



# Selective Enrollment Public Schools and District-Level Achievement Outcomes from 3rd to 8th Grade

Véronique Irwin

University of California, Berkeley

**Abstract:** Fierce local debates throughout the United States surround the equity of admitting students to public schools using academic criteria. Although research has evaluated the central assumption of these debates—that Selective Enrollment Public (SEP) schools enhance the welfare of students who attend them—none has addressed the district-level outcomes associated with these schools. This is important because the selectivity and scope of SEP schools produce tiered school systems (SEP districts). This district-level process, in turn, calls for an analysis of district-level achievement outcomes. To address this gap, I compile an original list of SEP schools using an innovative web scraping procedure. I combine these data with newly available district-level measures of third to eighth grade achievement from the Stanford Education Data Archive. Analyses follow a difference-in-differences design, using grade level as the longitudinal dimension. This approach facilitates a falsification test, using future treated districts, to reject spurious causation. I find evidence of overall slower growth in mean math achievement in SEP districts and for white, black, and Latinx racial/ethnic groups separately. SEP districts also see an increase in the white–Latinx math achievement gap. This work highlights the importance of considering SEP schools as part of a differentiated school system.

**Keywords:** Selective Enrollment Public schools; student differentiation; selectivity; scope; difference-in-differences

**Citation:** Irwin, Véronique. 2020. "Selective Enrollment Public Schools and District-Level Achievement Outcomes from 3rd to 8th Grade." *Sociological Science* 7: 100-127.

**Received:** September 24, 2019

**Accepted:** February 27, 2020

**Published:** April 2, 2020

**Editor(s):** Jesper Sørensen, Kim Weeden

**DOI:** 10.15195/v7.a5

**Copyright:** © 2020 The Author(s). This open-access article has been published under a Creative Commons Attribution License, which allows unrestricted use, distribution and reproduction, in any form, as long as the original author and source have been credited.

IN March 2019, news broke of wealthy parents going to extreme lengths to illegally secure their children admission to prestigious colleges, inciting national outrage about the entitlement of the ultrarich, and—through this shared outrage—demonstrating the importance all parents place on providing their children the best education. Although this highly publicized case focused on college admissions, similar pressures exist at the primary and secondary levels. In many medium and large school districts throughout the United States, this means securing a seat in a Selective Enrollment Public (SEP) school. In fact, just a few months prior to the “admissions scandal,” one of the nation’s best-known SEP schools, Boston Latin School, made headlines in the *Boston Globe*. The school was featured as part of a series called the “Valedictorians Project,” which documented the untapped potential of Boston’s public high school valedictorians—with the exception, that is, of graduates of the city’s SEP schools. The report described a two-tiered system, in which black and Latinx children overwhelmingly find themselves in schools where even valedictorians are told that their achievements “didn’t mean much compared to kids at [the SEP school,] Latin,” (Gay 2019). Concern about this inequity—in Boston and in other cities with SEP schools—manifests in debates about the racial justice of using academic admissions procedures to determine access to the district’s best schools (e.g., Ebbert 2016; Irizarry 2017; Rey 2017). The implicit assumption in these de-

bates is that access to SEP schools will enhance the well-being of underrepresented students who gain admissions.

Lost in this debate, however, is any consideration of the broader distributional consequences of having SEP schools in a district at all. Indeed, insofar as admissions processes shape the composition of SEP schools, they also shape the composition of traditional public schools throughout the district. SEP schools may therefore be an important element of district organization in determining the distribution of educational opportunities and outcomes districtwide. Yet we lack an understanding of what effect, if any, SEP schools have on achievement outcomes throughout the district. Thus, this article asks: Do traditional public districts with SEP schools (i.e., *SEP districts*) produce different (1) overall levels of achievement growth or (2) inequality of achievement than non-SEP districts? And, given the documented concerns around equity in access, (3) do these effects differ for students of different racial/ethnic backgrounds? By focusing on district-level outcomes, this article moves beyond a micro-level understanding of the effects of between-school student differentiation on the individual students and families who actively participate in school choice, in order to evaluate the meso-level impact of district organization on achievement outcomes.

To date, data limitations have prevented any such analysis. No existing data set on the universe of U.S. public schools gathers data on selective assignment procedures. Thus, I use an innovative web scraping technique to produce an original data set of SEP schools and the SEP districts that house them throughout the United States. These data are then combined with newly available national achievement data from the Stanford Education Data Archive (SEDA; Reardon et al. 2018). From SEDA, I draw on measures of district-level mean math achievement from third to eighth grade to evaluate achievement growth, as well as standard deviation of math achievement and racial achievement gaps in math as measures of inequality. Together, these data provide the unique ability to assess the role of SEP schools in shaping districts' overall ecology of educational opportunity.

Analyses employ a difference-in-differences (DiD) design, using grade level as the longitudinal dimension, in order to maximize the analytical leverage available from SEDA. Accordingly, the "treated" sample consists of 30 SEP districts that begin offering SEP schools between fourth and eighth grade, to allow for a third grade "pretreatment" observation. To evaluate whether the "dosage" of treatment matters and leverage the variation in these districts, I test for differences not only between SEP and non-SEP districts but between districts with different *prevalence* of SEP schools between fourth and eighth grade. Overall, I find that districtwide math achievement growth is slower in SEP districts than non-SEP districts. This result holds for white, black, and Latinx students, but not for Asian students. Additionally, I test the effect of SEP schools on district-level inequality of achievement using measures of standard deviation of achievement, as an indicator of total variance, as well as racial achievement gaps. Findings for each of these outcomes is mixed: I find marginal evidence of an increase in standard deviation only when using the dose treatment specifications, and I find evidence of an increase in race/ethnic achievement gaps only between white and Latinx students, not white and black students.

The remainder of the article is organized as follows. First, I introduce key elements of Sørensen's (1970) framework of student differentiation in order to situate SEP schools in the landscape of U.S. public education. I then briefly discuss what is known about SEP schools from the existing literature, which has focused on the effect of SEP schools on enrolled students. Next, I turn to research on the effects of between-school student differentiation on systemwide outcomes, in order to generate hypotheses for the subsequent analyses. This is followed by a discussion of my data collection procedure, secondary data from SEDA, and my difference-in-differences modeling approach. I then present results, before turning to a discussion, which highlights the questions that remain about mechanisms involved in this process.

## Academic Differentiation and the Place of Selective Enrollment Public Schools

Student differentiation refers to the practice of sorting students into subgroups for instructional purposes. The effect of student differentiation on achievement is the subject of longstanding and unsettled debate in the sociology of education. Sørensen (1970) argued that one reason for this is the inadequacy of theory to identify specific mechanisms through which differentiation might have its effect. He therefore theorized several dimensions of differentiation. In defining SEP schools, I am interested specifically in two dimensions of student differentiation: (1) selectivity of assignment procedures and (2) scope.<sup>1</sup>

SEP schools are unique in the public school sector in the United States in that they are both (1) selective in their assignment procedures and (2) high in scope. Selectivity refers to the degree of homogeneity the assignment procedure is intended to produce in terms of students' demonstrated performance or preparation. Scope refers to the extent to which differentiation confines students to educational spaces shared only by students of the same educational subgroup. In other words, a school is more selective when the assignment procedure is enforced by gatekeepers on the basis of academic criteria as opposed to student interest,<sup>2</sup> and scope is higher when differentiation occurs at the level of the school rather than the classroom. Figure 1 illustrates the position of several forms of student differentiation in the context of public school choice, along these two dimensions. Specifically, SEP schools are defined here as full schools, not within-school tracks or programs, that engage in reciprocal, that is, selective, choice on the basis of *academic* criteria.<sup>3</sup> SEP schools range in their level of selectivity, from requiring basic proficiency to gifted identification, and utilize admissions criteria that might include prior grades, writing samples, standardized testing, and/or specialized admissions exams. Although many SEP schools use a pure ranked admissions procedure, others enter all academically eligible students into a lottery.

Put simply, the "sector" of schools covered here is not a homogenous one, and it encompasses many kinds of learning environments.<sup>4</sup> For analytic purposes, however, I restrict the definition of SEP schools to schools that are funded and managed by a central, "traditional" public school district (i.e., districts that include

		Scope		
		Low	Medium	High
Selectivity of Assignment Procedure	Low	General Electives	Traditional Magnet Programs	Charters Traditional Magnet Schools
	High	Traditional Tracking	Selective Programs	SEP Schools

Figure 1: Defining SEP schools.

other traditional public schools with neighborhood-based catchment areas). This is essential to allow for the analysis of systemwide (here, district-level) achievement outcomes. Students from throughout the district compete for admissions, and the “brightest” students gain admission to the best schools, whereas the rest are relegated to neighborhood schools. This results in a clearly tiered system, bolstered by the guise of meritocracy, which might reasonably shape not only peer groups but also student exposure to expectations and resources throughout the district, key mechanisms theorized by Sørensen in the relationship between differentiation and student achievement. It is these whole districts, which I refer to as SEP *districts*, that constitute the educational systems of interest in the analyses below.

### Selective Schools: Effects on Enrolled Students and Remaining Questions

Although no research has specifically studied SEP districts as such, there is a tradition of research that investigates the impacts of academically selective schools on the students they serve. Below, I summarize the key findings of this literature in order to highlight the gaps that remain in our understanding of how these schools may shape achievement throughout the broader education systems of which they are a part.

One aim of academically selective schooling options, such as SEP schools, is unquestionably to enhance the potential of high-ability students. However, research has long complicated the notion that elite schools produce positive effects for their students in terms of academic self-concept (Marsh 1988; Marsh and Hau 2003),

attainment (Attewell 2001; Davis 1966; Espenshade, Hale, and Chung 2005), or achievement (Abdulkadiroğlu, Angris, and Pathak 2014; Allensworth et al. 2016; Dobbie and Fryer, Jr. 2011). By definition, only a small proportion of students at any school can be “top of the class.” Research shows that occupying lower positions in elite institutions—or being a “small frog in a large pond” (Davis 1966) can have negative consequences for high-achieving students. For example, (relatively) lower-performing students in extremely high-performing settings are less likely to take Advanced Placement (AP) classes in high school (Attewell 2001) and tend to apply to lower-prestige jobs after college than students of similar demonstrated ability in average institutions (Davis 1966). This, of course, is in addition to the fact that class rank itself can be of structural importance, as it is often considered during the admissions process at selective universities (e.g., Stevens 2007). As a result, the benefits of attending “star” schools reflect a winner-take-all phenomenon whereby benefits may only accrue to a very small number of students (Attewell 2001; Frank and Cook 1995).

More recently, scholars have conducted rigorous studies of the achievement effects of public “exam schools”<sup>5</sup> in Boston (Abdulkadiroğlu et al. 2014), New York City (Abdulkadiroğlu et al. 2014; Dobbie and Fryer, Jr. 2011), and Chicago (Allensworth et al. 2016) and grammar schools in the United Kingdom (Clark 2007). Using regression discontinuity designs, these studies find no important effect of SEP schools on high school achievement among marginally admitted students. It is important to consider, however, that marginally admitted SEP students may be the least likely to benefit from SEP schools, not only in SEP schools themselves but throughout SEP districts. First, the fact that marginal admissions winners do not benefit from SEP schools by no means suggests that the same is true of the top achievers in these schools.<sup>6</sup> Additionally, average-achieving students who get the opportunity to shine in traditional public schools because of the departure of SEP students may benefit from their relative star status. Importantly, such an effect would also make it less likely to observe a *relative* advantage for marginally admitted SEP students, even if their outcomes do improve. Together, these possibilities raise the question not only of SEP school effects on non-marginally admitted students, but of spillover effects at other schools throughout the district. The current article therefore contributes to the growing body of research on SEP schools by offering the first evaluation of the *districtwide* effects of SEP schools.

To situate this contribution, I turn next to research on the system-level effects of high scope (i.e., between-school) differentiation in the United States and selective differentiation abroad to generate hypotheses for achievement outcomes in SEP districts in the United States.

## Systemwide Effects of High-Scope Differentiation

No existing research, to my knowledge, has examined district-level effects of high-scope *selective* student differentiation in the United States. However, the proliferation of charter schools has motivated research on the effect of “competition” on traditional public school students. By law, charter schools are generally forbidden from using academic criteria in their assignment procedures,<sup>7</sup> but they

produce between-school (high-scope) differentiation on the basis of organizational, instructional, or disciplinary preferences. Research has evaluated what effect charter “competition” has on traditional public districts, but evidence is mixed, with studies finding positive (Booker et al. 2008; Jinnai 2014), negative (Ni 2009), or null (Zimmer and Buddin 2009) effects. Although these studies take the important step of testing system-level outcomes, they face geographic limitations because charter schools typically constitute their own “district” and do not have strict geographic attendance boundaries. This makes it difficult to define the differentiated *system* of which they are a part and is likely a contributing factor in producing such equivocal findings.

In contrast to the charter literature, which is challenged by a lack of clear school system boundaries, stratification researchers evaluate the effect of selective school system differentiation at the national level, thereby precluding this issue. Scholars in this field find that differentiated school systems are more likely to (re)produce inequality (Hanushek and Wößmann 2006; Pfeffer 2012; van de Werfhorst and Mijs 2010) while also producing some evidence of lower average achievement (Hanushek and Wößmann 2006; Pfeffer 2012). Furthermore, studies find that the more *rigidly* systems are differentiated, that is, the more difficult it is to switch between tracks once sorted (Pfeffer 2008), and the earlier differentiation happens (Horn 2009), the greater the inequality.

SEP districts can be considered relatively rigid in the American context because earning access to the top “track” in these districts is not a matter of changing classes within a single comprehensive school, but rather of gaining admissions to a SEP school in a specific grade. Although SEP districts are certainly not as rigidly differentiated as countries like Germany<sup>8</sup> (e.g., West and Nikolai 2013), for instance, they nevertheless present a contrast to what is otherwise considered a relatively open system throughout the United States. Based on this literature, then, we might expect SEP districts to demonstrate (1) higher inequality and (2) similar or perhaps somewhat lower levels of achievement than non-SEP districts.

## Data and Analysis

Until now, the ability to answer the questions posed in this article has been limited by two key data constraints: (1) the need for nationwide and nationally normed district-level achievement data and (2) the need to identify SEP districts and schools. The first requirement has recently been met by SEDA, which normalizes annual standardized test performance for students in grades 3 through 8 across districts throughout the United States. I personally address the second data requirement by generating an original list of SEP districts, which is the first of its kind to systematically include primary schools, making it compatible with the SEDA achievement data. With these data, I estimate DiD analyses of district-level average math achievement and inequality of achievement between third and eighth grade, in SEP versus non-SEP districts.



## *SEP Districts and Schools: Independent Variable and Sample Identification*

No formal list of Selective Enrollment Public schools exists. Information regarding public school admissions criteria is not collected by the Common Core of Data (CCD) or Civil Rights Data Collection (CRDC), and the one national survey that collects this information—the Schools and Staffing Survey (SASS)—has a small sample and insufficient detail for my purposes. Finn and Hockett (2012) begin the work of developing such a list but only include schools that grant high school diplomas. Thus, I collected original data and compiled my own list of SEP schools spanning the full kindergarten to 12th grade range. To do this, I used the programming language Python to conduct a systematic search for SEP schools by web scraping public school district websites for information regarding admissions procedures. In order to limit my search to a feasible number of districts, I cast my original net to include the 400 largest school districts in the country, which serve at least 20,000 students and can therefore a priori support high-scope (between-school) differentiation at both the primary and secondary levels. I then tested for saturation of my search in two ways: (1) first by expanding my list of districts to those with as few as 12,000 students but with at least one magnet school (as a signal of a favorable policy environment towards between-school differentiation) and (2) by drawing on lists of member institutions from the International Baccalaureate (IB) organization and the National Consortium of Specialized STEM Schools (NCSS).

Before beginning the formal web scrape, I manually searched through approximately 50 district websites to develop a sense of the language used to describe and provide information about SEP schools. The web scrape then proceeded in three steps. First, beginning from the home page of each district or school, I used the Python package BeautifulSoup to collect all of the links on the page and keep all those that were internal to the site. I then used this expanded list of web pages and repeated the procedure of extracting links, excluding duplicates. Second, I took this completed list of URLs and again used BeautifulSoup to search through them. This time, I searched the text on each page for a set of roughly 30 key words, such as “admission,” “application,” “gifted,” and “entrance exam,” developed during the initial manual search (see Appendix A in the online supplement for full list of keywords). Finally, I exported the list of URLs for the pages with at least one key word and searched through these web pages for evidence of SEP schools. During this step, I extracted more than 400 documents, including primarily school choice handbooks, statements of admissions procedures, and actual application materials, and qualitatively coded this information, as well as each relevant web page, for evidence of SEP schools.

As I combed through the district web pages that were flagged by my scraping procedure, I referenced Finn and Hockett’s (2012) list of exam schools,<sup>9</sup> as well as schools reported by the CRDC to enroll at least 75 percent of their students in gifted programming. This served to draw my attention to places where my scrape may have failed and triggered me to make adjustments to the program. After fine-tuning my search procedure using the original list of 400 districts serving at least 20,000 students, I identified 87 districts with at least one SEP school. Following this, the

saturation checks for medium size (12,000+) magnet districts and IB/NCSSS schools produced only an additional three SEP districts to my list.

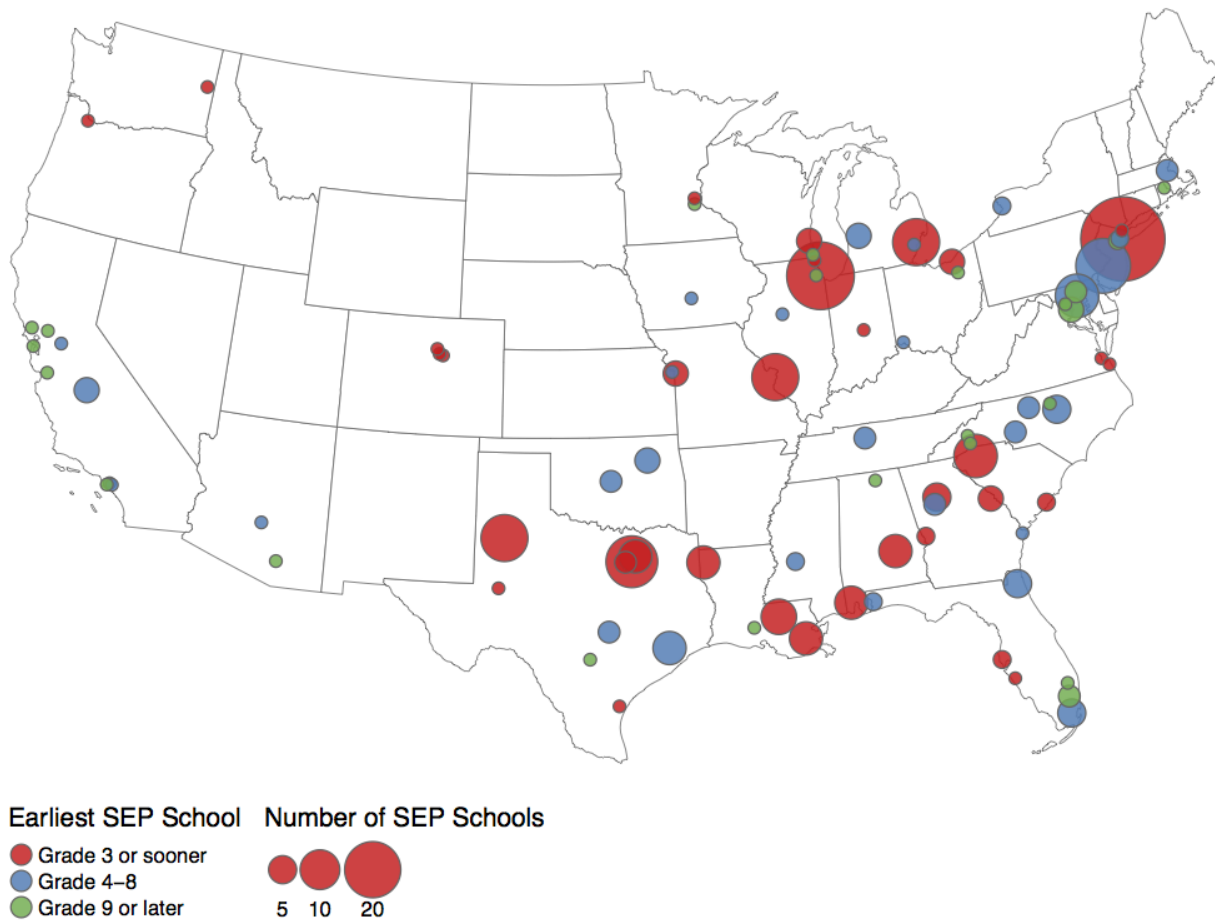
In total, this approach resulted in the identification of 370 SEP schools at all grade levels across 90 districts. In light of the existing research, which focuses on high schools, it is noteworthy that my search identified a near-equal split between primary and secondary SEP schools. SEP district locations and the grade level at which they are first offered are depicted in Figure 2. The average (median) SEP district offers two SEP schools, at least one of which begins by fifth grade.<sup>10</sup> There is a noticeable absence of SEP schools in the Great Plains and Mountain West, but this is not surprising. Because schools in these regions often serve very large areas, and districts themselves may contain very few schools, the feasibility of between-school differentiation is likely to be extremely limited. With this exception, SEP schools are geographically widespread throughout the country. This provides sufficient “treated” cases, as well as a reasonable pool of comparison districts, in the third to eighth grade range covered by SEDA.

The following analyses include 30 SEP districts that offer their first selective school options between fourth and eighth grade, depicted in blue in Figure 2. These districts begin the “treatment” of high-scope selective differentiation for the first time during the grade range for which SEDA, discussed below, provides achievement data.<sup>11</sup> Across these schools, the population of SEP school students is 32 percent white, 33 percent black, 16 percent Latinx, and 14 percent Asian. It is important to note, however, that although this appears to indicate an overrepresentation of black students relative to the national public school population, black students are actually underrepresented in SEP schools by 15 percent on average relative to the student population in their home district. Latinx students are underrepresented by 32 percent. By contrast, white students are overrepresented by 41 percent and Asian students by 167 percent. This reinforces local news reporting of underrepresentation of black and Latinx students in SEP schools and suggests the need to investigate the possibility of differential districtwide achievement effects of SEP schools for students of different racial/ethnic backgrounds.

### *SEDA: Dependent Variable and Controls*

The key dependent variables utilized here come from the district-level achievement measures in SEDA version 2.1 (Reardon et al. 2018). Every district in the United States is required to test students annually in grades 3 through 8 in math and English/Language Arts (ELA) and report this achievement in the national *EdFacts* data system; however, states establish their own tests and proficiency levels. Reardon and colleagues compile district-level data from these tests (for 2009 to 2015) and benchmark these against the National Assessment of Educational Progress (NAEP) by state, in order to create measures that are comparable for districts across states. This requires transforming achievement on state tests, which are often reported as coarse proficiency categories, onto the continuous NAEP scale to generate estimates of both mean and standard deviation of achievement. After benchmarking to NAEP, SEDA then scales these estimates for interpretability. The following analyses use the Cohort Standardized (CS) scale, which is calculated by mean-centering





**Figure 2:** SEP schools: location, level, and prevalence.

and dividing by the national grade-specific standard deviation from the original NAEP transformation<sup>12</sup> (Fahle et al. 2018; Reardon, Kalogrides, and Ho 2017). This scale is somewhat confusing to interpret for achievement *growth* because average achievement on this scale does not increase from one grade to the next. Where appropriate, I will reference parallel model results using grade-equivalent units to aid in interpretation. However, the CS scale is well suited to evaluate changes in inequality across grade levels because it is standardized by the amount of variation in each grade.<sup>13</sup> Moreover, by representing achievement in standard deviation units, results using this scale can be interpreted as effect sizes. For consistency, therefore, all model results are presented using the CS scale.

In addition to overall districtwide achievement, SEDA provides district achievement estimates by race. Because of the racially unequal access to SEP schools described in the previous section, it is important to test for potential differences in outcomes associated with SEP districts for students of different racial backgrounds. Accordingly, in addition to overall district mean achievement, I assess district mean math achievement for each race available from SEDA—white, Asian,

black, and Latinx—separately. I also examine inequality using SEDA’s measures of white–black and white–Latinx math achievement gaps, which are constructed by differencing race-specific mean achievement. Descriptive statistics for these variables (pretreatment) are presented in Table 1.

### *Analytic Approach: Difference-in-Differences for Math Achievement*

This article evaluates the effect of SEP schools on district-level mean and standard deviation of math achievement. I focus on math achievement for three reasons. First, math is a better indicator of the effect of school interventions because math is learned primarily in school, whereas language acquisition and reading skills are more heavily reinforced in the home (Bryk and Raudenbush 1988; Burkam et al. 2004; Murnane 1975; Parcel and Dufur 2001). This is reflected in the fact that summer losses tend to be greater in math than in reading, and this is increasingly true as students progress through school (see Cooper et al. 1996 for review). Given the potentially diffuse effects of SEP schools on achievement throughout a district, it is prudent to focus on an outcome that is more tightly coupled with schooling. Second, over the period from third to eighth grade, math instruction itself also becomes more differentiated, with students progressing at different speeds—and in different learning groups—through subject areas like General Math, Pre-Algebra, Algebra, or even Geometry by the end of eighth grade. Although differentiation in English/Language Arts certainly occurs, this tracking primarily distinguishes the *spacing* of classes—between “regular,” “honors,” or Gifted, for example—as opposed to distinct curricular content. Given the structure of these subjects during the focal grade levels for the present analyses, then, we might expect SEP schools to exert a more direct influence on the organization of math instruction.<sup>14</sup>

To evaluate the relationship between district-level selective differentiation and student achievement outcomes in math, I deploy a DiD design using grade level as the longitudinal dimension. DiD analyses compare the difference between treated and control units before treatment (the first difference) to their difference after treatment (the second difference) to estimate whether the change (if any) in the treated units is distinguishable from that of the control units and therefore attributable to treatment. Typically, treatments pertaining to policy changes or medical trials, for example, begin in a particular month or year, such that the periods “pre” and “post” treatment are defined by standard time. This is not possible for the present analyses because the SEDA data are too new to measure achievement before SEP schools were founded. Instead, I use grade as the longitudinal dimension. For the case of SEP schools, I argue that this approach actually offers some advantages over the traditional design. First, SEP schools are not the product of a targeted policy intervention that took effect in a single year, meaning that the particular temporal contexts of their founding could have different implications for their immediate impacts across districts. Moreover, whereas traditional longitudinal designs offer a single observation of the point of treatment, using grade level as the longitudinal dimension means that I observe the transition into treatment up to seven times for each SEP district in SEDA, which provides achievement data from 2009 to 2015. By averaging grade-level achievement across years, I am able to greatly reduce the

**Table 1:** Pretreatment (third grade) dependent variables, averaged 2009 to 2015.

	Range		Non-SEP	SEP	Total
<i>n</i>			755	30	785
Dependent variables, pretreatment (third grade)					
Mean achievement					
All	-1.310	1.127	-0.030	-0.209	-0.037
White ( <i>n</i> = 744)	-0.801	1.259	0.247	0.195	0.245
Black ( <i>n</i> = 626)	-1.189	0.286	-0.424	-0.502	-0.428
Latinx ( <i>n</i> = 739)	-1.107	0.564	-0.260	-0.293	-0.261
Asian ( <i>n</i> = 539)	-0.947	1.776	0.479	0.385	0.474
Inequality of achievement					
Standard deviation					
	0.777	1.324	0.967	0.979	0.967
White-black gap ( <i>n</i> = 612)	0.097	1.680	0.693	0.698	0.693
White-Latinx gap ( <i>n</i> = 708)	-0.108	1.434	0.506	0.488	0.506

Note: Non-SEP districts signifies all comparison districts, including future SEP districts.

risk that results are influenced by idiosyncratic shocks. Models, described below, therefore include one pretreatment (third grade) and one posttreatment (eighth grade) observation per district.<sup>15</sup>

Drawing on the grade-level variation in the provision of SEP schools, I define three treatment measures. First, I model treatment as a binary indicator of SEP district status, where 1 indicates at least one SEP school present in the district during the observation period and 0 indicates none. Districts are only considered treated if they offer their first SEP schools between fourth and eighth grade (*n* = 30), so that third grade functions as a true pretreatment observation for all districts in the analytic sample. Then, to test whether the “dosage” of treatment is important, I leverage the fact that different districts begin to offer SEP schools at different grade levels, and with greater or lesser prevalence. I operationalize the dosage of treatment in two ways: (1) the number of grades between fourth and eighth with at least one SEP school, ranging from zero to five, and (2) the percentage of fourth to eighth grade students throughout the district enrolled in a SEP school (with an interquartile range of 2 to 7 percent), based on enrollment data from the CRDC. In an ordinary least squares (OLS) regression framework, the effect of interest is operationalized as the interaction between the relevant treatment measure and grade level, as represented by  $\beta_3$  in the following equation.

$$Y_{ig} = \alpha + \beta X_{ig} + \beta_1 post_g + \beta_2 SEP\_Dist_i + \beta_3 (SEP\_Dist_i \times post_g) + \epsilon_{ig} \quad (1)$$

where  $Y_{ig}$  is the achievement outcome for district  $i$  in grade  $g$ ;  $\alpha$  is the intercept, which can be interpreted as the expected achievement outcome in third grade in comparison (i.e., non-SEP) districts;  $X_{ig}$  is a set of mean-centered district- (grade-) level covariates;  $post_g$  is an indicator for whether the observation occurs in eighth grade (equal to 1 in eighth grade, 0 in third);  $SEP\_Dist_i$  is one of three measures of the existence/prevalence of SEP schools in the given district, described above, and takes on the same value in third and eighth grade for each district; and  $\epsilon_{ig}$  is the

Huber–White heteroscedasticity-robust error term, clustered at the district level to account for nonindependence of errors at this level.<sup>16</sup>

Identification of a treatment effect in these models rests on the assumption that, in the absence of treatment, treated and control units would follow parallel trends. I address this assumption by carefully considering the definition of my treatment and comparison groups. For the comparison group, I restrict to “peer” districts by including only districts with some form of (nonselective) high-scope differentiation, which I define as districts with at least five schools serving students in each grade between grades 4 and 8, at least one of which is a traditional magnet *or* charter. There are 772 such districts with math achievement data from SEDA. These districts present an analytically useful comparison group because they are most similar in size and structure to SEP districts. In this way, I am comparing SEP districts with the districts they would be most likely to resemble in the absence of SEP schools. Although this necessarily limits the purview of the analysis and the conclusions that can be drawn from it, I argue that these bounds strategically sharpen the test of the mechanisms of interest—the scope and selectivity of student differentiation—in two important ways. First, it avoids comparing SEP districts with districts that logically cannot support high-scope (between-school) differentiation by virtue of their small size.<sup>17</sup> And second, by restricting the comparison sample to districts with intradistrict school choice, I focus my analysis specifically on the *selectivity* of high-scope differentiation.<sup>18</sup>

Still, selecting a strong comparison group does not fully address the assumption of parallel trends. To this end, as mentioned above, I omit districts that begin offering SEP schools before or during third grade. This reduces the likelihood of divergent trends emerging prior to the observation period—at least as a function of treatment—but does not permit an empirical test of this assumption because of the lack of achievement data prior to third grade. Instead, I draw on the variation in the grade level of treatment to generate a falsification test. Several SEP districts in my data offer SEP schools for the first time during or after ninth grade. This means that, during the fourth to eighth grade observation window for the current analysis, these districts are operating under a pretreatment district structure. Because these districts are likely to be the most similar to the analytic treatment sample, they offer useful leverage to evaluate the assumption that it is treatment itself, not the predisposition or preparation for treatment, that produces divergent trends between SEP districts and comparison districts. Thus, in addition to the 30 *treated* SEP districts, analyses below include 22 *future* SEP districts. These future SEP districts enter the model as an additional set of DiD terms, as expressed in Equation (2) below. In this setup, the falsification test asks whether future SEP districts differ from non-SEP comparison districts over the treatment period. In other words, confidence in the effect measured by  $\beta_3$  is enhanced if  $\beta_5$  is *not* statistically significant.

$$Y_{ig} = \alpha + \beta X_{ig} + \beta_1 post_g + \beta_2 SEP_{Dist_i} + \beta_3 (SEP_{Dist_i} \times post_g) + \beta_4 FutureSEP\_Dist_i + \beta_5 (FutureSEP_{Dist_i} \times post_g) + \epsilon_{ig} \quad (2)$$

Finally, each model includes a set of control variables ( $X_{ig}$ ), shown in Table 2. Control variables are the same in all models, with the exception of measures of

**Table 2:** Control variables with third grade (pretreatment) descriptive statistics.

<i>n</i>	Data source	Range		Non-SEP 755	SEP 30	Total 785
Total expenditures (log)	SEDA	8.742	10.488	9.307	9.391	9.310
Total teachers (log)	SEDA	4.137	10.303	6.750	7.949	6.796
% free or reduced lunch (0–1)	SEDA	0.038	0.962	0.490	0.656	0.496
% black students (0–1)	SEDA	0.000	0.978	0.160	0.366	0.168
% Latinx students (0–1)	SEDA	0.002	0.997	0.315	0.301	0.315
% of students in charters (0–1)	SEDA	0.000	0.500	0.180	0.252	0.183
50/10 income ratio	SEDA	1.731	9.925	3.820	4.165	3.834
Urban (0/1)	SEDA	0	1	0.405	0.800	0.420
Rural (0/1)	SEDA	0	1	0.113	0.000	0.108
% of school-age children who are foreign born (0–1)	EDGE	0.000	0.202	0.047	0.062	0.047
% of school-age children in female-headed households	EDGE	0.009	0.646	0.283	0.373	0.287
% of school-age children in owner-occupied homes	EDGE	0.129	0.905	0.596	0.499	0.593
Average value of owner-occupied homes (log)	EDGE	10.469	13.816	12.151	11.947	12.143
Total school-age population (log)	SEDA	6.913	13.559	9.870	11.089	9.917
% of white 5–17-year-olds in poverty	SEDA	0.000	0.579	0.124	0.163	0.125
White 50/10 income ratio	SEDA	1.098	17.680	3.649	4.039	3.664
% of black 5–17-year-olds in poverty	SEDA	0.000	1.000	0.300	0.359	0.302
Black 50/10 income ratio	SEDA	1.044	35.041	4.536	4.700	4.543
% of Latinx 5–17-year-olds in poverty	SEDA	0.000	1.000	0.284	0.356	0.286
Latinx 50/10 income ratio	SEDA	1.049	21.002	3.652	3.336	3.640
Number of charters in geographic district	SEDA	0.000	223.286	3.946	20.171	4.566

Note: In districts where schools offer “community coverage” for school lunches, meaning that 100 percent of students are provided free lunches, SEDA imputes the percentage that would be eligible based on traditional eligibility requirements.

poverty and socioeconomic inequality, for which race-specific measures are used in the models for race-specific district mean achievement. I include both percent black and percent Latinx in all models to account for the percent of students from racial/ethnic backgrounds that tend to be underserved by SEP schools ( $r = -0.347$  in analytic sample). Control variables from Census Education Demographic and Geographic Estimates (EDGE) are estimates for all school-aged children, regardless of public enrollment status. This is useful because families’ choice to opt into or out of public schools may be influenced by whether the district offers SEP schools. All control variables are mean-centered in the models, besides urbanicity, which is composed of a set of binary variables (the reference category is suburban). Accordingly, model intercepts can be interpreted as the third grade math achievement in a (suburban) district with average characteristics.<sup>19</sup>

Where there is an insufficient number of students (fewer than 20) with reported test data, SEDA does not provide race-specific achievement data. This results in a different number of observations in models for different racial groups. Furthermore, SEDA does not produce the race-specific poverty and socioeconomic segregation control measures for Asian students, so I rely on the full district measures for this group. Finally, of the 772 comparison districts with math achievement data, 17 are

missing covariate data and are dropped from the analytic sample. No SEP districts are missing covariate data, resulting in a final analytic sample of 785 districts.

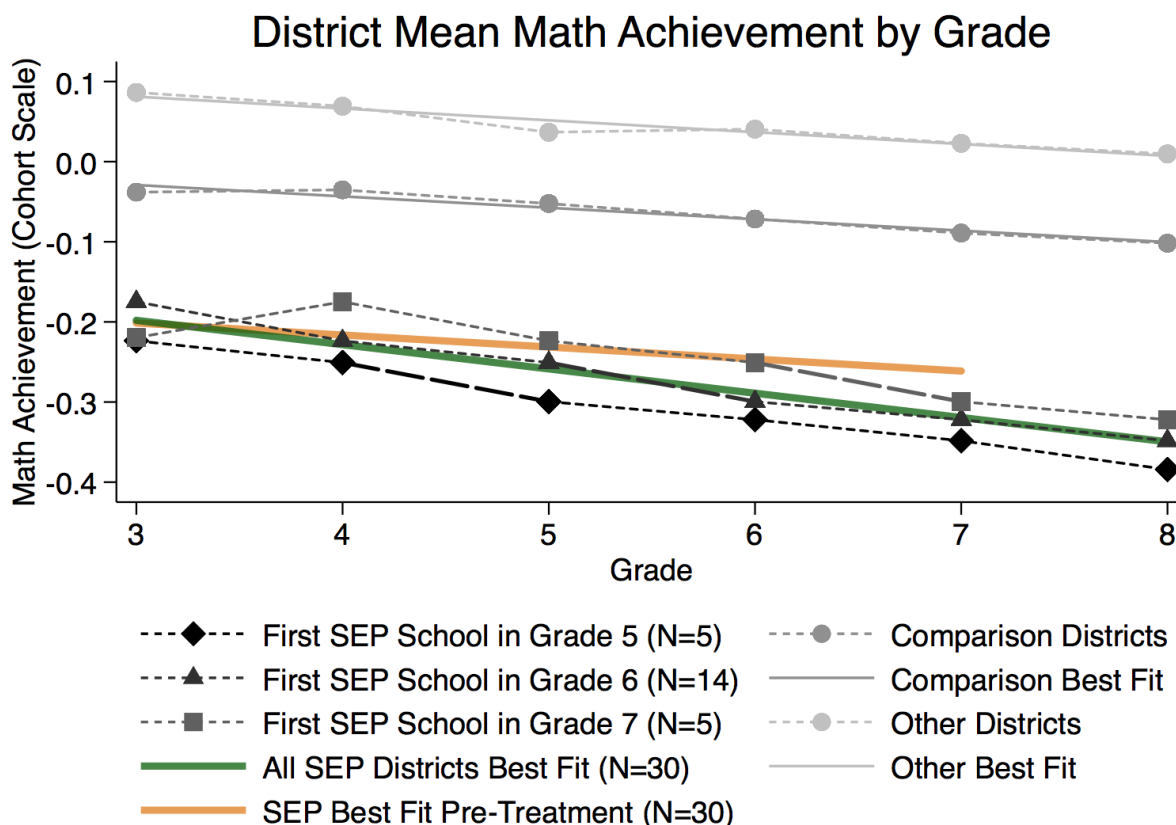
## Results

Before turning to model results, I present a graphical depiction of grade-level math achievement from third to eighth grade to establish the empirical phenomenon to be explained. Figure 3 shows the raw (i.e., uncontrolled) trend in average math achievement in SEP districts that begin offering SEP schools in fifth, sixth, or seventh grade,<sup>20</sup> as well as comparison districts. Overall, achievement in both SEP and comparison districts appears relatively low (below zero on a mean-centered scale) and appears to be trending downward (in relative terms). Note that this trend does not imply negative *absolute* growth from one grade level to the next.<sup>21</sup> An additional trend line is provided for smaller and/or undifferentiated districts (i.e., “other” districts that are omitted from the analytic sample). This additional trend line demonstrates that, although comparison districts in the analytic sample are relatively low achieving, their math achievement trajectory parallels that of districts throughout the United States. In other words, the apparent downward achievement trend is an artifact of the SEDA scaling<sup>22</sup> rather than a strange feature the analytic sample. The importance of Figure 3 lies in the relationship between these trend lines, not the specific slope of any individual line.

With this in mind, Figure 3 depicts relatively large drops in mean achievement at the grade level when SEP schools are first offered, with generally (but not always) flatter slopes at other grade levels. This could in part reflect an effect of school transitions, as research has shown that achievement tends to suffer initially when students enter a new school, before bouncing back (Grigg 2012). However, although nearly all students in the United States experience a school transition between third and eighth grade, the overall math achievement trend is steeper and more negative in SEP districts than in comparison districts on average across grade levels, as illustrated by the green line. Furthermore, the orange line plots the best fit trend only for district-grades in which SEP schools are not yet offered.<sup>23</sup> This line shows that before SEP districts begin to offer SEP schools, their achievement follows a very similar trend to comparison districts on average. This suggests that SEP schools may indeed be implicated in the lower average math achievement growth in these districts. By controlling for district characteristics and considering the prevalence of SEP schools, the DiD regression models, below, further test whether this divergence in district math achievement is associated with the provision of SEP schools.

In each table below, the coefficient of interest is called “SEP District DiD.” This is the interaction term between a district’s SEP treatment status (or dosage) and whether the observed achievement level is measured “pre” (third grade = 0) or “post” (eighth grade = 1) treatment. This is the additional difference between SEP and non-SEP districts that arises between third and eighth grade, beyond any difference that might have existed in third grade. Table 3 also includes coefficients for (1) the main effect of SEP district treatment status/dosage, which can be interpreted as the difference between treated and comparison districts in third grade, and (2) an indicator for eighth grade (third grade = 0, eighth grade = 1), which can be





**Figure 3:** Grade-level achievement in SEP districts versus other districts. *Note:* For pretreatment SEP district sample, the sample size at each grade level is as follows: third = 30, fourth = 27, fifth = 22, sixth = 8, seventh = 3.

interpreted as the expected change in non-SEP districts between third and eighth grade. These coefficients are omitted from subsequent tables for brevity. Results for the falsification test are represented as the coefficient for *future* SEP DiD.

### Slowing Growth in SEP Districts

First, I examine the association between high-scope selective differentiation and achievement growth at the district level. Model 1 shows the uncontrolled model where SEP district treatment is operationalized as a binary status. In this naïve model, the DiD coefficient is  $-0.084$ . Recall that, because achievement is measured in standard deviation units, this represents a  $-8.4$  percent effect size. This translates to the equivalent of nearly half a grade level less growth over the third to eighth grade period in SEP districts relative to non-SEP districts (see Appendix B in the online supplement for results using SEDA’s grade-equivalent scale). There is evidence that SEP districts tend to be different from non-SEP districts prior to treatment (the

**Table 3:** OLS DiD regression of mean math achievement on SEP district status.

	(1)	(2)	(3)	(4)	(5)
District controls		x	x	x	x
Treatment definition	Binary	Binary	Binary	Grade count	% enrolled
SEP district DiD	-0.084 <sup>†</sup> (0.027)	-0.085 <sup>†</sup> (0.028)	-0.085 <sup>†</sup> (0.028)	-0.022* (0.009)	-0.014 <sup>†</sup> (0.005)
SEP district (third grade difference)	-0.171 <sup>†</sup> (0.057)	0.027 (0.023)	0.025 (0.024)	0.008 (0.008)	0.006 (0.005)
Eighth grade (ref third grade)	-0.064 <sup>†</sup> (0.007)	-0.089 <sup>†</sup> (0.008)	-0.089 <sup>†</sup> (0.008)	-0.090 <sup>†</sup> (0.008)	-0.090 <sup>†</sup> (0.008)
Future SEP DiD			-0.014 0.024	-0.013 0.024	-0.013 0.024
Constant	-0.039 <sup>†</sup> (0.012)	-0.284 <sup>†</sup> (0.029)	-0.283 <sup>†</sup> (0.029)	-0.282 <sup>†</sup> (0.028)	-0.284 <sup>†</sup> (0.029)
<i>n</i>	1,604	1,570	1,570	1,570	1,570
<i>R</i> <sup>2</sup>	0.021	0.788	0.788	0.788	0.788

Notes: Huber-White heteroscedasticity-robust clustered standard errors. \* $p < 0.05$ , <sup>†</sup> $p < 0.01$ .

coefficient for the third grade difference is significant in model 1), but this difference is fully accounted for by the district-level control variables introduced in model 2.<sup>24</sup> Adding these controls has no impact on the estimated effect of district SEP status on achievement growth between third and eighth grade.

To further test that there are no unmeasured characteristics that increase *both* the propensity of districts to offer SEP schools *and* math achievement growth, model 3 includes the additional difference-in-differences test for districts that will begin selectively differentiating students into SEP schools at the high school level. The DiD for these *future* SEP districts in model 3 is not statistically significant, meaning that there is no expected difference in math achievement trends between these districts and other comparison districts. Moreover, there *is* a statistically significant difference between the achievement trends in future SEP districts and current treated SEP districts ( $p = 0.044$ , Wald test comparing DiD coefficients not shown). Together these findings lend confidence to the assumption of parallel trends prior to treatment and imply that the divergence in trends between SEP districts and comparison districts between third and eighth grade is not being spuriously driven by the *propensity* for districts to offer SEP schools, but rather by the logistics and practice of providing SEP schools.<sup>25</sup>

Models 4 and 5 further assess the relationship between SEP offerings and districtwide achievement by operationalizing treatment as the district-level “dosage,” or prevalence of SEP schools over the fourth to eighth grade treatment period. In model 4, SEP districts’ treatment dosage is defined as the number of grades in which at least one SEP school is offered, ranging from zero to five. In model 5, SEP school dosage refers to the percentage of fourth to eighth grade students enrolled in SEP schools in the district. In model 4, each additional grade for which districts offer at least one SEP option is associated with a 0.022 decrement in math achievement, or a -2.2 percent effect size. Among SEP districts, the average number of treated grades

is three (i.e., SEP schools begin in sixth grade), meaning that this translates to an average expected decrement of 0.066, or the equivalent of approximately one-third of a grade level of learning (see model B3 in the online supplement). Similarly, model 5 estimates a growth penalty of 0.014 standardized achievement units for each additional percentage of fourth to eighth grade students enrolled in a SEP school. With an average SEP school enrollment of 4.77 percent in SEP districts, this again equates to an average effect size of  $-6.6$  percent.

Overall, the estimated DiD coefficient for SEP districts is quite consistent across models in Table 3. However, note that although this increases confidence in the robustness of the result, it does not necessarily imply that SEP school dosage exerts a truly linear effect. Because of the relatively small number of treated districts (30), it is difficult to interrogate these relationships in detail. One possibility is that these patterns reflect a “tipping point” rather than a linear effect. For instance, estimated effects may be driven by districts with above-average prevalence of SEP schools (i.e., fourth to eighth grade district enrollment of 5 percent or higher). Additional regressions, not shown, suggest that this may be the case. Such an effect is more difficult to evaluate for the grade level of differentiation, because of clustering of treatment onset at sixth grade. Nevertheless, these findings present a high degree of confidence that a relationship exists between district provision of SEP schools and overall math achievement.

### *Race-Specific Achievement*

Next, I test whether this effect differs by race. Each panel in Table 4 parallels models 2 to 5 from Table 3 for a different racial/ethnic group. Where relevant, I will refer to these models collectively as race-specific (RS) models. Models 2-W, 2-B, and 2-L show that the provision of SEP schools is associated with lower levels of math achievement growth for white, black, and Latinx students. The substantive story is the same for each of these three racial/ethnic groups, although the relative effect size appears somewhat larger for Latinx students (11 percent, compared with about 9 percent for white and black students). Model 2-A shows no observed effect of SEP district status on the math achievement of Asian students.

Like the pooled race models, models 3-RS include a second DiD test for districts that offer SEP school beginning in high school. Again, the coefficients for the future SEP DiD are not significant. This suggests that it is unlikely that some predisposition towards selective differentiation is spuriously driving the findings in this analysis. One important caveat is that, although the DiD for future SEP districts is not itself significant for black students, it is also not statistically distinguishable from the treated SEP DiD. This may suggest that some part of the achievement penalty for black students in SEP districts begins to arise prior to treatment. I will return to this possibility in the discussion.

Models 4-RS and 5-RS test the operationalization of treatment as continuous indicators of the prevalence of SEP schools throughout the district. For white and Latinx students, the average implied treatment effect over the third to eighth grade period remains consistent and statistically significant. For black students, the estimated effect falls just below traditional levels of significance ( $p = 0.057$  and  $p =$

**Table 4:** DiD of mean math achievement on SEP status by race.

Treatment definition	Binary	Binary	Grade count	% enrolled
Panel 1: White ( $n = 1,488$ )	(2-W)	(3-W)	(4-W)	(5-W)
SEP district DiD	-0.091 <sup>†</sup> (0.034)	-0.091 <sup>†</sup> (0.034)	-0.029* (0.013)	-0.016* (0.007)
Future SEP DiD		-0.005 (0.030)	-0.005 (0.030)	-0.004 (0.030)
Panel 2: Asian ( $n = 1,078$ )	(2-A)	(3-A)	(4-A)	(5-A)
SEP district DiD	-0.033 (0.041)	-0.033 (0.042)	-0.004 (0.013)	0.000 (0.008)
Future SEP DiD		0.012 (0.064)	0.013 (0.064)	0.014 (0.064)
Panel 3: Black ( $n = 1,250$ )	(2-B)	(3-B)	(4-B)	(5-B)
SEP district DiD	-0.087* (0.036)	-0.088* (0.036)	-0.020 (0.011)	-0.010 (0.005)
Future SEP DiD		-0.040 (0.026)	-0.039 (0.026)	-0.038 (0.026)
Panel 4: Latinx ( $n = 1,478$ )	(2-L)	(3-L)	(4-L)	(5-L)
SEP district DiD	-0.110 <sup>†</sup> (0.026)	-0.110 <sup>†</sup> (0.026)	-0.030 <sup>†</sup> (0.009)	-0.017 <sup>†</sup> (0.005)
Future SEP DiD		-0.003 (0.026)	-0.002 (0.026)	-0.002 (0.026)

Notes: All models include district controls. Huber–White heteroscedasticity-robust clustered standard errors. Models designated -W, -A, -B, and -L indicate models specific to white, Asian, black, and Latinx students, respectively. \* $p < 0.05$ , <sup>†</sup> $p < 0.01$ .

0.064 for models 4-B and 5-B, respectively). Overall, however, both the magnitude and significance of these effects is very similar across model specifications for all racial/ethnic groups. In fact, in supplemental models (see Appendix B in the online supplement for replication of key coefficients from Tables 3 and 4), the effect for black students remains statistically significant ( $p < 0.01$ ) for both dose treatment specifications. It is especially striking that these effects are not only consistent across model specifications but are also quite consistent for students of different racial/ethnic backgrounds.

Importantly, just because effects are similar between students of different racial/ethnic backgrounds on average *across* districts does not mean they are necessarily the same *within* districts. This is a separate question, which requires evaluation of district-level racial achievement gaps. I turn to this next, along with the question of overall inequality of achievement.

### *Mixed Evidence of Increasing Inequality*

I evaluate the association between SEP district status and inequality of achievement at the district level, using standard deviation and racial achievement gaps as measures of inequality, for a total of three inequality outcomes: white–black achievement gaps, white–Latinx achievement gaps, and districtwide standard deviation

of achievement. As a reminder, the scale used for these analyses is standardized by the national grade-level standard deviation and therefore accounts for the total amount of variation at each grade. Again, Table 5 includes models using both the binary status and dose treatment measures of district provision of SEP schools.

Overall, findings regarding the association between SEP schools and district-level inequality of achievement are quite mixed. There is no evidence of an effect of SEP schools on standard deviation of achievement using the binary treatment specification in model 6-SD. However, there is marginal evidence of an increase in inequality when using the dose specifications. Model 7-SD predicts an average effect size of 2.4 percent ( $0.008 \times 3$ ), whereas model 8-SD predicts an average effect size of 1.7 percent ( $0.003 \times 4.77$ ). These effects are relatively small compared with the effects on average achievement and, again, achieve only marginal significance ( $p = 0.068$  and  $p = 0.072$ , respectively). Nevertheless, it is interesting to note the difference from Table 3, where greater significance was achieved using the binary than the dose treatment specifications. This could suggest that the appropriate functional form (and perhaps the relevant mechanisms) are different for standard deviation of achievement than for average achievement.

Additionally, there is mixed evidence of worsening racial/ethnic gaps. There is evidence that the achievement gap between white and Latinx students increases between third and eighth grade in SEP districts relative to comparison districts. This effect is moderate in magnitude, on the order of a 4 percent effect size on average based on results from both model 6-WLG and 7-WLG. The average effect appears somewhat smaller, at about 2.9 percent ( $0.006 \times 4.77$ ) in model 8-WLG. For white and black students, by contrast, there is no evidence of a change in the achievement gap in SEP districts relative to comparison districts between third and eighth grade. Not only is the DiD nonsignificant, but the point estimate is actually negative (but essentially zero) in the dose treatment models. However, the future SEP DiD for black students is marginally significant ( $p = 0.079$  to  $0.087$  across white–black achievement gap models) and positive. Although not too much should be made of these coefficients themselves, this does suggest the need for caution in interpreting results causally for black students and to consider the potential for effects of “preparing” for differentiation, particularly for these students.

## Discussion

This study is motivated by the fact that, to the extent that selective assignment procedures shape the educational experiences of students in SEP schools, they also shape the educational experiences of the students shut out of them. Rather than isolated institutions then, it is important to consider SEP schools as part of a selectively differentiated school system—a SEP district. Accordingly, the intent of this article is to evaluate district-level achievement outcomes associated with high-scope (i.e., between-school) selective differentiation, rather than the effect of SEP schools on enrolled (or excluded) students separately. This question is newly answerable by combining the district-level data available from SEDA with the original list of SEP schools and SEP districts compiled here. Difference-in-

**Table 5:** DiD of district-level achievement inequality on SEP status.

Treatment definition	Binary	Grade count	% enrolled
Panel 1: Districtwide standard deviation ( $n = 1,570$ )	(6-SD)	(7-SD)	(8-SD)
SEP district DiD	0.022 (0.016)	0.008 (0.004)	0.004 (0.002)
Future SEP DiD	0.026 (0.020)	0.026 (0.020)	0.026 (0.020)
Panel 2: White–black gap ( $n = 1,222$ )	(6-WBG)	(7-WBG)	(8-WBG)
SEP district DiD	0.024 (0.034)	−0.001 (0.013)	−0.002 (0.006)
Future SEP DiD	0.058 (0.033)	0.057 (0.033)	0.057 (0.033)
Panel 3: White–Latinx gap ( $n = 1,416$ )	(6-WLG)	(7-WLG)	(8-WLG)
SEP district DiD	0.041 <sup>†</sup> (0.014)	0.013 <sup>†</sup> (0.005)	0.006* (0.003)
Future SEP DiD	0.003 (0.030)	0.003 (0.030)	0.002 (0.030)

Notes: All models include district controls. Huber–White heteroscedasticity-robust clustered standard errors. Models designated -SD indicate standard deviation; those designated -WBG and -WLG indicate white–black gap and white–Latinx gap, respectively. \* $p < 0.05$ , <sup>†</sup> $p < 0.01$ .

differences analyses produce findings that are noteworthy not only in their novelty, but in their (dis)alignment with existing research on similar topics.

In terms of average achievement, existing research on the impact of SEP schools generally suggests neutral or slightly negative effects on achievement levels, whether that research focuses on enrolled students (Abdulkadiroğlu et al. 2014; Dobbie and Fryer, Jr. 2011) or on the entire differentiated school system (Hanushek and Wößmann 2006; Horn 2009). The present study estimates a slowing of math achievement growth by roughly 6.6 to 8.5 percent of a standard deviation in SEP districts relative to comparison districts, overall. This is a substantively large effect, equivalent to nearly half a grade level’s worth of learning over five years. This effect is relatively consistent for white, black, and Latinx students, on the order of a −9 to −11 percent effect size using the binary treatment specification. The exception to this rule is for Asian students, for whom there is no discernable relationship between SEP schools and district-level achievement growth. Although these results are somewhat in keeping with existing research, the consistency and magnitude of the finding that SEP districts produce significantly *less* achievement growth between third and eighth grade than non-SEP districts is nevertheless striking.

Next, the relatively weak evidence of a relationship between high-scope selective differentiation and inequality at the district level in the United States presents a contrast to the dominant findings from the nation-level social stratification literature, discussed above. Although there is some suggestion of a marginal increase in the districtwide standard deviation of achievement as the prevalence of SEP schools increases throughout the district, the only consistent evidence of an effect on inequality is for the achievement gap between white and Latinx students. In interpreting this finding, it is worth noting both the advantages and disadvantages



of the present analyses relative to existing work. On the one hand, unlike the stratification literature, whose units of analysis are whole countries with distinct educational regimes, the present analysis has the advantage of observing SEP and non-SEP districts within same states. This may allow the current study to better account for other policies considered important to student achievement, such as testing standards and the inclusivity of (public) postsecondary education (Ayalon and Gamoran 2000; Kerckhoff 2001). On the other hand, the present study may be limited by the fact that SEDA's estimates of standard deviation rely heavily on coarse proficiency categories, which might obscure achievement differences at the extreme. Estimates of standard deviation may therefore be conservative. In other words, although these estimates present important improvements over existing work, they also face important limitations that warrant investigation in future research.

Another point of consideration around the analysis of inequality is that the measures used here, particularly standard deviation of achievement, may obscure differences in the meaning of inequality in districts that have different average levels of achievement. For instance, if mean achievement is high, then a relatively high standard deviation may not in itself be "bad," as it may simply imply exceptionally high achievement among students at the top of the distribution. Similarly, the same white–minority achievement gap would signify different educational experiences in a district where average-achieving white students perform at the national average compared with one where they perform a full grade level above the mean. Using SEDA's standardized scaling helps to address issues of comparability by adjusting each dependent variable using national grade-level standard deviations, but it does not address differences in districts' average levels of achievement. Yet, even with these differences and limitations, there is strong evidence of growing inequality between white and Latinx students in SEP districts relative to non-SEP districts. Given that mean achievement is simultaneously *decreasing* in SEP districts, it is unlikely that this inequality could be characterized as a "good" thing.

These findings were produced with a unique DiD design that uses grade level as the longitudinal dimension. This was a useful approach to extract analytic leverage from the data but also raised potential concerns based on the inability to evaluate pretreatment trends. This is important because a divergence between SEP and non-SEP districts prior to treatment might suggest unobserved differences in districts that have the *propensity* to offer SEP schools, or perhaps that there is some effect of "preparing" for selective differentiation. If either were the case, we would reasonably expect similar processes to characterize the trend in *future* SEP districts—those that have not begun offering SEP schools in the grade range of the current analysis but that offer SEP schools in high school. Analyses test this possibility, and the estimated DiD coefficients for future SEP districts do *not* reach conventional levels of significance ( $p < 0.05$ ) in any of the models presented above. Overall, then, analyses in this article present strong evidence of a slowing of math achievement growth and a widening white–Latinx achievement gap in SEP districts relative to similar non-SEP districts.

Importantly, however, this DiD test for future SEP districts also raises some questions about the effect of SEP districts on black students. Although future SEP

status itself is not significantly predictive of math achievement growth relative to comparison districts, there is overlap in the estimated growth trend for black students in treated SEP and future SEP districts. Moreover, for the white–black achievement gap, future treatment is marginally predictive ( $p < 0.1$ ) of an increase in the achievement gap, whereas there is no evidence of a treatment effect of SEP districts from third to eighth grade. Rather than indicating something about the *type* of district that offers SEP schools (i.e., a spurious effect, which would imply an effect in both current *and* future treated districts), this may imply an effect of *preparation* for selective differentiation. For instance, it could be that middle schools that typically send very few of their students to SEP high schools are less proactive about preparing their students for the admissions process (or, vice versa—SEP “feeder” schools may be especially proactive), which could have spillovers for achievement on standardized tests. If there are systematic inequalities in who these middle schools serve, whereby SEP feeder schools disproportionately underserve black students, then this could result in relatively lower achievement outcomes for black students prior to differentiation. In order to establish whether these findings may indeed be evidence of negative externalities in the *preparation* for differentiation, particularly for black students, future research must consider the mechanisms through which the provision of SEP schools generate districtwide effects.

Although the present study represents an important starting point in our understanding of the potential spillover effects of SEP schools, made possible by two new sources of data, the analyses here are still limited by their inability to identify the mechanisms at work. SEP schools are likely to shape students’ educational experiences by shaping the peers they share classrooms with, the curricula taught in those classrooms, and the expectations teachers hold for them, both for students who attend SEP schools and those who attend non-SEP schools in SEP districts. Importantly, only a very small percentage of students in SEP districts typically attend SEP schools. For neighborhood schools that serve primarily poor and minority populations—for whom structural inequalities limit both information and preparation for SEP school admissions—this may mean only a handful of students are lost to SEP schools. In other words, a *direct* peer effect from the composition of students’ classrooms or schools may be unlikely. However, teacher expectations for their students may be broadly shaped by the district’s arbitrary definition of “bright” students. Districts may also make strategic decisions about how to allocate resources based on this process of student differentiation, perhaps concentrating college preparatory resources in a smaller number of schools. Such mechanisms are difficult to measure with the national data sets, like the Civil Rights Data Collection or Common Core of Data, particularly at the primary level.<sup>26</sup> Future research should consider these and other mechanisms in order to better understand the relationship between SEP schools and district-level outcomes.

Finally, it is important to acknowledge that SEP schools currently operate under a system of unequal opportunity. Under these less-than-ideal conditions, SEP districts may not be producing their best potential outcomes for achievement growth and inequality. Of course, achieving equal access is no doubt a sticky issue (e.g., Orfield and Ayscue 2018), and the likelihood of positive results under such a

regime is made somewhat questionable by the evidence from this study that SEP schools currently hinder average districtwide achievement growth even among white students, who tend to be *over*represented in SEP schools. However, this article far from closes the debate on the merits of SEP schools. Rather, it highlights that the effects of these schools should not be sought solely among choice participants. This, in turn, should motivate continued research on the conditions under which SEP schools are founded, the place they occupy in district-level processes, and the mechanisms through which they exert their effects.

## Notes

- 1 Other important dimensions, according to Sørensen, are inclusiveness and whether differentiation is horizontal or vertical. Inclusiveness refers to the number of opportunities available at a given level of schooling (i.e., K-8 education) and is of little analytic value here, because schooling is compulsory in the United States until students are 16. Vertical differentiation reduces variation in students' presumed capacity to learn course content, whereas horizontal differentiation reduces the variation in the content delivered in a particular course setting. These latter concepts overlap significantly with the selectivity of assignment procedures.
- 2 Sørensen's framework defines electivity—the extent to which subgroup placement is a choice for students—as a distinct characteristic of assignment procedure. In theory, an elective process could produce a homogenous (i.e., select) group. In practice, however, when assignment is not purely the prerogative of students, this either results in or is the result of selective criteria.
- 3 As distinct from schools for the performing arts that use auditions or portfolios.
- 4 These schools are referred to by many different names by their operating districts, including selective enrollment schools, specialized schools, exam schools, criteria schools, and so on. Notably, some of these schools are designated as magnets, whereas others are not. Magnets are public schools that offer specialized programs intended to attract students from throughout a district. Canonically, these programs are designed to foster integration and do not use admissions criteria (Fuller and Elmore 1996; Wells 1993). Variation in magnet status among SEP schools may therefore reflect different levels of emphasis on student integration associated with these schools.
- 5 Exam schools are a specific subtype of SEP schools that utilize exam scores as the sole (or primary) criterion in their admissions process.
- 6 Some research has attempted to estimate effects for SEP students further from the margin and reaches similar conclusions about the weak effect of SEP schools on enrolled students (Angrist and Rokkanen 2015).
- 7 However, academic self-selection into charter schools is possible.
- 8 That is, traditional public school students in SEP districts are not structurally precluded from selective four-year colleges, as is the case for vocational school students in Germany.
- 9 My list differs significantly from Finn and Hockett's because (1) they only include schools that offer high school diplomas, whereas mine includes primary schools, and (2) they include several schools run at the state level or by district consortia, which I omit to facilitate district-level analysis. However, this filter is only relevant at the high school level and does not factor into the sample for the current article.

- 10 The mean number of SEP schools offered in SEP districts is 4.2. This differs significantly from the median because of the large number of SEP schools offered in very large districts, namely, Philadelphia, Dallas, Chicago, and New York City. However, all but Philadelphia begin differentiating students before third grade and are therefore excluded from these analyses.
- 11 Note that this grade-level restriction pertains to the analytic sample, only, and did not factor into the design of the web scrape procedure to identify SEP schools.
- 12 At each grade level, every cohort is mean-centered and standardized relative to a single cohort (at the appropriate grade level) to allow for analysis of absolute differences in achievement levels over time. The reference cohort is the cohort that participated in NAEP in fourth grade in 2009.
- 13 SEDTA provides an additional scale, called the Grade Cohort Scale (GCS), which is standardized relative to the change in NAEP scores from one grade to the next, each within a given cohort. This scale produces grade-equivalent units, which increase from an average of 3 for students achieving at grade level in third grade to 8 in eighth grade. On this scale, the average standard deviation of math achievement increases with each grade level, making it somewhat difficult to compare inequality between pre- and posttreatment. For questions of achievement growth, however, the GCS is quite useful because scores increase by approximately one unit with each additional grade, providing an intuitive measurement for achievement growth. Models for achievement growth are replicated in Appendix C of the online supplement using the GCS scale.
- 14 Because of this tracking structure, it also is easier to identify advanced coursework in math than in reading using available national data. In fact, although the CRDC provides relatively detailed information on mathematics courses, beginning with Algebra in seventh grade, it contains no information about English courses. This information on student participation in advanced coursework is tested as a mediator in Appendix B of the online supplement.
- 15 Averaging across years also avoids the issue of states with missing data in eighth grade for multiple years because of the use of end-of-course (as opposed to end-of-year) tests, which could therefore not be benchmarked to state-level NAEP scores. For instance, California has only one year of complete eighth grade data and Texas only two (out of seven), so running regressions with one observation for each year could importantly underweight these large states.
- 16 Additionally, to account for potential state-level differences in proficiency standards or other policies that may influence student test outcomes, models are estimated with state fixed effects by including a state indicator as a factor variable in the models, representing 42 states and the District of Columbia; I set the largest state—California—to be the base category.
- 17 In addition to being substantively meaningless to attribute achievement differences in a district with one school to the lack of SEP schools, these small districts, if included, would also receive disproportionate weight by virtue of the district-level structure of the analysis.
- 18 Additional analyses, not shown, define the comparison group solely by district size and produce substantively similar results.
- 19 Such an “average” district may be more or less typical for students of different racial backgrounds.
- 20 Three districts each begin offering SEP schools in fourth and eighth grade. Individual trend lines for these sets of districts are omitted to avoid overcrowding the graph, but they are included in the average lines of best fit.

- 21 Rather, it may imply relatively faster growth in NAEP-participating cohort, against which other cohorts in the data were standardized, producing an artifact of relatively slower growth overall.
- 22 This downward trend from one grade to the next persist for all cohorts except the reference cohort of 2009 fourth grade NAEP participants. This suggests there is some artifact of benchmarking other cohorts against this cohort that produces this trend. This artifact is unique to this cohort-mean-standardized scale and does not affect model results. Results for models using the grade-equivalent scale (which increases one unit per grade level on average) are presented in Appendix B of the online supplement.
- 23 That is, there is a diminishing sample from one grade to the next, with 30 in third grade and only three districts that are still “pretreatment” by seventh grade.
- 24 Note that in Appendix B of the online supplement, which presents model results using the GCS (grade-equivalent) scale, the difference between SEP and comparison districts in third grade remains significant in models 2 to 5. It is possible that the collinearity between grade level itself and the GCS scale may artificially produce this relationship. However, this does not affect the DiD coefficients, which are consistent both in significance and relative effect size in models using the GCS and CS scale.
- 25 By examining districts that are likely to be most similar to the treated districts in this analysis, this test also provides reassurance that results are not likely to be significantly affected by the omission of any districts that the scrape may have failed to identify.
- 26 Appendix C of the online supplement provides results for an attempt at such a mediation analysis, using available data from the CRDC and SEDA—(1) the percentage of students in the district attending a school where Algebra is offered in seventh grade, as an indicator of the (un)evenness of curricular rigor throughout the districts, and (2) student exposure to low-income peers, based on free and reduced price lunch status, as an indicator of the concentration of exposure to more/less privileged peers. Information on advanced math courses is the only available data on curricular differentiation from the CRDC prior to high school. One might reasonably expect a school’s advanced math offerings to be directly shaped by its selective standing within a district, and for that in turn to shape math achievement, making it a reasonable mediator. Likewise, SEP schools might shape the distribution of more/less privileged students throughout the district, in turn shaping things like PTA resources or teacher expectations at different schools, which might then shape achievement. Student socioeconomic segregation is therefore perhaps more an indicator of other mechanisms than a true mediator in itself.

## References

- Abdulkadiroğlu, Atila, Joshua D. Angrist, and Parag A. Pathak. 2014. “The Elite Illusion: Achievement Effects at Boston and New York Exam Schools.” *Econometrica* 82(1):137–96. <https://doi.org/10.3982/ECTA10266>.
- Allensworth, Elaine M., Paul T. Moore, Lauren Sartain, and Marisa de la Torre. 2017. “The Educational Benefits of Attending Higher Performing Schools: Evidence from Chicago High Schools.” *Educational Evaluation and Policy Analysis* 39(2):175–97. <https://doi.org/10.3102/0162373716672039>.
- Angrist, Joshua D., and Miikka Rokkanen. 2015. “Wanna Get Away? Regression Discontinuity Estimation of Exam School Effects Away from the Cutoff.” *Journal of the American Statistical Association* 110(512):1331–44. <https://doi.org/10.1080/01621459.2015.1012259>.

- Attewell, Paul. 2001. "The Winner-Take-All High School: Organizational Adaptations to Educational Stratification." *Sociology of Education* 74(4):267–95. <https://doi.org/10.2307/2673136>.
- Ayalon, Hanna, and Adam Gamoran. 2000. "Stratification in Academic Secondary Programs and Educational Inequality: Comparisons of Israel and the United States." *Comparative Education Review* 44(1):54–80. <https://doi.org/10.1086/447591>.
- Booker, Kevin, Scott M. Gilpatric, Timothy Gronberg, and Dennis Jansen. 2008. "The Effect of Charter Schools on Traditional Public School Students in Texas: Are Children Who Stay Behind Left Behind?" *Journal of Urban Economics* 64(1):123–45. <https://doi.org/10.1016/j.jue.2007.10.003>.
- Bryk, Anthony S., and Stephan W. Raudenbush. 1988. "Toward a More Appropriate Conceptualization of Research on School Effects: A Three-Level Hierarchical Linear Model." *American Journal of Education* 97(1):65–108. <https://doi.org/10.1086/443913>.
- Burkam, David T., Douglas D. Ready, Valerie E. Lee, and Laura F. LoGerfo. 2004. "Social-Class Differences in Summer Learning between Kindergarten and First Grade: Model Specification and Estimation." *Sociology of Education* 77(1):1–31. <https://doi.org/10.1177/003804070407700101>.
- Clark, Damon. 2010. "Selective Schools and Academic Achievement." *B.E. Journal of Economic Analysis & Policy* 10(1):1917. <https://doi.org/10.2202/1935-1682.1917>.
- Cooper, Harris, Barbara Nye, Kelly Charlton, James Lindsay, and Scott Greathouse. 1996. "The Effects of Summer Vacation on Achievement Test Scores: A Narrative and Meta-Analytic Review." *Review of Educational Research* 66(3):227–68. <https://doi.org/10.3102/00346543066003227>.
- Davis, James A. 1966. "The Campus as a Frog Pond: An Application of the Theory of Relative Deprivation to Career Decisions of College Men." *American Journal of Sociology* 72(1):17–31. <https://doi.org/10.1086/224257>.
- Dobbie, Will, and Roland G. Fryer, Jr. 2011. *Exam High Schools and Academic Achievement: Evidence from New York City*. Working Paper No. 17286, National Bureau of Economic Research. <https://doi.org/10.3386/w17286>.
- Ebbert, Stephanie. 2016. "Testing Program to Expand, Increase Latin School Access." *Boston Globe*, April 27.
- Espenshade, Thomas J., Lauren E. Hale, and Chang Y. Chung. 2005. "The Frog Pond Revisited: High School Academic Context, Class Rank, and Elite College Admission." *Sociology of Education* 78(4):269–93. <https://doi.org/10.1177/003804070507800401>.
- Fahle, Erin, Benjamin Shear, Demetra Kalogrides, Sean Reardon, Richard DiSalvo, and Andrew Ho. 2018. *Stanford Education Data Archive: Technical Documentation*. Version 2.1.
- Finn, Chester E., Jr., and Jessica A. Hockett. 2012. *Exam Schools: Inside America's Most Selective Public High Schools*. Princeton, NJ: Princeton University Press. <https://doi.org/10.1515/9781400844579>.
- Frank, Robert H., and Philip J. Cook. 1995. *The Winner-Take-All Society: Why the Few at the Top Get So Much More Than the Rest of Us*. New York, NY: Random House.
- Fuller, Bruce, and Richard Elmore, eds. 1996. *Who Chooses? Who Loses? Culture, Institutions and the Unequal Effects of School Choice*. New York, NY: Teachers College Press.
- Gay, Malcolm. 2019. "The Valedictorians Project: Part 3. The Race to Get into Boston's Exam Schools." *The Boston Globe*, January 17.
- Grigg, Jeffrey. 2012. "School Enrollment Changes and Student Achievement Growth: A Case Study in Educational Disruption and Continuity." *Sociology of Education* 85(4):388–404. <https://doi.org/10.1177/0038040712441374>.



- Hanushek, Eric A., and Ludger Wößmann. 2006. "Does Educational Tracking Affect Performance and Inequality? Differences-in-Differences Evidence across Countries." *Economic Journal* 116(510):C63–76. <https://doi.org/10.1111/j.1468-0297.2006.01076.x>.
- Horn, Daniel. 2009. "Age of Selection Counts: A Cross-Country Analysis of Educational Institutions." *Educational Research and Evaluation* 15(4):343–66. <https://doi.org/10.1080/13803610903087011>.
- Irizarry, Yasmiyn. 2017. "Commentary: Healing Won't Begin until LASA Represents All of Austin ISD." *Statesman*, February 24.
- Jinnai, Yusuke. 2014. "Direct and Indirect Impact of Charter Schools' Entry on Traditional Public Schools: New Evidence from North Carolina." *Economics Letters* 124(3):452–6. <https://doi.org/10.1016/j.econlet.2014.07.016>.
- Kerckhoff, Alan C. 2001. "Education and Social Stratification Processes in Comparative Perspective." *Sociology of Education* 74(Extra):3–18. <https://doi.org/10.2307/2673250>.
- Marsh, Herbert W. 1991. "Failure of High-Ability High Schools to Deliver Academic Benefits Commensurate with Their Students' Ability Levels." *American Educational Research Journal* 28(2):445–80. <https://doi.org/10.3102/00028312028002445>.
- Marsh, Herbert W., and Kit Tai Hau. 2003. "Big-Fish-Little-Pond Effect on Academic Self-Concept: A Cross-Cultural (26-Country) Test of the Negative Effects of Academically Selective Schools." *American Psychologist* 58(5):364–76. <https://doi.org/10.1037/0003-066X.58.5.364>.
- Murnane, Richard. 1975. *The Impact of School Resources on Learning of Inner City Children*. Cambridge, MA: Ballinger.
- Ni, Yongmei. 2009. "The Impact of Charter Schools on the Efficiency of Traditional Public Schools: Evidence from Michigan." *Economics of Education Review* 28(5):571–84. <https://doi.org/10.1016/j.econedurev.2009.01.003>.
- Orfield, Gary, and Jennifer B. Ayscue, eds. 2018. *Discrimination in Elite Public Schools: Investigating Buffalo*. New York, NY: Teachers College Press.
- Parcel, Toby L., and Mikaela J. Dufur. 2001. "Capital at Home and at School: Effects on Student Achievement." *Social Forces* 79(3):881–911. <https://doi.org/10.1353/sof.2001.0021>.
- Pfeffer, Fabian T. 2008. "Persistent Inequality in Educational Attainment and Its Institutional Context." *European Sociological Review* 24(5):543–65. <https://doi.org/10.1093/esr/jcn026>.
- Pfeffer, Fabian T. 2012. *Equality and Quality in Education*. Research Report 12-774, Population Studies Center.
- Reardon, Sean F., Andrew D. Ho, Benjamin R. Shear, Erin M. Fahle, Demetra Kalogrides, and Richard DiSalvo. 2018. *Stanford Education Data Archive*. Version 2.1.
- Reardon, Sean F., Demetra Kalogrides, and Andrew Ho. 2017. "Linking U.S. School District Test Score Distributions to a Common Scale." Working Paper No. 16-09, Stanford Center for Education Policy Analysis.
- Rey, Jay. 2017. "African-American Enrollment Dips at City Honors, Despite Efforts to Correct Disparity." *Buffalo News*, September 26.
- Sørensen, Aage Böttger. 1970. "Organizational Differentiation of Students and Educational Opportunity." *Sociology of Education* 43(4):355–76. <https://doi.org/10.2307/2111838>.
- Stevens, Mitchell L. 2007. "Creating a Class: College Admissions and the Education of Elites." Cambridge, MA: Harvard University Press.

- van de Werfhorst, Herman G., and Jonathan J. B. Mijs. 2010. "Achievement Inequality and the Institutional Structure of Educational Systems: A Comparative Perspective." *Annual Review of Sociology* 36:407–28. <https://doi.org/10.1146/annurev.soc.012809.102538>.
- Wells, Amy Stuart. 1993. *Time To Choose: America at the Crossroads of School Choice Policy*. East Rutherford, NJ: Putnam Publishing Group.
- West, Anne, and Rita Nikolai. 2013. "Welfare Regimes and Education Regimes: Equality of Opportunity and Expenditure in the EU (and US)." *Journal of Social Policy* 42(3):469–93. <https://doi.org/10.1017/S0047279412001043>.
- Zimmer, Ron, and Richard Buddin. 2009. "Is Charter School Competition in California Improving the Performance of Traditional Public Schools?" *Public Administration Review* 69(5):831–45. <https://doi.org/10.1111/j.1540-6210.2009.02033.x>.

**Acknowledgments:** I thank Samuel Lucas, Daniel Schneider, Anthony Jack, and Zawadi Rucks-Ahadiana for their thoughtful comments. This work also benefited from excellent feedback at the Improving Education and Reducing Inequality Conference at the Russell Sage Foundation and was supported by funding from Russell Sage Foundation Grant #83-18-14.

**Véronique Irwin:** Department of Sociology, University of California, Berkeley. E-mail: [virwin@berkeley.edu](mailto:virwin@berkeley.edu).