-951*	0.56	0.59*	-0.69*	-0.74*
	(153)	(151)	(0.29)	(0.28)	(0.26)	(0.25)
	Chicago					
	Median Household Income		% FRPL		% White	
Intercept	84,136*	86,423*	30.68*	28.69*	46.91*	48.13*
	(1,074)	(1,162)	(1.06)	(1.12)	(0.77)	(0.81)
Achievement	1,182	944	1.39	1.68	-1.92	-2.08
	(1,582)	(1,532)	(1.55)	(1.49)	(1.11)	(1.07)
Growth	-7,232*	-7,597*	1.18	1.75	-1.87	-2.26*
	(1,413)	(1,385)	(1.48)	(1.44)	(1.06)	(1.03)
Both	-6,640*	-7,113*	3.92*	4.44*	-4.02*	-4.36*
	(1,496)	(1,449)	(1.56)	(1.51)	(1.11)	(1.07)
	Houston					
	Median Household Income		% FRPL		% White	
Intercept	76,996*	77,574*	40.05*	39.53*	30.71*	31.42*
	(656)	(724)	(0.59)	(0.63)	(0.55)	(0.61)
Achievement	105	88	0.21	0.18	1.99*	1.91*
	(1,000)	(974)	(0.89)	(0.87)	(0.81)	(0.79)
Growth	-2,390*	-2,665*	0.84	1.06	-4.41*	-4.70*
	(935)	(912)	(0.83)	(0.80)	(0.76)	(0.75)
Both	-768	-1,032	0.60	0.84	-0.08	-0.25
	(998)	(975)	(0.90)	(0.88)	(0.82)	(0.80)
	Phoenix					
	Median Household Income		% FRPL		% White	
Intercept	76,537*	77,657*	32.94*	31.69*	62.93*	63.51*
	(467)	(483)	(0.54)	(0.55)	(0.39)	(0.41)
Achievement	965	806	-0.32	-0.13	-2.39*	-2.51*
	(635)	(621)	(0.72)	(0.71)	(0.51)	(0.50)
Growth	1,569*	1,378*	-0.81	-0.59	-2.37*	-2.52*
	(621)	(607)	(0.70)	(0.69)	(0.50)	(0.49)
Both	1,311*	1,042	-0.42	-0.11	-2.67*	-2.83*
	(626)	(615)	(0.71)	(0.69)	(0.50)	(0.49)

Notes. Values are OLS coefficients (robust standard errors); for each outcome, estimates in the second column control for all available covariates; * $p < 0.05$

Table B5. Replication: Metropolitan Area Results, Comparison to Achievement Group (Individuals = 2,500)

	Median Household Income		% FRPL		% White	
New York						
Intercept	67,579*	69,269*	43.92*	41.88*	54.89*	57.67*
	(815)	(879)	(0.97)	(1.05)	(1.29)	(1.40)
Control	-2,221	-1,886	2.95*	2.55	-4.00*	-3.46
	(1,151)	(1,118)	(1.39)	(1.35)	(1.84)	(1.79)
Growth	-11,180*	-11,305*	13.67*	13.81*	-18.66*	-18.84*
	(1,146)	(1,124)	(1.37)	(1.35)	(1.84)	(1.80)
Both	-1,596	-1,670	2.01	2.07	-2.79	-2.88
	(1,159)	(1,133)	(1.39)	(1.36)	(1.85)	(1.81)
Los Angeles						
Intercept	47,746*	47,749*	71.95*	71.90*	13.12*	13.16*
	(111)	(118)	(0.21)	(0.23)	(0.19)	(0.20)
Control	534*	525*	-0.21	-0.20	0.32	0.31
	(147)	(142)	(0.30)	(0.29)	(0.26)	(0.26)
Growth	-989*	-1,040*	1.59*	1.63*	-1.33*	-1.38*
	(163)	(159)	(0.29)	(0.29)	(0.26)	(0.25)
Both	-376*	-426*	0.34	0.39	-0.37	-0.43
	(163)	(159)	(0.30)	(0.29)	(0.26)	(0.26)
Chicago						
Intercept	85,318*	87,367*	32.07*	30.37*	44.98*	46.05*
	(1,161)	(1,227)	(1.13)	(1.20)	(0.80)	(0.85)
Control	-1,182	-944	-1.39	-1.68	1.92	2.08
	(1,582)	(1,532)	(1.55)	(1.49)	(1.11)	(1.07)
Growth	-8,415*	-8,541*	-0.21	0.08	0.06	-0.18
	(1,480)	(1,454)	(1.53)	(1.50)	(1.08)	(1.06)
Both	-7,823*	-8,058*	2.52	2.76	-2.09	-2.28*
	(1,559)	(1,516)	(1.61)	(1.57)	(1.13)	(1.10)
Houston						
Intercept	77,101*	77,663*	40.27*	39.71*	32.69*	33.33*
	(755)	(799)	(0.67)	(0.70)	(0.59)	(0.63)
Control	-105	-88	-0.21	-0.18	-1.99*	-1.91*
	(1,000)	(974)	(0.89)	(0.87)	(0.81)	(0.79)
Growth	-2,495*	-2,754*	0.63	0.88	-6.40*	-6.61*
	(1,007)	(978)	(0.89)	(0.86)	(0.80)	(0.78)
Both	-873	-1,120	0.39	0.66	-2.07*	-2.16*
	(1,066)	(1,042)	(0.96)	(0.94)	(0.85)	(0.83)
Phoenix						
Intercept	77,502*	78,463*	32.62*	31.56*	60.53*	61.00*
	(430)	(454)	(0.48)	(0.51)	(0.33)	(0.36)
Control	-965	-806	0.32	0.13	2.39*	2.51*
	(635)	(621)	(0.72)	(0.71)	(0.51)	(0.50)
Growth	604	572	-0.49	-0.46	0.02	-0.01
	(593)	(585)	(0.66)	(0.65)	(0.45)	(0.45)
Both	346	236	-0.10	0.02	-0.28	-0.32
	(599)	(591)	(0.66)	(0.65)	(0.45)	(0.45)

Notes. Values are OLS coefficients (robust standard errors); for each outcome, estimates in the second column control for all available covariates; * $p < 0.05$

Table B6. Replication: Heterogeneous Treatment Effects (Individuals = 2,500; Observations = 12,500)

	Median Household Income				% FRPL				% White			
Intercept	70,758*	71,363*	66,111*	68,443*	43.79*	43.23*	49.69*	46.37*	41.12*	41.86*	34.99*	39.22*
	(684)	(622)	(962)	(594)	(0.70)	(0.65)	(1.02)	(0.64)	(0.71)	(0.65)	(0.92)	(0.62)
Achieve	1,447	1,227	-478	898	-0.92	-0.62	0.49	-0.70	1.10	0.62	0.49	0.76
	(983)	(902)	(1,499)	(877)	(1.03)	(0.95)	(1.62)	(0.94)	(0.98)	(0.90)	(1.43)	(0.87)
Growth	-3,432*	-4,529*	-3,605*	-3,576*	2.16*	3.29*	2.70	2.71*	-4.48*	-5.75*	-3.61*	-4.93*
	(891)	(816)	(1,332)	(784)	(0.93)	(0.86)	(1.44)	(0.85)	(0.96)	(0.88)	(1.27)	(0.83)
Both	-1,484	-1,342	-2,649*	-871	0.85	0.79	1.75	0.23	-1.32	-1.39	-1.64	-0.87
	(968)	(871)	(1,341)	(836)	(1.04)	(0.93)	(1.48)	(0.92)	(1.01)	(0.90)	(1.30)	(0.87)
Child	-737				1.02				-0.09			
	(978)				(1.03)				(1.01)			
A×Child	-1,500				1.46				-1.76			
	(1,431)				(1.50)				(1.42)			
G×Child	-660				1.31				-1.09			
	(1,288)				(1.37)				(1.37)			
B×Child	440				-0.27				0.16			
	(1,384)				(1.50)				(1.43)			
Public		-2,352*				2.61*				-1.85		
		(1,000)				(1.06)				(1.04)		
A×Public		-1,240				0.99				-1.02		
		(1,460)				(1.54)				(1.45)		
G×Public		1,853				-1.16				1.71		
		(1,319)				(1.41)				(1.40)		
B×Public		32				0.02				0.26		
		(1,424)				(1.54)				(1.47)		
White			5,827*				-7.33*				8.31*	
			(1,109)				(1.17)				(1.08)	
A×White			1,476				-0.78				-0.56	
			(1,692)				(1.81)				(1.62)	
G×White			-518				0.54				-2.31	
			(1,513)				(1.63)				(1.49)	
B×White			1,953				-1.46				0.63	
			(1,550)				(1.70)				(1.52)	
$75k				6,778*				-7.19*				6.53*
				(960)				(0.97)				(0.99)
A×$75k				-1,175				2.20				-2.31
				(1,414)				(1.45)				(1.41)
G×$75k				-1,228				1.02				-0.99
				(1,265)				(1.30)				(1.37)
B×$75k				-1,942				2.38				-1.88
				(1,405)				(1.48)				(1.43)

Notes. Values are OLS coefficients (robust standard errors clustered at the individual level); * $p < 0.05$