Doing More with Less: School Management and Education Production*

Sam Stemper†

September 19, 2022

Abstract

The school superintendent is the highest-ranking executive in U.S. school districts, responsible for managing personnel decisions and overseeing regular school operations. To estimate the causal effect of superintendents on district performance, I collect data on the tenures of over 18,000 school superintendents covering over half of American public school children using a model of test score value-added that allows a superintendent's effect to emerge over the course of their tenure within a district. Superintendent transitions between districts are leveraged to validate these estimates and to identify common practices of effective superintendents. I show that superintendents have large effects on school district performance, accounting for one-fourth of the observed differences in learning rates across districts. Top management matters most in districts where managerial flexibility is ex-ante largest: smaller districts and districts with weaker teachers unions. Effective superintendents do not change levels of district spending or staffing but instead make changes in school operations, increasing teacher turnover and reducing teacher absences. Finally, I find evidence that the link between value-added and salary for superintendents is strongest in districts with higher levels of local interdistrict competition.

*First version: October 21, 2021. This version: September 19, 2022. I am grateful to Christopher Avery, Andrew Bacher-Hicks, Raj Chetty, Will Dobbie, Joshua Goodman, and Lawrence Katz and seminar participants at Harvard Public/Labor Lunches, Harvard Kennedy School Applied Micro Seminar, and the Brandeis University Faculty Seminar for helpful comments and suggestions. All remaining errors are mine.

†Harvard University. Email: stemper@g.harvard.edu

Top management is a crucial input in production. While survey evidence indicates that better-managed firms are more productive and better-managed schools have higher levels of student achievement, causal evidence supporting these relationships is scarce.[1] The limited causal evidence we have from field experiments does not explain how and why good management arises in practice, and to what degree observed differences in productivity are a consequence of differences in management quality. This paper examines the role of school superintendents, the top management at the school district level in U.S. K-12 education.

There are a number of reasons why school management matters. First, effective schools have long-run benefits. Across historical and contemporary settings, measures of school quality bear strong relationships with measures of well-being later in life.[2] Second, government spending on public education is massive: in the US, governments spend over $700 billion on K-12 education, accounting for over 3.5 percent of GDP. To the degree that management is technological in nature (Bloom et al. (2016)), better management can enhance the productivity of these inputs. Finally, several features of the public education system suggest that good management may not arise automatically. Relative to institutions in the private sector, public schools may face less competitive pressure (Brueckner and Neumark (2014) and Diamond (2017)) or have special interest groups (Hoxby (1996)) that constrain their capacity for productivity-increasing reforms.

However, estimating the causal effects of management is difficult. To date, most evidence is cross-sectional, and shows that better managed schools have, on average, higher levels of student achievement. This approach gives rise to two issues.

First, there is the issue of measurement. Indices of school management are typically based on surveys in which enumerators ask school principals (or other school managers) questions about their management practices and score their responses based on their alignment with pre-defined "best practices."[3] While these "best practices" are typically selected to correspond with

[1] Bloom et al. (2013) and Fryer Jr (2014) are notable exceptions with respect to firms and schools, respectively.

[2] See, e.g. Card and Krueger (1992).

[3] For example, the World Management Survey's Educational Questionnaire asks the following questions about "Standardization of Instructional Practices":

a. How structured or standardised are the instructional planning processes across the school?

b. What tools and resources are provided to teachers (e.g. standards-based lesson plans and textbooks) to ensure consistent level of quality in delivery across classrooms?

c. What are the expectations for the use of these resources and techniques?

d. How does the school leader monitor and ensure consistency in quality across classrooms?

well-established best practices in other sectors, they require researchers to choose which types of educational behaviors constitute “good" or “bad" management ex-ante.

Even if measurement is perfect, another issue arises; management is not assigned randomly. Specifically, the processes that determine management quality are likely related to other processes that determine measures of student achievement. For example, suppose that high- and low-income parents sort into different sets of schools. If (a) high-income parents demand a higher level of management quality from local schools and (b) children born into high-income families score higher on standardized tests, researchers may observe a positive relationship between management and test scores, even in the absence of any causal connection. These challenges of endogenous management quality and reverse causality caution against a causal interpretation of the estimates in cross-sectional comparisons.

In this paper, I attempt to overcome these hurdles by inverting this typical approach: rather than quantifying what we believe to be good management practices and testing whether these practices correlate with student outcomes, I identify the effect of good *managers* based on outputs alone (namely, standardized test scores). Doing so allows me to quantify the effect of management on outcomes without taking a stance on which practices constitute good or bad management. From there, I can examine how the practices employed by “good" managers differ from those employed by “bad" managers, and what circumstances give rise to effective and ineffective management.

I do so by assembling the largest-known database of district superintendent rosters from U.S. school districts. This database is the first of its kind and represents over 18,000 unique superintendents, over 7,000 school districts, and covers over 55 percent of American K-12 students as of 2014. I link this data to multiple sources of data on student demographics, district characteristics, and student achievement, and quantify the effect that superintendents have on test scores by estimating superintendent valued-added models. These models, which have been used broadly for estimating the effects of teachers, aim to identify the causal effect of superintendents.

Estimating value-added in a managerial context is complicated by the fact that superinten-

Top-scoring schools are those in which “[s]chool has implemented a clearly defined instructional planning process designed to align instructional strategies and materials with learning expectations and incorporate flexibility to meet student needs; these are followed up on through comprehensive monitoring or oversight." “2009 Education Survey Instrument," World Management Survey.

dents likely affect student achievement beyond their tenure in a district. For example, an effective superintendent may hire skilled teachers or introduce a high-quality curriculum; these and other changes to district operations may persist after a superintendent leaves. Similarly, when a new superintendent enters a district, their effect on district performance may be not realized immediately. Instead, it may take time for district outcomes to fully reflect the quality of their management.

I account for this possibility directly in my estimation, modeling a superintendent's effect as a function of their value-added as well as their tenure within the district. Similar to models that incorporate dynamics in similar settings (Coelli and Green (2012) and Kinsler (2012), for example), I estimate these dynamic parameters directly, allowing me to test how the distribution of superintendent effects varies as superintendents spend more time in a district.

My findings are as follows.

First, variation in superintendent value-added is economically meaningful. My estimates suggest that a permanent one standard deviation improvement in superintendent value-added increases test scores by 0.011 standard deviations. Moving from the 10th to the 90th percentile of superintendent value-added generates larger gains: 0.024 test score standard deviations. These estimates are roughly one-tenth the size of typical estimates of teacher value-added. Still, for many districts, the potential test score gains from improvements in top management exceed gains from much more ambitious and costly improvements in teacher value-added, because most districts hire hundreds of teachers but only one superintendent.

However, identifying effective superintendents is complicated by the managerial dynamics described above; I find that these dynamics are important in estimation. My estimates suggest that superintendents realize less than 50 percent of their "full" potential effect on the district in the first two years. I argue that these two effects—the smaller effects of management (relative to teachers) as well as managerial dynamics—may in part explain the limited evidence for superintendent effects in earlier research (Chingos et al. (2014), for example).

I validate my value-added estimates by studying how test scores change in response to the arrival of new superintendent. A superintendent's value-added estimate from one district predicts changes in test scores after she arrives in a new district, but does not predict changes prior to superintendent entry, bolstering the case for a causal interpretation of superintendent effects. Heterogeneity analyses suggest that the effects of management are largest in small districts, districts with lower baseline levels of achievement, and districts in states with weaker teachers unions.

Second, high value-added superintendents affect district practices, but have limited effects on district inputs. Using the same design as above, which compares outcomes in a district before and after the arrival of a high- or low-value added superintendent, I test for changes across a broad range of district inputs and school practices. I find no evidence that superintendent value-added is associated with increases in district resources in aggregate, or shifts in district spending in favor of one input or another (instruction spending or capital spending, for example). Instead, I find evidence that effective superintendents change district practices. In the years immediately following the arrival of a higher value-added superintendent, rates of teacher absences fall, the share of new teachers increases, and schools are more likely to offer a gifted and talented program.

Finally, I examine whether superintendents are rewarded for performance in the form of longer tenures or higher pay. I find some evidence in favor of this hypothesis generally, and strong evidence of hetereogeneity based on the degree of local interdistrict competition. Consistent with a model of Tiebout competition and empirical results in Hoxby (2000), I find that superintendents with higher value-added have longer tenures and higher pay in districts with higher levels of local competition.

Broadly, my work relates to two distinct sets of literature.

First, my work builds on the body of evidence with respect to management in education. Much of this work is primarily descriptive, aiming to introduce new measures of management, based on detailed survey data (as in Bloom, Lemos, Sadun and Van Reenen (2015)) or administrative data (as in Leaver et al. (2019)). These measures exhibit large positive relationships with data on student achievement even after controlling for a number of relevant student and school characteristics. Still, these papers are primarily descriptive and do not purport to estimate causal effects of management. Another set of papers uses randomized experiments to assess the effectiveness of management interventions on student outcomes. These papers include Fryer Jr (2014) and Fryer et al. (2017). These papers demonstrate that large scale, ambitious management interventions have positive, causal effects on student outcomes. My work bridges the gap between non-experimental and experimental work by estimating the causal impact of management outside of an experimental setting.

Further, all four of the aforementioned papers focus on schools and principals as the source of managerial effects. In Bloom, Lemos, Sadun and Van Reenen (2015) and Leaver et al. (2019), management indices are constructed based on school-level survey responses. In Fryer Jr (2014) and Fryer et al. (2017), experimental randomization takes place at the school level. I compliment this work by investigating the role of district management, who manage more students in total but typically have a smaller role in overseeing everyday school operations.

Second, my empirical work relates to the use of value-added modeling, both within and outside of the educational realm. Interpreting fixed effects as measures of "management quality" has its historical roots in Mundlak (1961). More recently, Bertrand and Schoar (2003) estimate similar models to characterize CEO management styles. Value-added models have been used most broadly in education, frequently in the context of quantifying teacher effects. A relatively smaller set of papers aims to estimate the effects of various managerial inputs, including principals (Coelli and Green (2012), Bartanen and Husain (2022), Bartanen et al. (2022), Loeb and Grissom (2013), and others) and superintendents (Chingos et al. (2014), Lavy and Boiko (2017)). To my knowledge, my value-added estimates involve a largest set of students of any such study, allowing me to estimate managerial effects and quantify the practices of effective managers with more precision than prior work.

This paper proceeds as follows. In the section below, I briefly describe my setting: Superintendents in American school districts. Next, I introduce and describe my main data sources. In Section 3, I characterize the market for school superintendents. Sections 4 describes my method for estimating and validating superintndent value-added, and Section 5 presents results. I explore mechanisms in Section 6 and test for effects on superintendent tenures and pay in Section 7. Section 8 concludes.

1 Superintendents in US School Districts

In the US, school districts are the administrative body responsible for education. Districts vary wildly in size; some enroll fewer than 100 students while the largest districts enroll over 1 million students. School boards, small groups of locally-elected officials, are responsible for the administration of education within a district—responsibilities that typically include hiring a superintendent.[4] While most school board members are part-time volunteers, superintendents are full-time employees of the school district, are typically the highest earning employee in a school district, and are often referred to as the "CEO" of local education.

Consistent with their role as the top executive, superintendents are responsible for a broad range of functions within a school district. These functions vary slightly across districts, but typically include coordinating with district stakeholders including the school board and district staff, monitoring the district budget, managing personnel decisions such as hiring, firing, evaluation, and compensation, and overseeing the daily operations of district schools. A 2018 Gallup survey of 1,892 US superintendents provides useful context on the issues that draw superintendent's attention. The three issues that respondents cited as challenges for their district were "[i]mproving the academic performance of underprepared students," "[t]he effects of poverty on student learning," and "recruiting and retaining talented teachers." As top executives, superintendents have multiple avenues through which they impact district operations and improve performance.[5]

Data also indicate that superintendents turnover is a persistent issue. Survey data from the School Superintendents Association find that the average yearly turnover rate is between 14 and 16 percent.[6] Moreover, less than 40 percent of superintendents surveyed by Gallup have been in their districts for five years or more, and news reports frequently describe superintendents "leaving their posts in droves."[7] Superintendents often leave their jobs to pursue superintendent jobs in other districts, a phenomenon discussed in detail in Grissom and Andersen (2012). In the section below, I describe the data I collect to study the effects of district superintendents systematically.

[4] In a small number of states, districts can choose to select superintendents via local election. While little systematic data exist, Schuh and Herrington (1990) finds that, as of 1990, 97.8 school superintendents were appointed, and that all elected superintendents came from six states: Alabama, Florida, Georgia, Mississippi, Tennessee, and South Carolina. Of these states, only Georgia appears in my sample. In 1992, Georgia enacted a constitutional amendment mandating appointed superintendents. "Lawmakers shoot down option of electing school superintendents to slow revolving door," Atlanta Journal Constitution, March 17, 2017.

[5] "The Gallup 2018 Survey of K-12 School District Superintendents," Gallup.

[6] "Superintendent and District Data," AASA.

[7] "The Gallup 2018 Survey of K-12 School District Superintendents," Gallup. "Who wants to lead America's school districts? Anyone? Anyone?" Hechinger Report, January 6, 2022. "Superintendent Turnover Is a Real Thing. How Bad Is It?" Education Week, February 28, 2022.

2 Data

2.1 Superintendent Roster Data

I collect superintendent roster data from 20 states between 2008 and 2018. At present, these data exist across a number of publicly available state-specific reporting systems. For example, Pennsylvania reports district superintendent rosters within full-count school payroll data, Texas reports superintendent rosters in PDF school directories, and other states, such as Missouri, publish only the current year's superintendent roster online. In cases such as Missouri's, I use the Wayback Machine to retrieve rosters from prior years. Appendix C describes my process for collecting superintendent roster data, separately for each state.

Table 1 summarizes the coverage of my data across states. In total, my data contain 7,822 unique districts. I assess the coverage of my data according to both the share of enrollments covered in these districts as well as the share of districts covered. In 13 of 20 states in my sample, my superintendents data covers over 90 percent of total state enrollment. In 18 of 20 states, my data covers over 80 percent of state enrollment. Coverage with respect to districts is slightly lower, due to relatively lower coverage of rural districts. Still, my data covers roughly 55 percent of total K-12 enrollment in the US and 47 percent of US public school districts.

There are two states where my data covers a relatively lower share of state enrollments population: New York and Illinois. In New York, my data excludes districts in New York City, which together comprise a very large share of total enrollments; in 2014 over 35% of all school children in New York State attended a school in one of the 32 New York City Geographic Districts. In Illinois, my data includes the largest district in the state—City of Chicago SD 299, which enrolled 392,558 students in 2014—but excludes many medium-sized districts with reporting inconsistencies over the sample period. This excluded group includes Schaumburg CCSD 54 and Cicero SD 99, both of which are located in the Chicago suburbs and enrolled over 10,000 students in 2014.

Figure 1 shows where districts in my sample are located. Districts in my data are shown in black, and districts missing from my sample are shown in blue. The size of each displayed point is proportionate to 2014 enrollments. Figure 1 illustrates that most districts missing from my sample are rural and relatively small, and that, consistent with Table 1, my data contain the vast majority of state enrollments for most states in my sample.

Panel A of Table 2 displays summary statistics of my sample districts. The first two columns of Table 2 display district-level averages for all districts in the US and all districts in my sample, respectively. As of 2014, the average district in the US enrolled roughly 3,000 students across 6 schools. Districts in my sample are, on average, reasonably similar: the average district in my sample enrolls roughly 3,500 students across 6.7 schools. Staffing totals are similarly comparable.

The next two columns of Table 2 displays enrollment-weighted averages, which can be interpreted as student level averages. Because larger districts enroll more students, student level average district sizes are substantially larger. The average student in the US is enrolled in a district that enrolls over 40,000 students total and maintains over 75 schools. Enrollment-weighted averages for my sample districts are slightly smaller but similar in magnitude.

Superintendent roster data includes superintendent names and the districts in which they worked each year, but does not identify unique superintendents across years. To link superintendents over time, I implement a record linking algorithm within each state. Appendix D details the linking procedure that I use to identify superintendents over time across and within districts. Broadly, this process involves three steps. First, I make reasonable changes to reported names to make them more consistent over time. This step includes removing middle initials and converting names reported as “Last, First" to “First Last." Second, I identify potential matches within each state using a measure of string similarity. Finally, I break potentially erroneous matches by separating instances in which matched superintendents are working in two districts simultaneously. Later, in Section 3, I use these links to characterize superintendent tenures and transfers.

2.2 Test Score Data

In my analyses of test score data, I link superintendent roster data to district-by-grade-by-year test score data from the Stanford Education Data Archive (“SEDA") (Reardon et al. (2021)). SEDA data contain, for each district, grade (3 through 8), and year between 2008 and 2017, average normed test scores in math and reading and language arts (“RLA") on state-administered standardized tests. Over this period, US states reported test results in the form of counts a small number of ordered categories (e.g. “Not Proficient," “Proficient," “Advanced"). Reardon et al. (2021) translate these scores into normed means using Heteroskedastic Ordered Probit models, accounting for differences in state proficiency thresholds by rescaling to performance on nationally-administered

National Assessment of Educational Progress test.

SEDA data also contains demographic information merged from the National Center for Education Statistics (“NCES"). In my analyses, the set of baseline demographic characteristics I use in value-added estimates includes grade-specific racial shares (percent Asian, Black, and Hispanic), grade-specific share of free lunch students and reduced lunch students, and SEDA's “standardized SES composite," which is meant to proxy for socioeconomic status, and is “computed as the first principal component factor score of the following measures: median income, percent with a bachelor's degree or higher, poverty rate, SNAP rate, single mother headed household rate, and unemployment rate."

At different points in my analysis, I consult two additional sources of test score data. First, to control for levels of achievement prior to 2008 in my value-added estimates, I use data on proficiency rates from the 2003 and 2004 rounds of state-administered No Child Left Behind tests, used in Reback et al. (2014). These data contain school level proficiency rates in math and reading, which I aggregate to the district level. Second, to estimate school level superintendent effects, I use school-by-grade proficiency rate data from EdFacts. This data reports the share of students scoring at or above the state proficiency threshold on state tests between 2009 and 2018. This data provides a higher level of granularity than my district level data, but its units—rates of proficiency—vary substantially across states.

Panel B of Table 2 displays averages with respect to test score data, which also includes average demographic characteristics of students. Mean scores in my sample are reasonably close to nationally normed averages. Relative to all US districts, demographics of students in my sample are quite similar. 13 percent of test-takers are Black and 30 percent are Hispanic. 51 percent of students in my test score data qualify for free or reduced lunch. District averages are slightly lower for these percentages because smaller districts tend to be more White and have smaller shares of students eligible for free and reduced lunch.

2.3 District Characteristics and Financial Data

I measure school inputs—school spending and school staffing—using data reported to the NCES Common Core of Data. In particular, school finance data comes from the School District Finance Survey (Form F33) survey. I focus on levels of spending per pupil and spending shares across the three largest categories of spending: instruction, support services, and capital. I adjust all spending levels to 2021 dollars using the Consumer Price Index retroactive series using current methods. School staffing data comes from the Local Education Agency Universe Survey. As with spending, I represent staffing totals in per pupil terms.

Panel C of Table 2 displays summary statistics with respect to school spending. The average district in my sample spent over $16,000 per student in 2014. Districts allocated approximately half of this spending towards instruction, approximately 30 percent towards support services, and seven percent towards support services.

2.4 School Practices Data

Finally, I use data from the Civil Rights Data Collection ("CRDC") to measure school practices. CRDC data is collected every two years by the U.S. Department of Education's Office for Civil Rights, and measures a range of school-level characteristics, related to discipline, bullying, athletics, curriculum, school finance, and personnel. Given the responsibilities of school superintendents, I focus on outcomes related to curriculum and personnel.

My CRDC data covers the universe of schools in the US every two years from 2009 to 2017. Table 3 summarizes this data for districts in my sample, separately by year and by school level. Curricular choices vary substantially across school levels, but are reasonably consistent over time. Algebra courses are extremely common in high schools and slightly less so in middle schools. Advanced Placement ("AP") courses are found only in high schools, with few exceptions. Across all levels, gifted and talented programs are offered in 50 to 80 percent of schools, with slightly higher rates in middle schools and slightly lower rates in high schools. Across all school levels, average shares of teachers in who were chronically absent—defined as absent more than 10 days of the school year—falls between 20 and 30 percent. Inexperienced teachers are also common; roughly 10 percent of teachers are in their first or second year of teaching.

3 Describing the Superintendent Market

Prior to laying out my strategy for estimating superintendent value-added, I first characterize the superintendent labor market more generally.

3.1 Superintendent Tenures

Table 4 summarizes the structure of superintendent tenures in my data. Panel A of Table 4 shows the distribution of superintendent tenures at the district level. There are 7,822 unique districts in my data, and between 2008 and 2018, the average district had 2.7 superintendent tenures.

Panel B summarizes my data at the superintendent level. There are over 18,000 unique superintendents in my data, most of whom work in only one district over this period. My record linking algorithm identifies a small share of superintendents, 11.7 percent, who work in more than one district. This underscores the necessity of using large-scale data from multiple states in this context. Within each state, the number of multi-district superintendents is quite small. However, across the 20 states in my data, I am able to identify thousands of superintendent transfers across districts.

The statistics in Panel B likely understate the frequency of across-district transitions, because my data include many newly-hired superintendents who will eventually move across districts in the years after my sample has ended. To get a better sense of the long-run rate of across-district transfers, Panel C repeats this exercise, restricting to superintendents who were employed in 2008. Within this sample, a much larger share of superintendents work in more than one district: 17.6 percent.

Finally, Panel D of Table 4 summarizes the distribution of tenures at the superintendent-by-district level. The average tenure in my data is 3.3 years and the median is three years; these figures are roughly consistent with survey evidence from Gallup, in which the median superintendent was employed in their district for between three and five years.

In a small set of states—Georgia, Pennsylvania, and Wisconsin—I source my superintendent roster data from full-count personnel data that includes salary information. The final row of Panel D summarizes the distribution of mean salaries across district tenures. In these states, average annual salaries were approximately $160,000. Around this mean, salaries vary substantially; the standard deviation is over $48,000 and the maximum is over $500,0000.

3.2 Characterizing Superintendent Job Transitions

Linking superintendents across allows me to identify when a superintendent exits one district and enters another. To characterize the nature of these job transitions, I identify every instance in which a superintendent left one district and later worked for a different district. There are over 2000 such transitions in my data.

Table 5 summarizes these transitions by comparing the characteristics of districts that a superintendent left—the exiting district—to the characteristics of districts that a superintendent entered—the entering district. The first and second columns present means and standard deviations of characteristics for the exiting and entering district, respectively, where characteristics are measured in the relevant year—in the exiting year for the exiting district and in the entering year for the entering district. The final column displays the results of a linear regression on a stacked dataset in which each transition appears in two rows, each row representing either the exiting or entering district. With this data, I test for differences between exiting districts and entering districts by regressing each characteristic on a dummy variable identifying the entering districts. I include fixed effects for each transition in each regression.

Table 5 highlights two facts about superintendent transitions. First, superintendent transitions often entail moving to a larger district. On average, enrollments in entering districts are higher by over 1,000 students, an increase of over 35 percent. (Differences in the log of enrollments imply an increase of similar magnitude.) Entering districts have 1.3 more schools, and higher shares of underrepresented minority students, though the magnitude of this latter effect is reasonably small: two percentage points.

Second, while superintendents typically leave to slightly higher-scoring districts, these changes are not coincident with trends in district test scores. The final two rows in Table 5 illustrate this point. Average scores in entering districts are roughly 0.017 standard deviations higher than scores in exiting districts. This fact is not particularly surprising, given the differences in district characteristics described above. However, I find no evidence that transitions are correlated with changes in test score trends. The bottom row of Table 5 shows that cohort level changes in test scores—defined as the year-over-year change in a cohort's test scores (for example, the difference between 4th grade class's scores in 2012 and the 3rd grade class's scores in 2011) in both exiting and entering districts are extremely small in magnitude. The difference between these two values is similarly small and statistically insignificant. In a study of superintendent turnover in California, Grissom and Andersen (2012) document similar patterns: superintendent turnover is uncorrelated with short-term test score growth, but typically involves moves to larger and more urban school districts.

3.3 Superintendent Tenure and Changes in District Practices

Finally, I provide suggestive evidence that superintendents meaningfully affect district practices. To do so, I link my superintendent roster data to CRDC school practices data. With this linked data, I test whether early-tenure superintendents are more likely to make changes in school policies.

I focus on the five school practices variables in Table 3. For all practices other than the share of inexperienced teachers, I calculate the absolute change in school i in consecutive surveys between year t and year $t-2$. (CRCD data is reported every two years.) For binary variables, this is simply a binary variable indicating whether school i changed a given practice between two consecutive survey years. The share of inexperienced teachers already measures changes, so I use it directly in my analysis.

For a given practice—for example, whether the school offers algebra—I test whether higher tenure superintendents are less likely to introduce changes in that practice. I estimate the equation below via OLS.

$$\mathbb{I}\{\text{Changed Algebra}\}_{sdt} = |\text{Algebra}_{sd,t} - \text{Algebra}_{sd,t-2}| = \beta_0 + \beta_1 \text{Tenure}_{dt} + \gamma X_{sdt} + \varepsilon_{sdt}, \quad (1)$$

where $\mathbb{I}\{\text{Changed Algebra}\}_{sdt}$ indicates whether school s in district d changed algebra practices between year t and year $t-2$. Tenure_{dt} represents the tenure (in years) of the superintendent in district d in year t. X_{xdt} is a matrix of covariates. If superintendents exercise some authority over school practices, we might expect that earlier-tenure superintendents are more likely to change school practices. Over the long-term, schools “settle in" to a set of practices that align with the superintendent’s preferences. In this case, we expect β_1 to be negative: the probability of changing practices is a decreasing function of superintendent tenure.

My results are shown in Table 6. Column 1 displays estimates that control only for state-by-year-by-school-level fixed effects. Estimates are uniformly negative and statistically significant for four of five practices, suggesting that increases in tenure decrease the likelihood of a change in practices. Per-year magnitudes are generally in the range of two basis points, suggesting that a

5-year difference in superintendent tenure decreases the probability of a given practice changing by one percentage point.

Column 2 of Table 6 adds fixed effects for districts. These fixed effects control for differences in the year-over-year variability in practices within districts—for instance, some districts may be consistently more or less variable over time, regardless of the superintendent's tenure. Adding these fixed effects increases the explanatory power of my mode, but reduces estimated effect sizes substantially. Still, effects are uniformly negative but statistically imprecise. Column 3 repeats this exercise, adding school fixed effects, with similar effects.

Broadly, these results are consistent with a model in which superintendents affect school practices, but these effects are concentrated earlier in their tenure. As superintendents spend more time in a district, educational practices align more closely with their vision, and the likelihood that a given practice changes in a given year falls.

Thus, superintendents have scope to change education production in school districts. However, this descriptive analysis cannot identify whether superintendents affect district outputs–that is, student outcomes—rather than just inputs, and by how much. Moreover, if superintendents do indeed affect student outcomes, this analysis cannot determine whether effective or ineffective superintendents make different choices about district operations.

To answer these more pressing questions, I use tools from the value-added literature. Specifically, I estimate a model of superintendent value-added and use out-of-sample value-added estimates to validate my model and to identify best practices: practices that reliably effective superintendents tend to enact upon entering a new district. However, estimating superintendent value-added is complicated by the dynamics that I describe above: superintendent-induced changes are dynamic, and are not realized immediately upon entering a district. In the section below, I describe my method for identifying superintendent value-added.

4 Estimating Dynamic Superintendent Effects

In this section, I describe my approach to estimating, validating, and characterizing superintendent value-added. Throughout, I focus my discussion on my main test score value-added estimates. In Appendix A, I introduce and summarize additional value-added estimates, which differ in the set of covariates included in the estimation or the assumptions made during estimate. These estimates are described in more detail in the body of Appendix A, but the setup and estimation is extremely similar in nature to that of the estimates described below.

4.1 Setup

To begin, let average student achievement in district d in year t is given by $\bar{A}_{dt}$. I assume that achievement is determined by the following equation.

$$\bar{A}_{dt} = X_{dt}\beta + v_{dt}, \qquad \text{where } v_{dt} = \theta_{dt} + \varepsilon_{dt} \tag{2}$$

where X_{dt} is a matrix of measures of prior district performance—which may include lagged test scores—and contemporaneous district demographics. θ_{dt} represents the district level superintendent effect, whose structure is described in detail below. All other variation in test scores is assumed to be attributable to an idiosyncratic district-year shock ε_{dt}.

As in Coelli and Green (2012) for school principals, district level superintendent effects θ_{dt} are assumed to evolve over time based on current and prior district superintendents. Denote the superintendent in district d in year t as $j(d,t)$, and denote superintendent j's "full" effect as $\mu_{j(d,t)}$. (For brevity, I will often use μ_j in place of $\mu_{j(d,t)}$.) This effect is revealed over a superintendent's tenure. Specifically, I assume that the district-level superintendent effect θ_{dt} is a weighted average of the superintendent effect in the prior year and the current superintendent's full effect $\mu_{j(d,t)}$, where weights are given by a decay parameter, ρ.

$$\theta_{dt} = \rho\theta_{d,t-1} + (1-\rho)\mu_{j(d,t)} \tag{3}$$

Equation 3 can also be represented as a function of the full history of superintendents in a given district, shown below.

$$\theta_{dt} = \rho^t\theta_{d,0} + (1-\rho)\sum_{t'=1}^{t'=t}\rho^{(t-t')}\mu_{j(d,t')} \tag{4}$$

In both equations above, ρ is a measure of persistence. To build intuition, consider two extremes. First, if $\rho = 0$, a superintendent's full effect is realized immediately upon entering a new district. Mathematically, this implies that $\theta_{dt} = \mu_{j(d,t)}$. While this assumption may be reasonable for teachers, whose effect on students is year-specific, this is unlikely to be the case for superintendents, if superintendents impose changes to district operations only gradually. Alternatively, if $\rho = 1$ changes in superintendents have no effect on district operations. Mathematically, $\theta_{dt} = \theta_{d,0}$.

For intermediate values of ρ between 0 and 1, higher values of indicate more persistence in prior superintendent effects and slower diffusion of new superintendent effects. In these cases, θ_{dt} approaches $\mu_{j(d,t)}$ as a superintendent's tenure in a district increases. (Note that this approach necessarily shuts down the possibility of experience effects.)

Figure 2 illustrates how θ_{dt} changes in response to a one-unit change in superintendent value-added, under different assumptions about ρ. The top line in Figure 2, shown in solid blue, is the immediate-realization case described above, in which $\rho = 0$. As ρ increases, the speed with which a superintendent's "full" effect is reflected in district outcomes becomes slower. At $\rho = 0.75$, superintendents "reveal" less than half of their full capability after three years.[8]

On an individual basis, Equation 3 can be represented as a function of two forces: tenure and decay. To do so, consider the effect in year t of a superintendent who worked in a district for x years, most recently in year q. (Note that this means that, during a superintendent's tenure $t = q$.) The coefficient on this superintendent's μ_j term in year $t \geq q$ can be represented as as below.

$$(1-\rho) \sum_{t'=q-x+1}^{t'=q} \rho^{t-t'} \tag{5}$$

[8]Mathematically, if $\rho = 0.75$, $1 - \rho^2 = 0.4375$.

Distributing $(1-\rho)$, this term simplifies as follows.

$$
\begin{aligned}
(1-\rho)\sum_{t'=q-x+1}^{t'=q}\rho^{t-t'} &= \sum_{t'=q-x+1}^{t'=q}\rho^{t-t'} - \sum_{t'=q-x+1}^{t'=q}\rho^{t-t'+1} \\
&= \sum_{i=t-(q-x+1)}^{i=t-q}\rho^{i} - \sum_{i=t-(q-x+1)+1}^{i=t-q+1}\rho^{i} \\
&= \left[\rho^{t-(q-x+1)} + \rho^{t-(q-x+1)-1} + \rho^{t-(q-x+1)-2} + \ldots + \rho^{t-q}\right] \\
&\quad - \left[\rho^{t-(q-x+1)+1} + \rho^{t-(q-x+1)} + \rho^{t-(q-x+1)-1} + \ldots + \rho^{t-q+1}\right] \\
&= \rho^{t-q} - \rho^{t-q+x} \\
&= \underbrace{(1-\rho^{x})}_{\text{Effect of Within-District Tenure}} \times \underbrace{\rho^{t-q}}_{\text{Effect of Decay}}
\end{aligned} \tag{6}
$$

The equation above illustrates the two drivers of superintendent effects. First, within-district tenure increases the magnitude of a superintendent's effect on district outcomes. In the limit (as a superintendent's tenure, x, approaches infinity) a superintendent's effect is fully realized: this term approaches 1. Second, after a superintendent leaves a district, their effect on district operations decays. Mathematically, this is reflected in the relationship between the year of observation, t, and the second term above. Holding a superintendent's year of exit, q, constant, increases in t decay their effect on district operations. (Note that during a superintendent's tenure, this second term is equal to one, as $t-q=0$.)

Returning to Equation 2, plugging in Equation 4 yields the equation below.

$$
\bar{A}_{dt} = X_{dt}\beta + v_{dt}, \qquad \text{where } v_{dt} = \rho^{t}\theta_{d,0} + (1-\rho)\sum_{t'=1}^{t'=t}\rho^{(t-t'+1)}\mu_{j(d,t')} + \varepsilon_{dt} \tag{7}
$$

4.2 Estimation

Changes in superintendents within districts over time allow me to separately identify ρ and μ_j in Equation 7. I define the least squares estimation problem as:

$$
\min_{\beta,\mu,\rho}\sum_{d=1}^{N_d}\sum_{t=0}^{T}\left(\bar{A}_{dt} - X_{dt}\beta - (1-\rho)\sum_{t'=1}^{t'=t}\rho^{(t-t'+1)}\mu_{j(d,t')}\right)^{2}. \tag{8}
$$

This equation is estimable via nonlinear least squares. However, in practice, estimation presents two difficulties. First, given the scale of my data—over 18,000 individual superintendent tenures across over 7,000 districts—computation is extremely costly. Second, because the identifying equation does not have well-defined asymptotic standard errors, inference on my parameter of interest, ρ, is difficult.

I overcome these difficulties by estimating Equation 8 in two steps. First, to estimate the distribution of $\hat{\rho}$, I estimate Equation 8 via nonlinear least squares using 100 bootstrapped subsamples of 1,000 districts each.[9] These subsamples are much smaller than my full sample, and therefore they overestimate the variance in my estimates of ρ. In practice, bootstrapped distributions are centered around a reasonably small range of values. In the second step, I rely on the fact that, for a given value of ρ, Equation 8 is linear and can be estimated via ordinary least squares.[10] To obtain value-added estimates, I estimate value-added terms via ordinary least squares after setting ρ equal to the mean value from the distribution of bootstrap estimates.

Practically, with test score data, which is available annually at the district-by-grade-by-subject level, I construct district-by-year residuals via OLS prior to estimation. In my main estimates, mt vector of controls, X_{dt} includes the log of enrollment, the SEDA socioeconomic status composite measure, racial shares, fixed effects for locale type (urban, suburb, town, and rural), and three measures of prior test score performance: mean performance on 2003-2004 No Child Left Behind tests from Reback et al. (2014), 2008 grade-by-subject specific test scores from SEDA, and cohort-specific lagged test scores, which measure the score of students in the prior grade in the prior year (under normal circumstances, these will mostly be the same set of students).[11]

In estimation, I set the initial leadership effect, θ_{d0}, equal to zero and the first superintendent effect, μ_{d0}, equal to zero. In addition, I omit all effects for superintendents who work in a district for only one year, who are primarily hired on an interim basis.

Moreover, in my analysis I estimate district-by-superintendent effects, rather than superintendent effects. This distinction affects only superintendents who work in more than one district (e.g. a superintendent who works in one district from 2009 to 2013 and another from 2014 to 2018), but

[9] Specifically, I use the nls function in R, setting the initial value of ρ equal to 0.5.

[10] To see this, note that x and $t - q$ in Equation 6 are values observed in superintendent roster data.

[11] Because I include 2008 scores and cohort lagged scores as controls, I exclude observations from the first year (2008) and first grade (grade 3) from my analysis.

has two distinct advantages. First, it allows me to estimate the within-superintendent variance of my estimates, which allows me to scale my estimates to account for sampling error. Second, it allows me to validate my results by testing whether a superintendent's estimated value added in one district predicts changes in test scores upon entering another district. I describe both of these processes below.

As is the case with teacher value-added, my raw estimates of superintendent value-added from Equation 8 are inflated by sampling error.[12] I form empirical Bayes estimates of value-added using the equation below.

$$\hat{\mu}_j^{EB} = a\hat{\mu}_j + (1-a)\bar{\mu}, \tag{9}$$

where $\bar{\mu}$ is my estimate of the average superintendent value-added, and a is a weighting parameter. I estimate a as below.

$$a = \frac{\hat{\sigma}^2}{\hat{\sigma}^2 + \hat{\lambda}}, \tag{10}$$

where $\hat{\sigma}^2$ is an estimate of the variance of the true value-added and $\hat{\lambda}$ is the estimated variance of the error in the value-added estimate. Following Kane and Staiger (2008), I estimate $\hat{\sigma}^2$ as the covariance between estimated value-added in the superintendent's current district and estimated value-added in their prior district (using the subsample of superintendents who worked in multiple districts).

4.3 Validation

To validate my estimates, I focus on instances in which a superintendent enters a new district. My validation exercise tests whether a superintendent's out-of-district effects predict changes in a new district, and whether the magnitude of these effects are consistent with the model predictions described above.

Figure 3 illustrates this approach using a simple example. In the example, Superintendent J worked in District A for years 2 and 3. Between years 3 and 4, Superintendent J moved from

[12] See Koedel and Rockoff (2015) for a discussion of shrinkage procedures in value-added modelling, and Morris (1983) for a more general discussion.

District A to District B. In the first step of my approach, I estimate μ_J^A: Superintendent J's effect on outcomes in District A. In the second step, I test whether μ_J^A predicts changes in District B.

More technically, consider the effect of a superintendent who worked in a district for x years, most recently in year q. For simplicity of notation, denote the superintendent's first year in the district as $r = q - x + 1$. If my estimate of that superintendent's value-added is unbiased, changes in θ_{dt}, the district level superintendent effect, should evolve according to the equation below.

$$\Delta\theta_{dt} = \begin{cases} (1-\rho^{t-r})\mu_{j,-d} & \text{if } t \in [r, ..., q] \\ \left[(1-\rho^{x}) \times \rho^{t-q}\right]\mu_{j,-d} + (1-\rho)\sum_{t'=q+1}^{t'=t} \rho^{(t-t')}\mu_{j(d,t')} & \text{if } t > q \end{cases} \tag{11}$$

In the equation above, I use $\mu_{j,-d}$ to denote superintendent j's estimated value-added in districts other than district d. The first line above corresponds to superintendent effects during a superintendent's tenure. As superintendent tenure within a district increases, the superintendent realizes more of their full effect, and this term approaches μ_j. The second line concerns effects after a superintendent has left. As time passes following a superintendent's exit from a district, their impact on district outcomes dissipates, and the effect of more recent superintendents increases.

The second term in the bottom line, $(1-\rho)\sum_{t'=q+1}^{t'=t} \rho^{(t-t')}\mu_{j(d,t')}$, denotes the effects of more recent superintendents: superintendent who succeed the denoted superintendent. In my validation exercise, I present results making two extreme assumptions about the relationship between present superintendent quality and future superintendent quality, which I denote with μ^{+}. First, I simply assume that an exiting superintendent is replaced by an average superintendent who remains in the district indefinitely: $\mu^{+} = 0$. Second, I assume that an exiting superintendent is replaced by a superintendent of exactly the same quality who remains in the district indefinitely: $\mu^{+} = \mu_{j,-d}$.

To test these predictions, I construct a test score panel consisting of all available years before and after a superintendent enters a district. I calculate predicted changes in test scores, $\Delta\hat{\theta}_{dt}$, using Equation 11. (For pre-entry years, $t < r$, I set $\Delta\hat{\theta}_{dt} = 0$.) With these data, I estimate the equation below and test whether $\beta_1 = 1$. γ_{dj} and ξ_t represent district-superintendent (i.e. event-specific) and time fixed effects, respectively.

$$y_{dt} = \beta_0 + \beta_1 \left[\Delta\hat{\theta}_{dt}\right] + \gamma_{dj} + \xi_t + \varepsilon_{dt} \tag{12}$$

I estimate Equation 12 via OLS. I cluster standard errors at the district level.

This estimate is akin to a difference-in-difference estimate, comparing changes in districts that are "treated" with superintendents with different value-added estimates at different times. In this spirit, I use this setup to estimate event study models, which allow me to estimate the time path of superintendent effects nonparametrically, and are slightly more explicit about the comparisons used to identify causal effects.

However, recent work in the econometrics literature raises concerns about the use of canonical two-way fixed effects models when treatment timing is staggered and treatment effects vary over time (Goodman-Bacon (2021)). To ameliorate such concerns, I estimate stacked difference in difference models (Baker et al. (2022), Cengiz et al. (2019)). To do so, I construct my analysis dataset as follows. First, I identify each instance of superintendent entry for which I have an out-of-sample value-added estimate $\mu_{j,-d}$. I denote events with e. Next, for each event e, I identify all districts in the same commuting zone with no such entry over the entire sample period. Commuting zones are "geographic units of analysis intended to more closely reflect the local economy where people live and work."[13] These districts are my control districts and are shown on the right in Figure 3. Finally, I estimate the equation below, which controls for event-by-time and event-by-district fixed effects.

$$y_{edt} = \alpha_{ed} + \lambda_{et} + \sum_{k \neq -1} \beta_k [D_{edk} \times \mu_{j(e),-d}] + \varepsilon_{edt}, \tag{13}$$

where D_{edk} is a binary variable equal to one in year k relative to the year of superintendent entry. (D_{edk} is equal to 0 for all control districts in all years.) My coefficients of interest, β_k, trace out the dynamic effect of superintendent value-added in the years before and after a superintendent enters a new district.

[13] There are 709 commuting zones in the US. "Commuting Zones and Labor Market Areas," USDA Economic Research Service.

The analogous difference-in-difference equation is given below.

$$y_{edt} = \alpha_{ed} + \lambda_{et} + \beta[D_{edt} \times \mu_{j(e),-d}] + \varepsilon_{edt}, \tag{14}$$

where D_{ed} is a binary variable identifying the treated district in years after superintendent entry. In my results, I present estimates of Equation 13 in the form of event study plots and estimates of Equation 14 in tables. I cluster standard errors at the district level.

This approach has three distinct benefits, relative to the validation exercise described above. First, by saturating my specification with event indicators, treatment effects are identified based only on comparisons between treated districts and untreated districts in the same year. As Baker et al. (2022) notes, this approach "circumvent[s] the problems introduced by staggered treatment timing and treatment effect heterogeneity." Second, this approach is more explicit about the comparison units I use to construct my estimates: districts in the same commuting zone with no eligible superintendent entry events during my sample period. Finally, my event study specification in Equation 13 places no parametric restrictions on the effect of superintendents on outcomes by year, allowing me to test the degree to which these estimates conform to the predictions of my model (and shown visually in Figure 2).

Identification of causal effects in this context rests on the assumption that, absent the entry of a high- or low-value-added superintendent, outcomes in the treated district would follow the same path as those in the control districts. While I cannot test this assumption directly, I confirm that my estimates do not exhibit significant differences in trends in the pre-period, both with respect to test scores as well as a host of demographic characteristics. Additionally, I test for treatment effects on school demographics by estimating Equation 13 with student demographic characteristics on the right hand side. In doing so, I confirm that the entry of a high- or low-value-added superintendent is not coincident with changes in student demographics.

4.4 Identifying Practices of Effective Superintendents

The validation exercise described above can be applied to answer a related question: what do higher value-added superintendents do? To answer this question, I test how the arrival of a high or low value-added superintendent affects various outcomes at the district or school level: district spending or teacher absences, for example.

Throughout, I use the stacked difference-in-differences approach, where treated districts are compared over time to districts in the same commuting zone in the same year. My estimates from Equations 13 and 14 reflect comparisons between districts in which a high- or low-value-added superintendent entered and nearby districts where no such event occurred.

In measuring effects for school-level outcomes in CRDC data, I make two small changes to my approach. First, because districts vary wildly in the number of schools they manage, unweighted school level results will implicitly place more weight on larger districts. To gauge the degree to which this affects my estimates, I produce two estimates that apply different weights to each school. In one set of estimates, I apply equal weight to each district in my sample, weighting each school by its proportion of total district enrollment in each year. In another set of estimates, I weight each observation by total school enrollment. The second change results from the structure of CRDC data. CRDC data is collected every two years. As such, I pool event study estimates to include two year periods relative to the year of superintendent entry (years -2 and -1, years 0 and 1, etc.).

5 Superintendent Value-Added

5.1 Distribution of Superintendent Value-Added

I first summarize the distribution of superintendent value-added and contextualize it relative to effect sizes in other educational settings. Figure 4 summarizes the distribution of my parameter estimates: ρ, my measure of managerial persistence, and μ_j, superintendent value-added.

In Panel A of 4, I show the sampling distribution of bootstrap estimates of ρ. This distribution is centered around a mean of 0.71, indicating that, in a superintendent's first year within a district, only 36 percent of their full effect is realized. 95 percent of bootstrap estimates fall between 0.51 and 0.88. This distribution is wide, but it is notable that none of the bootstrap samples produce estimates consistent with a model of immediate realization of superintendent effects: $\rho = 0$.

Panel B shows the distribution of superintendent value-added. These values are produced by first estimating 8 via OLS, setting ρ equal to its mean bootstrap estimate, 0.71. In the second step, I shrink these raw estimates using Equation 10.

I display summary statistics for my value-added estimates in the top left corner of Figure 4. The standard deviation of my value-added measure is 0.011 standard deviations. The difference between the 90th and 10th percentile value-added is much larger: 0.024 standard deviations. I contextualize these estimates first relative to the overall distribution of observed district-level variation in learning rates as well as relative to two well-established literatures: the teacher value-added literature and the literature on the effects of school spending on student outcomes.

Appendix Figure B1 shows the observed distribution of learning rates—defined as the raw difference in normed test scores between consecutive years for the same cohort—across districts in the US. In the US as well as within the subset of states in my superintendent roster data, the standard deviation of district-level learning rates is 0.044; roughly four times the standard deviation of estimated superintendent effects.

Teacher value-added estimates also provide a useful point of comparison. Hanushek and Rivkin (2010) summarize the magnitude of teacher value-added estimates, finding that the average the average (across-study) standard deviation of teacher effects is 0.11 student-level standard deviations in reading and 0.15 in math. My estimates are roughly one-tenth the size of these effects. However, school districts are much larger than individual classrooms. To provide a more intuitive comparison, one might ask: how many high-quality teachers would have the same effect as a high-quality superintendent?

In Figure 5, I calculate the total district level effect of replacing 20 teachers with teachers with one standard deviation higher value-added. If the standard deviation of teacher value-added is given by σ_{TVA}, the effect of such a policy on average district scores is:

$$\sigma_{TVA} \times 20 \times \frac{\text{Avg. Classroom Size}}{\text{Total District Enrollment}}$$

In Panel A of Figure 5, I display results assuming an average classroom size of 25 and setting the upper and lower bounds to reflect estimates from Hanushek and Rivkin (2010) in math and reading, respectively.

The effect of such a policy at the district level depends crucially on the size of the district in question. In smaller districts with fewer teachers, improving the value-added of 20 teachers can have very large effects; these teachers comprise a large share of the overall teaching staff. However, in larger districts, this effect gets smaller, as the share of classrooms affected by the policy falls. Comparing these estimated policy effects of large-scale improvement in teacher value-added to a hypothetical increase in superintendent value added (shown in Panel C), I estimate that the two policies would have equal effects in a district with total enrollment of 5,000.[14] (For larger districts, increasing superintendent quality would have larger effects than increasing teacher quality, and vice versa.)

How much might such a policy cost, if school districts use higher salaries to attract more effective teachers? Estimates will depend on the degree to which changes in salaries translate to changes in teacher value-added; in their study of how recessions affect that quality of entrant teacher cohorts, Nagler et al. (2020) provide a sense of magnitudes. Recessions increase the relative pay of a teaching job by between $2,379 and $7,140, and increase average value-added in math by roughly 0.1 test score standard deviations, roughly equal to σ_{TVA}. Taking these estimates at face value, a policy that recruits 20 teachers higher value-added teachers with higher salaries would cost between $47,580 and $142,800.

Another useful comparison is to consider the effect of school spending on student achievement and ask: what amount of spending would have the same effect as replacing an average superintendent with an above-average superintendent? Jackson and Mackevicius (2021) summarize the literature on school spending and student achievement, arguing that on average, a $1,000 increase in school spending for four years increases test scores by 0.0352 standard deviations. In Panel B of Figure 5, I show the estimated effect of a $1 million increase in district spending as a function of district size.

I calculate effects on average district test scores as:

$$\frac{1,000,000}{\text{Total District Enrollment}} \times \frac{\delta}{1,000}$$

where δ is the estimated effect of a $1,000 increase in school spending for four years. Figure 5

[14] Relevant calculations are below.

$$\sigma_{SVA} = \sigma_{TVA} \times 20 \times \frac{\text{Avg. Classroom Size}}{\text{Total District Enrollment}}$$

$$0.011 = 0.11 \times 20 \times \frac{25}{\text{Total District Enrollment}} \quad \rightarrow \quad \text{Total District Enrollment} = 5,000$$

shows upper and lower bounds that correspond to the 95 percent confidence interval in Jackson and Mackevicius (2021): -0.00725 to 0.07767. Taking the mean estimate in Jackson and Mackevicius (2021)—0.0352 standard deviations—I estimate that the two policies would have equal effects in a district with total enrollment of 3,200.[15]

While these comparative cost estimates are inexact, they highlight the potential for improvements in management quality to improve school quality at relatively lower cost that the aforementioned interventions. In three states for which I have salary data, Table 4 shows that average salaries are approximately $160,000, far below the $1 million cost associated with spending-driven improvements, and in the same range as teacher-driven improvements. Of course, this comparison assumes that districts are able to identify and reward effective superintendents; I return to this question in Section 7.

5.2 Testing for Out-of-District Validity

I next evaluate the out-of-district validity of my estimates. To do so, I test whether changes in student test scores following superintendent entry are in line with the magnitude of my value-added estimates. Formally, I estimate Equation 12 using out-of-district estimates of superintendent value-added and test whether the estimated β_1 coefficient is equal to 1. There are roughly 1,500 eligible instances of superintendent entry in my data.

My estimates are shown in Table 7. As noted above, I estimate Equation 12 under two extreme assumptions about μ^+, the value-added of subsequent superintendents after the entering superintendent ultimately leaves the district. In the first three columns of Table 7, I assume that exiting superintendents are replaced with superintendents of average quality: $\mu^+ = 0$. Estimates center around 1, but are imprecise absent controls for lagged cohort test scores. Adding lagged cohort test scores in Column 3 yields estimates of 1.02; I reject the null hypothesis that this estimate is equal to zero and cannot reject that this estimate is equal to 1.

Columns 4 to 6 of Table 7 repeat these estimates, here assuming that exiting superintendents

[15] Relevant calculations are below.

$$\sigma_{SVA} = \frac{1,000,000}{\text{Total District Enrollment}} \times \frac{\delta}{1,000}$$
$$0.011 = \frac{1,000,000}{\text{Total District Enrollment}} \times \frac{0.0352}{1,000} \quad \rightarrow \quad \text{Total District Enrollment} = 3,200$$

are replaced with superintendents of exactly the same quality, $\mu^+ = \mu_{j,-d}$. Columns 4 and 5 are well above 1, indicating that a one-unit increase in superintendent value-added increases test scores by 2 times the predictions implied by my model. In Column 6, I add lagged cohort test scores; resulting estimates are 1.3. Again, I cannot reject the null hypothesis that this estimate is equal to 1.

Figure 6 presents the equivalent event study estimates of superintendent effects, which provide nonparametric estimates of dynamic superintendent effects. Moreover, unlike the estimates in Table 7, these estimates use my stacked difference-in-differences approach to estimate causal effects. This approach compares the evolution of test scores in districts where a superintendent entered to districts in the same commuting zone with no such event over the sample period.

In Panel A of Figure 6, the dashed horizontal line is drawn at $\beta_k = 1$, which corresponds to forecast-unbiased estimates. The first series, shown as blue circles, shows the dynamic estimates of superintendent entry over time, controlling for baseline demographic controls. Prior to superintendent entry, district outcomes show little relationship with the value-added of future superintendents. In the years following superintendent entry, test scores increase steadily for roughly three years and are approximately flat thereafter. Adding lagged scores—shown in yellow triangles—reduces the magnitude of these estimates only slightly.

In Panels B and C of Figure 6, I estimate similar models, replacing my scaled value-added estimates with re-scaled measures of test score value-added. In Panel B, I show that, consistent with the magnitudes discussed above, a one standard deviation increase in superintendent value-added leads to an improvement in test scores on the order of 0.01 to 0.02 test score standard deviations. In Panel C, estimated effects reflect the impact of a 100 percentile increase in superintendent value-added. Estimated effects sizes in Panel C are between 0.02 and 0.05 test score standard deviations.

Table 8 disaggregates this analysis into comparisons between equal-sized discrete groups of superintendents: superintendents in each quartile of the value-added distribution. I operationalize these comparisons by estimating Equation 13, replacing $\hat{\mu}_{j,-d}$ with binary variables indicating whether the district hired a superintendent with 2nd-quartile, 3rd-quartile, or 4th-quartile value-added. (Hires whose value-added is in the bottom quartiler are the reference group.) Reassuringly, the results in Panel B indicate that superintendent effects are driven both by test score losses in districts that hire low value-added superintendents as well as test score gains in districts that hire high value-added superintendents.

Before discussing heterogeneity and robustness, I situate my estimates relative to other estimates of top school management in the literature. In the economics literature, two studies estimate superintendent value-added directly. Lavy and Boiko (2017) study school CEOs in Israel, managers who oversee approximately 15 schools at a time. Similar to mine, their estimation strategy leverages CEO transitions between districts. While their sample size is quite small—fewer than 60 superintendents per year—the authors find that superintendents have significant and large effects on school performance; in their sample, a one standard deviation improvement in school superintendent value-added increases test scores by 0.04 standard deviations. This estimate is nearly four times the size of my estimate, though I note that the institutional context is quite different and, as detailed in Section 5.3, some subsamples of my data produce estimates of roughly this magnitude.

In addition, Chingos et al. (2014) use data from Florida and North Carolina, and argue that "[s]uperintendents are largely indistinguishable," based on two facts. First, overall achievement changes very little in response to superintendent turnover. Second, in North Carolina, average differences in test scores between successive superintendents are extremely noisy and that "the difference in the academic performance of students under two superintendents who serve successively in the same district, cannot be estimated with sufficient precision to permit the reliable identification of winners and losers."

In light of this work, I offer three responses. First, changes in student achievement are a function of superintendent quality rather than superintendent tenures. In Section 7, I show that school districts often fail to retain high value-added superintendents, suggesting that the link between superintendent tenure and superintendent quality may be weak. Thus, while replacing one superintendent with another may fail to generate test score gains on average, these effects likely differ based on the quality of the exiting and entering superintendent. Second, my work demonstrates that the dynamics of superintendent effects are extremely important in estimation. Comparing two-year averages (as in Chingos et al. (2014)) between successive superintendents may understate actual differences in effectiveness by as much as 50 percent.[16] Finally, while I

[16]To see this, note the first two years of a superintendent's tenure will reflect $1-\rho$ and $1-\rho^2$ of the superintendent's effect, respectively. Thus, an equal-weighted average of these values is $\frac{1-\rho+1-\rho^2}{2}$. At $\rho = 0.71$, this value is equal to

argue that superintendent value-added estimates are economically meaningful, I am in agreement with Chingos et al. (2014) that they are not sufficiently precise to be employed in the "reliable identification of winners and losers" on an individual basis.

5.3 Heterogeneity

Next, I explore heterogeneous effects across districts and student populations. To do so, I estimate Equation 14 with different subsamples and different student populations. Specifically, I identify 8 (overlapping) subsamples of interest: test subject (math and RLA), district size (large district and small districts), achievement level (districts with high versus low baseline achievement), and union status (states with high versus low union power, as described in Brunner et al. (2020)). Within these 8 subsamples, I estimate effects on 6 student populations: all students, economically disadvantaged and non-economically disadvantaged students, Black students, Hispanic students, and White students. I estimate Equation 14 using my baseline set of covariates plus lagged cohort test scores, and report the effects of a one standard deviation increase in superintendent value-added. Figure 7 displays my results. (I provide definitions of each subsample in the notes for Figure 7.)

I highlight three patterns that emerge from Figure 7. First, superintendent value-added is not subject-specific. Both math and RLA scores exhibit significant improvements across many subsamples. Second, districts with more managerial flexibility—namely, those that are smaller and districts in states with less union power—exhibit larger responses to superintendent effects. Finally, superintendent effects are most pronounced among generally lower-achieving subgroups. Districts with lower levels of baseline achievement, economically disadvantaged students, and minority students (particularly Hispanic students), often exhibit larger responses to superintendent value-added.

5.4 Robustness and Extensions

The identifying assumption underlying this work is that, absent the arrival of a high- or low-value-added superintendent, district outcomes would have followed the same trend as outcomes in control districts. In Appendix Figure B2 I test for effects of superintendent value-added on student characteristics. To do so, I simply re-estimate the test score event study regressions above,

0.393.

replacing test scores with student demographics. Figure B2 demonstrates two facts. First, the arrival of a high- or low-value-added superintendent is not preceded by differential trends with respect to student demographics. High value-added superintendents don't systematically seek out districts with decreasing shares of disadvantaged students, or vice versa. Moreover, effects after superintendent entry indicate that the characteristics of test takers do not change substantially after the arrival of a high- versus low-value-added superintendent. My estimates are quite small, ruling out changes larger than one percentage point in response to a one standard deviation change in superintendent value-added.

I also confirm that my validation estimates are not driven by any one state. To do so, I estimate my validation equation, Equation 12, 20 times, once after excluding each state individually. The results are shown in Appendix Figure B3, which shows that my results are stable, regardless of which state is excluded in validation.

In addition, in Appendix A, I examine how two changes to my value-added model affect my estimates. First, I produce value-added estimates assuming no dynamics: setting ρ equal to zero during estimation. These value-added estimates are highly correlated with those presented in the body of this paper: the rank-rank correlation between these two series is 0.89. As such, using these estimates in validation produces very similar results. However, the magnitude of value-added estimates that assume away tenure dynamics is, on average, 1.9 times smaller than those in the body of this paper. This is sensible; estimates that do not account for tenure dynamics will weight early-tenure dynamics equal to those later in a superintendent's tenure. Because these early-tenure observations are less reflective of a superintendent's full effect, estimates will generally be smaller in magnitude than estimates with tenure dynamics.

Next, I test the degree to which a simpler model of superintendent value-added has out-of-sample validity. To do so, I estimate a model of superintendent value-added identically to the one described in text, with one change: I exclude cohort lagged test scores. The validation estimates resulting from this exercise are shown in Table A1. Across all specifications, out-of-district value added correlates positively with test score increases upon entering a new district. However, many of these estimates are statistically insignificant, and all fail tests for forecast-unbiasedness. I conclude that, similar to teacher value-added (as documented in Chetty et al. (2014) and elsewhere), including lagged test scores is essential to yielding forecast-unbiased estimates of superintendent value-added.

6 What Explains Management Effects?

In this section I explore potential mechanisms behind superintendent effects. I consider three potential explanations, which are largely informed by the literature on education production in the US. First, I test whether superintendent effects reflect changes in within-school test score performance or are caused by reallocating students to different schools. Second, I test whether effective superintendents systematically change the levels or allocation of district spending and staffing. Finally, I test for more subtle changes in school-level staffing and curriculum, which I broadly refer to as school practices.

6.1 Within- or Between-School Changes?

The limited work on quantifying management practices has focused primarily on quantifying differences across schools rather than differences within the same school over time (Leaver et al. (2019), Bloom, Lemos, Sadun and Van Reenen (2015)). Moreover, recent work argues that the distribution of school value-added is large, and that policies that reassign students to schools with higher value-added (via closing schools with low value-added and/or opening additional schools) may yield large achievement gains (Angrist et al. (2017)). One such example comes from Newark, New Jersey. In their study of school reforms that started in 2011 in Newark Public School district, Chin et al. (2019) attribute much of the test score gains to "shifting enrollment from lower- to higher-growth" schools. Naturally, this prompt the questions: do effective superintendents realize test score gains by making changes within schools or by reallocating students across schools? To distinguish between these two explanations, I use school-by-grade level proficiency data from EdFacts.

With these data, I estimate two stacked difference in difference models.[17] First, a model that controls for district fixed effects and second, a model that controls for school effects. (Both models include state-by-year-by-grade fixed effects.) To the degree that superintendent effects are driven by the reallocation of students across schools, the latter model ought to attenuate my estimates substantially: the inclusion of school fixed effects would control for across-school differences in

[17] To limit the size of the stacked dataset, I define control districts as districts in the same commuting zone and the same state-specific quartile of 2009 proficiency.

productivity.

My results, separately for math and reading subjects, are shown in Figure 8. Results excluding school fixed effects are shown as blue. Consistent with my results using SEDA data, I find that effective superintendents indeed raise district test score performance upon entering a district. Here, effect sizes suggest that a one standard deviation increase in superintendent value-added increases proficiency rates by roughly one percentage point.

Adding school fixed effects has no effect on the magnitude of my estimated treatment effects. These estimates are shown in yellow in Figure 8. Consistent with this result, Panel A of Table 9 shows that difference-in-difference estimates are extremely similar, regardless of whether they are estimated with or without school fixed effects.

In Panel B of Table 9, I test for differential effects based on relative baseline (within-district) performance. To do so, I calculate each school's percentile within the 2009 distribution of proficiency rates within the district. I interact this measure of baseline performance with my treatment indicator. Results in Columns 1 and 3, which exclude school fixed effects, exhibit small, imprecise, negative coefficients. Columns 2 and 4 add school fixed effects, which generates interaction terms that are larger in magnitude and more precise. The estimates suggest schools at the top of the baseline district performance distribution are unaffected by top management. Put differently, management affects lowest-performing schools the most.

Thus, I conclude that superintendents affect student achievement through within school improvements in test scores, rather than from shifts in enrollments across schools.

6.2 Management and District Inputs

Historically, the question of school resources and student outcomes has received substantial attention from researchers (Coleman (1968), Jackson et al. (2015)). More recent work has focused on allocation of these resources across categories of spending (Baron (2022), Brunner et al. (2020)). Naturally, in this context, I ask whether whether effective superintendents differentially change district-level inputs.

My estimates are shown visually in Figure 9. The top three panels of Figure 9 show the effects of superintendent test score value-added on the composition of spending. I show results with respect to the three largest components of district budgets: instruction (which includes teacher salaries and benefits), support services (which includes various staff and non-staff support services, such as guidance counselors, administration, and student transportation), and capital spending (which includes building construction and maintenance). All three series show little evidence of large effects; effect sizes are consistently insignificant and small in magnitude. The corresponding difference-in-difference coefficients are displayed in Table 10, and rule out even small effect sizes of 0.01, or one percent of total spending in response to a one standard deviation change in superintendent value-added.

The bottom two panels estimate the effects of superintendent quality on levels of district resources. I use two measures of district resources: the log of teachers per pupil and the log of total spending per pupil (in 2021 dollars). While neither input appears to move systematically following the hiring of a high- versus low-quality superintendent, I do find small negative effects on the log of teachers per pupil: effects that increase slightly in the first three years following a superintendent's entry. While the estimated effect is statistically significant in Table 10, it is quite small. As in the estimates described above, I can rule out meaningfully large effect sizes of 0.02; approximately two percent.

In Appendix Figure B4, I show the results of a more comprehensive search for superintendent effects on district finances and staffing. Across a range of different outcomes and different measures (spending shares, spending levels, and logs of spending levels), I find little systematic evidence of effects on district resources. What limited evidence I find is consistent with the results above: higher value-added superintendents may induce slightly lower teachers per pupil counts. Consistent with this result, I find some evidence that average teacher salaries increase in response to the arrival of a higher value-added superintendent.

Still, altogether these results suggest that effective superintendents do not appear to substantially change the share of resources allocated to broad types of school inputs nor the level of school spending or staffing. These results are consistent with the "management as technology," model of management effects: given the same inputs, effective superintendents are capable of producing more learning than ineffective superintendents. In the section that follows, I examine whether school practices may explain these effects.

6.3 Management and District Practices

The analyses of district inputs above may fail to identify more subtle differences in the school operations; these differences are often highlighted in management surveys. For example, the World Management Survey's Educational Questionnaire, used in Bloom, Lemos, Sadun and Van Reenen (2015) to quantify management quality across a number of schools in different countries, attempts to measure how much schools reward high performing staff, remove poor performing staff, and "personaliz[e ...] instruction based on student needs," While these questions do not have exact analogues in large-scale survey data, I use CRDC data to quantify staffing and curricular choices.

My results are shown visually in Figure 10. CRDC data is available only in odd-numbered years, so my event study estimates group observations into two-year increments. Panel A of Figure 10 displays effects on teachers experience and absences. Relative to low value-added superintendents, high value-added superintendents increase the share of teachers in their first or second year of teaching. This effect is most pronounced in the first two years following a superintendent transition, suggesting that effective superintendents may increase teacher turnover in the short term. The precision associated with this effect varies slightly depending on weighting; district-weighted estimates are less precise that estimated weighted by total enrollment. Still, magnitudes are reasonably large: a one standard deviation increase in superintendent value-added increases the share of teachers in their first year by nearly one percentage point, relative to a sample mean of approximately 10 percent.

Effects on the rate of chronic teacher absences—defined as absent more than 10 days of the school year—are similarly large. Increases in superintendent value-added decrease the share of teachers who are chronically absent. In the first two years after superintendent entry, this effect size is roughly two percentage points for every one standard deviation increase in superintendent value-added: between five and 10 percent of the sample mean. I report the corresponding difference-in-difference results in Table 11, which comport with the visual evidence in Figure 10. Increases in superintendent value-added increase the share of inexperienced teachers and decrease the rate of teacher absences.

In Panel B of Figure 10, I examine effects on school curricular decisions. Notably, rates of gifted and talented programs increase in the years immediately following the arrival of a higher value-added superintendent. Effects on my two other measures of curricula—indicators for algebra and AP course offerings—show little evidence of short-run effects. (Effects on algebra are positive and statistically significant in later years, but, given these time patterns, it is difficult to disentangle the effect of superintendents directly versus the effect of higher levels of achievement in later years.)

Overall, these results are consistent with a large body of literature that emphasizes personnel management as a key aspect of management (Bloom and Van Reenen (2011)), particularly in education management. Notably, two management interventions that produced large test score gains (Fryer Jr (2014) and Fryer et al. (2017)) both include large-scale changes in personnel or personnel management practices. Fryer Jr (2014) conducts a intervention that “implemented a bundle of best practices" into a set of traditional public schools in Houston, Texas. In treatment schools, “nineteen out of twenty principals were removed and 46 percent of teachers left or were removed before the experiment began." Similarly, Fryer et al. (2017) argues that the positive effects of a high-dosage principal training program were driven in part by increasing the amount of time principals spent on “investing in human capital": “meeting with teachers and school-based staff, reviewing teacher lesson plans, [and] observing classroom instruction," among other activities.

7 Are Superintendents Rewarded for Performance?

Prior to concluding, I briefly examine whether superintendents are rewarded for performance, either via longer tenures or through higher pay. This specific question dates back to Ehrenberg et al. (1988), but researchers in many other domains have examined the determinants of the quality of the public sector workforce, both within educational environments (Nagler et al. (2020)) and elsewhere (Ferraz and Finan (2009), Dal Bó et al. (2017)).

To do so, I test whether superintendents with higher value-added estimates have longer tenures and higher pay. Specifically, I estimate the model below.

$$y_{jd} = \beta_0 + \beta_1 \hat{\mu}_{jd} + \gamma X_{jd} + \varepsilon_{jd}, \tag{15}$$

where y_{jd} is an outcome—either tenure (in years) or the log of annual salary—for superintendent j in district d. $\hat{\mu}_{jd}$ is my estimate of value-added for the same superintendent-district pair. X_{jd} is a set of district level or district-by-year level covariates. β_1 measures the effect of superintendent value-added on superintendent tenure or pay.

This setup also allows me to test for heterogeneity in the degree to which superintendents are rewarded for performance in different types of districts. The amount of local interdistrict competition is a natural candidate for heterogeneity analysis for two reasons.

First, the role of competition in education quality has been a pressing question for researchers for decades (Hoxby (2000), Rothstein (2007)). More recently, advocates for charter schools (and school choice more broadly) argue that increased competition from alternative schools will drive improvements in the management of lower-performing public schools. To the degree that competition affects the decisions of school management, there is perhaps no single discrete decision that draws more of the school board's attention than selection of the school superintendent.

Expanding beyond the education realm, recent literature based on management surveys highlights the importance of competition in raising management quality. For instance, in a review of survey evidence on management practices across firms and countries, Bloom and Van Reenen (2010) conclude that, among other factors, "imperfectly competitive markets [...] allow bad management practices to persist." One case study in health care (Bloom, Propper, Seiler and Van Reenen (2015)) suggests a causal interpretation of this relationship, but other evidence is scarce.

To measure the degree of local competition, I calculate the number of districts within a 10 mile radius of the district's center. I then take the inverse hyperbolic sine of this value, then standardize the measure to have mean 0 and standard deviation 1. I refer to this as the "nearby districts index." Appendix Figure B5 shows how this value varies across districts in my sample. Across most states (with the exception of Georgia and Virginia) urban and suburban districts have more local competition, whereas rural districts have less. In all regressions, I control for state fixed effects interacted with fixed effects for locale type, which takes one of 12 values indicating its locale ("City large", "City midsize", ... "Rural remote"). The bottom panel of Figure B5 shows the variation in my nearby districts index after residualizing on these fixed effects and rescaling. By construction, these patterns do not vary systematically across types of locales.

My regression results are shown in Table 12. In Panel A, I test whether superintendents with higher value-added have longer tenures. In this analysis, I restrict my data to superintendents who were hired prior to 2014, and thus have greater potential for long tenures. Results in Column 1 show that superintendent value-added correlates positively with tenure in this sample, but the estimated effect is small and statistically insignificant. In Column 2, I test for an interaction between district competition and value-added. My results indicate that a one standard deviation increase in the amount of local competition increases tenures of superintendents with one standard deviation higher value-added by 0.06 years. Results in Column 3, which add district fixed effects, are similar directionally, but imprecise; few schools have more than one superintendent with value-added estimates prior to 2014.

This evidence is suggestive, in large part because the value-added estimates included as explanatory variables come from the same district-superintendent combinations as my tenure data. Using salary data—in particular, salary data from early in a superintendent's tenure—provides a cleaner test: first-year salaries reflect a district's belief about the quality of their superintendent, rather than the realized outcomes in subsequent years (which are used to produce estimates of value-added). However, using salary data limits my sample to the three states for which such data exist: Georgia, Pennsylvania, and Wisconsin.

Anecdotally, school boards use salaries to attract candidates who they believe are high-quality. The Dallas Independent School District "lure[d]" Mike Moses from his job as Deputy Chancellor at Texas Tech University with a contract that "probably made him the highest paid superintendent in the country."[18] After the board hires a superintendent, salary is often a focal point of contract negotiations between the superintendent and the school board.[19] Here, I test whether superintendents with higher value-added are rewarded with higher salaries.

Panels B and C show results with respect to superintendent salaries in the year of hiring and the year after. Results in Column 1 suggest that superintendents with higher value-added yield two percent higher salaries in their first and second year in a district. Column 2 shows how this varies according to district competition. Districts with higher levels of competition reward superintendents who are (subsequently) more effective in raising test scores. Effect sizes suggest that the returns to value-added are 1.4 and 1.7 percent higher in districts with one standard deviation more competition.

Finally, in Column 3 I test whether this effect shows up within districts. I do so by including district fixed effects, meaning that my analysis compares the salaries of superintendents with dif-

[18] "Tying the Contract Knot," AASA, The School Superintendents Association.
[19] "Who Should Negotiate Your Contract?" AASA, The School Superintendents Association.

ferent value-added estimates who worked in the same district at different times. These estimates yield results of similar direction and magnitude: higher levels of local competition increase the returns to superintendents with higher value-added.

In Appendix Tables B1 and B2, I perform two robustness checks for these results. First, in Appendix Table B1, I confirm that these results hold under a number or various combinations of fixed effects with respect to entry year, state, district, and locale type. Second, in Table B2, I test whether my results are sensitive to the radius used to calculate the number of nearby districts. Results in Table B2 demonstrate that distance radii of 2, 10, or 20 miles produce similar results.

At a high level, these results are consistent with a model of Tiebout competition between districts. In more competitive areas, districts must compete more vigorously for students, whose parents can more easily "vote with their feet" if a school district fails to meet their expectations. Thus, the potential costs of managerial hiring mistakes—overpaying for a low-quality superintendent or letting an effective superintendent leave—are greater in districts with more local competition. This is the argument laid out in Hoxby (2000), who argues that competition causally increases levels of achievement.

8 Conclusion

In spite of an increasing amount of academic focus on management across a number of domains, causal evidence on the role of management in education remains scarce. In this work, I bring new, large-scale data to bear on the question, and show that changes in school management have significant and persistent effects on student achievement.

I explore potential mechanisms behind explaining superintendent effects, and rule out two simple explanations related to school resources or changes in the composition of school spending. Instead, my results support a "management as technology" interpretation; effective superintendents are capable of increasing achievement without increasing inputs. I find some evidence of changes in the working environment for teachers, supporting this interpretation.

My work highlights a number of important considerations for policy as well as avenues for future research.

First, my results highlight the importance of incorporating dynamics in estimates of value-added in managerial contexts, and suggest that researchers ought to consider these dynamics when estimating value-added in similar contexts. The presence of these dynamics, in combination with the relatively short tenures of many superintendents, suggests that school boards may be wise to scale their evaluations of superintendent quality in proportion to the length of a superintendent's tenure in a district. The estimates in this paper suggest that outcomes in a superintendent's first year are relatively uninformative about their long-term effect on district outcomes.

Second, my heterogeneity analyses point to managerial flexibility as potentially important in amplifying the effect of management in education. Whether such amplification generates efficiency gains through better identification of effective or ineffective managers may be a fruitful question in future work. Regardless, policy interventions aimed at improving management quality are likely to be most effective in settings where managers have relatively more flexibility.

Finally, my results with respect to superintendent tenures and salary suggest one avenue through which competition in education (and perhaps competition more broadly) may contribute to gains in productivity: through a more efficient market for management.

References

Angrist, J. D., Hull, P. D., Pathak, P. A. and Walters, C. R. (2017), 'Leveraging Lotteries for School Value-Added: Testing and Estimation', *The Quarterly Journal of Economics* **132**(2), 871–919.

Baker, A. C., Larcker, D. F. and Wang, C. C. (2022), 'How Much Should We Trust Staggered Difference-in-Differences Estimates?', *Journal of Financial Economics* **144**(2), 370–395.

Baron, E. J. (2022), 'School Spending and Student Outcomes: Evidence from Revenue Limit Elections in Wisconsin', *American Economic Journal: Economic Policy* **14**(1), 1–39.

Bartanen, B. and Husain, A. N. (2022), 'Connected Networks in Principal Value-Added Models', *Economics of Education Review* **90**, 102292.

Bartanen, B., Husain, A. N. and Liebowitz, D. D. (2022), 'Rethinking Principal Effects on Student Outcomes'.

Bertrand, M. and Schoar, A. (2003), 'Managing with Style: The Effect of Managers on Firm Policies', *The Quarterly Journal of Economics* **118**(4), 1169–1208.

Bloom, N., Eifert, B., Mahajan, A., McKenzie, D. and Roberts, J. (2013), 'Does Management Matter? Evidence from India', *The Quarterly Journal of Economics* **128**(1), 1–51.

Bloom, N., Lemos, R., Sadun, R. and Van Reenen, J. (2015), 'Does Management Matter in Schools?', *The Economic Journal* **125**(584), 647–674.

Bloom, N., Propper, C., Seiler, S. and Van Reenen, J. (2015), 'The Impact of Competition on Management Auality: Evidence from Public Hospitals', *The Review of Economic Studies* **82**(2), 457–489.

Bloom, N., Sadun, R. and Van Reenen, J. (2016), Management as a Technology?, Technical report, National Bureau of Economic Research.

Bloom, N. and Van Reenen, J. (2010), 'Why Do Management Practices Differ across Firms and Countries?', *Journal of Economic Perspectives* **24**(1), 203–24.

Bloom, N. and Van Reenen, J. (2011), Human Resource Management and Productivity, *in* 'Handbook of Labor Economics', Vol. 4, Elsevier, pp. 1697–1767.

Brueckner, J. K. and Neumark, D. (2014), 'Beaches, Sunshine, and Public Sector Oay: Theory and Evidence on Amenities and Rent Extraction by Government Workers', *American Economic Journal: Economic Policy* **6**(2), 198–230.

Brunner, E., Hyman, J. and Ju, A. (2020), 'School Finance Reforms, Teachers' Unions, and the Allocation of School Resources', *Review of Economics and Statistics* **102**(3), 473–489.

Card, D. and Krueger, A. B. (1992), 'Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States', *Journal of Political Economy* **100**(1), 1–40.

Cengiz, D., Dube, A., Lindner, A. and Zipperer, B. (2019), 'The Effect of Minimum Wages on Low-Wage Jobs', *The Quarterly Journal of Economics* **134**(3), 1405–1454.

Chetty, R., Friedman, J. N. and Rockoff, J. E. (2014), 'Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates', *American Economic Review* **104**(9), 2593–2632.

Chin, M., Kane, T. J., Kozakowski, W., Schueler, B. E. and Staiger, D. O. (2019), 'School District Reform in Newark: Within-and Bbetween-School Changes in Achievement Growth', *ILR Review* **72**(2), 323–354.

Chingos, M. M., Whitehurst, G. J. and Lindquist, K. M. (2014), 'School Superintendents: Vital or Irrelevant?', *Brown Center on Education Policy at Brookings* pp. 1–16.

Coelli, M. and Green, D. A. (2012), 'Leadership Effects: School Principals and Student Outcomes', *Economics of Education Review* **31**(1), 92–109.

Coleman, J. S. (1968), 'Equality of Educational Opportunity', *Integrated Education* **6**(5), 19–28.

Dal Bó, E., Finan, F., Folke, O., Persson, T. and Rickne, J. (2017), 'Who Becomes a Politician?', *The Quarterly Journal of Economics* **132**(4), 1877–1914.

Diamond, R. (2017), 'Housing Supply Elasticity and Rent Extraction by State and Local Governments', *American Economic Journal: Economic Policy* **9**(1), 74–111.

Ehrenberg, R. G., Chaykowski, R. P. and Ehrenberg, R. A. (1988), 'Are School Superintendents Rewarded for "Performance"?'.

Ferraz, C. and Finan, F. (2009), Motivating Politicians: The Impacts of Monetary Incentives on Quality and Performance, Technical report, National Bureau of Economic Research.

Fryer Jr, R. G. (2014), 'Injecting Charter School Best Practices into Traditional Public Schools: Evidence from Field Experiments', *The Quarterly Journal of Economics* **129**(3), 1355–1407.

Fryer, R. G. et al. (2017), Management and Student Achievement: Evidence from a Randomized Field Experiment, Technical report, National Bureau of Economic Research.

Goodman-Bacon, A. (2021), 'Difference-in-Differences with Variation in Treatment Timing', *Journal of Econometrics* .

Grissom, J. A. and Andersen, S. (2012), 'Why Superintendents Turn Over', *American Educational Research Journal* **49**(6), 1146–1180.

Hanushek, E. A. and Rivkin, S. G. (2010), 'Generalizations about Using Value-Added Measures of Teacher Quality', *American Economic Review* **100**(2), 267–71.

Hoxby, C. M. (1996), 'How Teachers' Unions Affect Education Production', *The Quarterly Journal of Economics* **111**(3), 671–718.

Hoxby, C. M. (2000), 'Does Competition Among Public Schools Benefit Students and Taxpayers?', *American Economic Review* **90**(5), 1209–1238.

Jackson, C. K., Johnson, R. C. and Persico, C. (2015), 'The Effects of School Spending on Educational and Economic Outcomes: Evidence from School Finance Reforms', *The Quarterly Journal of Economics* **131**(1), 157–218.

Jackson, C. K. and Mackevicius, C. (2021), The Distribution of School Spending Impacts, Technical report, National Bureau of Economic Research.

Kane, T. J. and Staiger, D. O. (2008), Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation, Technical report, National Bureau of Economic Research.

Kinsler, J. (2012), 'Beyond Levels and Growth: Estimating Teacher Value-Added and its Persistence', *Journal of Human Resources* **47**(3), 722–753.

Koedel, C. and Rockoff, J. E. (2015), 'Value-Added Modeling: A Review', *Economics of Education Review* **47**, 180–195.

Lavy, V. and Boiko, A. (2017), Management Quality in Public Education: Superintendent Value-Added, Student Outcomes and Mechanisms, Technical report, National Bureau of Economic Research.

Leaver, C., Lemos, R. F. and Scur, D. (2019), 'Measuring and Explaining Management in Schools: New Approaches using Public Data', *World Bank Policy Research Working Paper* (9053).

Loeb, S. and Grissom, J. A. (2013), 'What Do We Know about the Use of Value-Added Measures for Principal Evaluation? What We Know Series: Value-Added Methods and Applications. Knowledge Brief 9.', *Carnegie Foundation for the Advancement of Teaching* .

Morris, C. N. (1983), 'Parametric Empirical Bayes Inference: Theory and Applications', *Journal of the American Statistical Association* **78**(381), 47–55.

Mundlak, Y. (1961), 'Empirical Production Function Free of Management Bias', *Journal of Farm Economics* **43**(1), 44–56.

Nagler, M., Piopiunik, M. and West, M. R. (2020), 'Weak markets, strong teachers: Recession at career start and teacher effectiveness', *Journal of Labor Economics* **38**(2), 453–500.

Reardon, S. F., Fahle, E. M., Chavez, B., Kalogrides, D., Shear, B. R. and Ho, A. D. (2021), 'Stanford Education Data Archive (Version 4.1)'.

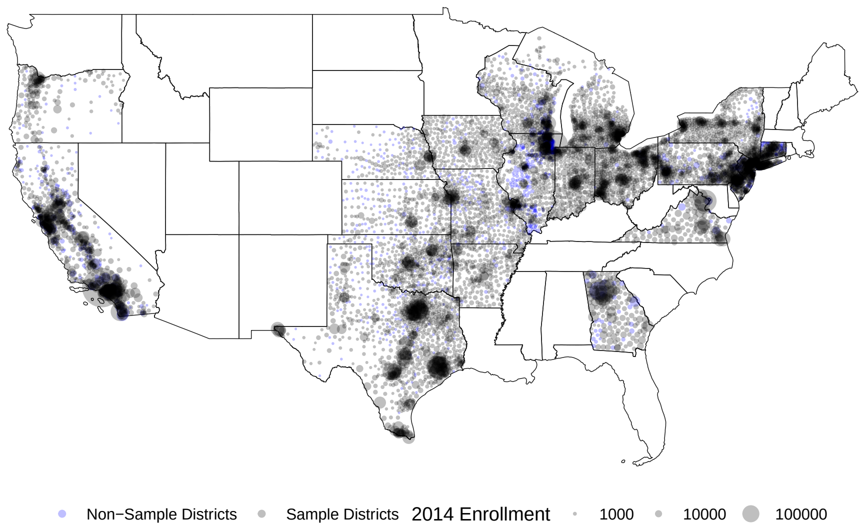
Reback, R., Rockoff, J. and Schwartz, H. L. (2014), 'Under Pressure: Job Security, Resource Allocation, and Productivity in Schools Under No Child Left Behind', *American Economic Journal: Economic Policy* **6**(3), 207–41.

Rothstein, J. (2007), 'Does Competition Among Public Schools Benefit Students and Taxpayers? Comment', *American Economic Review* **97**(5), 2026–2037.

Schuh, J. and Herrington, C. (1990), 'Electing versus Appointing District Superintendents of Schools.'.

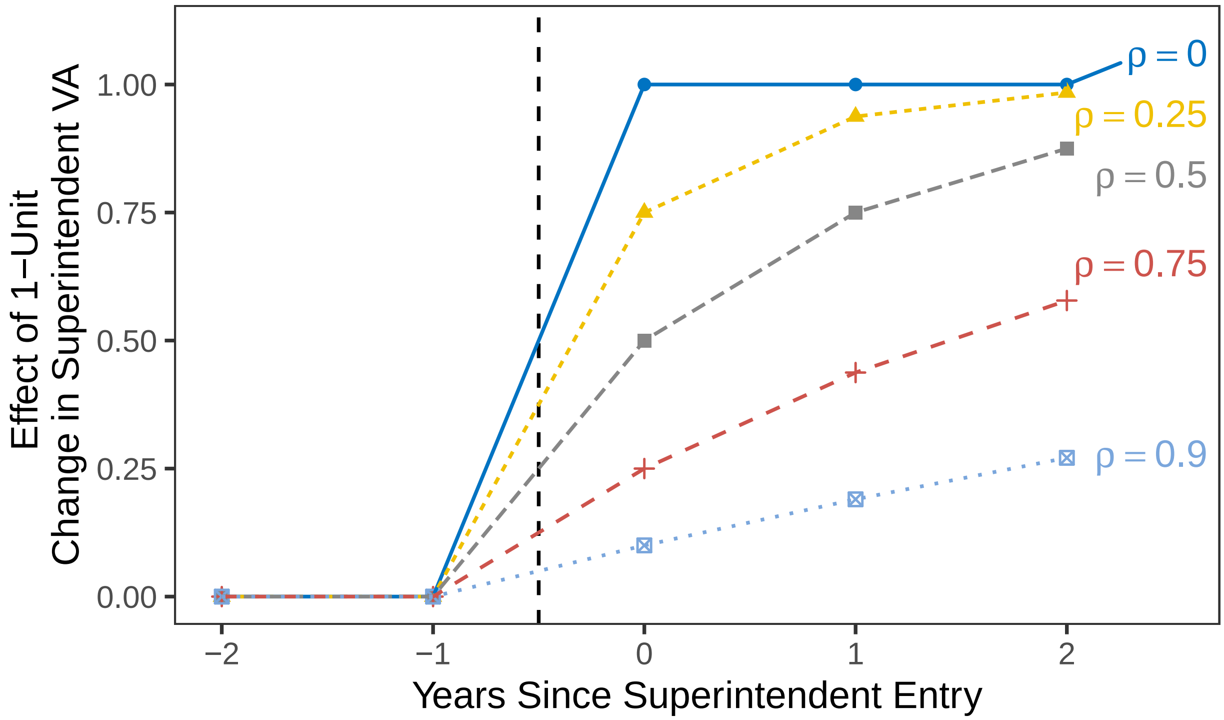
Figures and Tables

Figure 1: Location of Districts in Superintendent Roster Data



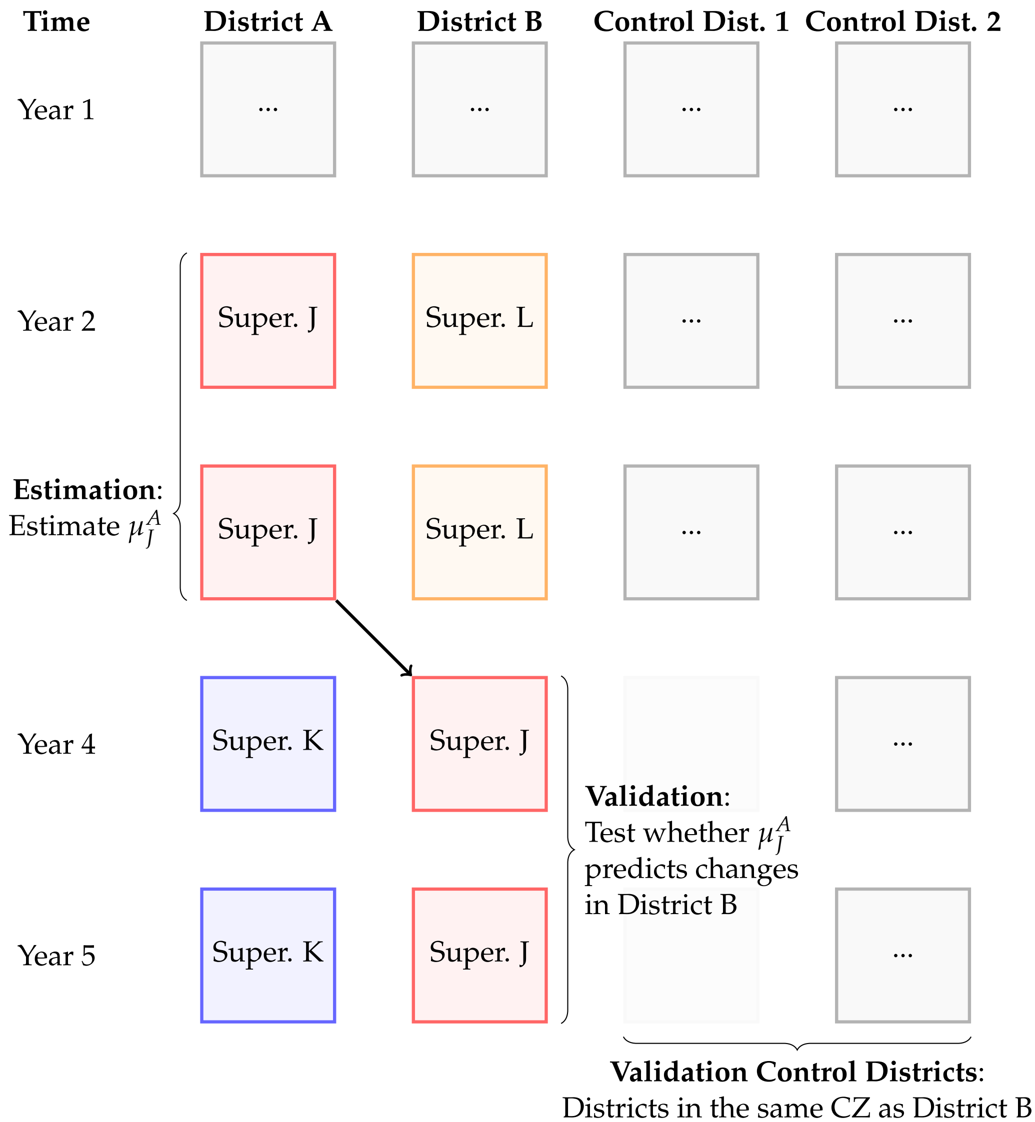
Source: Superintendent Roster Data; NCES Common Core of Data
Notes: Figure displays the location of sample and non-sample districts in states that appear in superintendent roster data. Sample districts are displayed in black and non-sample districts are displayed in blue. Sizes are proportional to 2014 enrollments.

Figure 2: Effect of ρ on District Outcomes



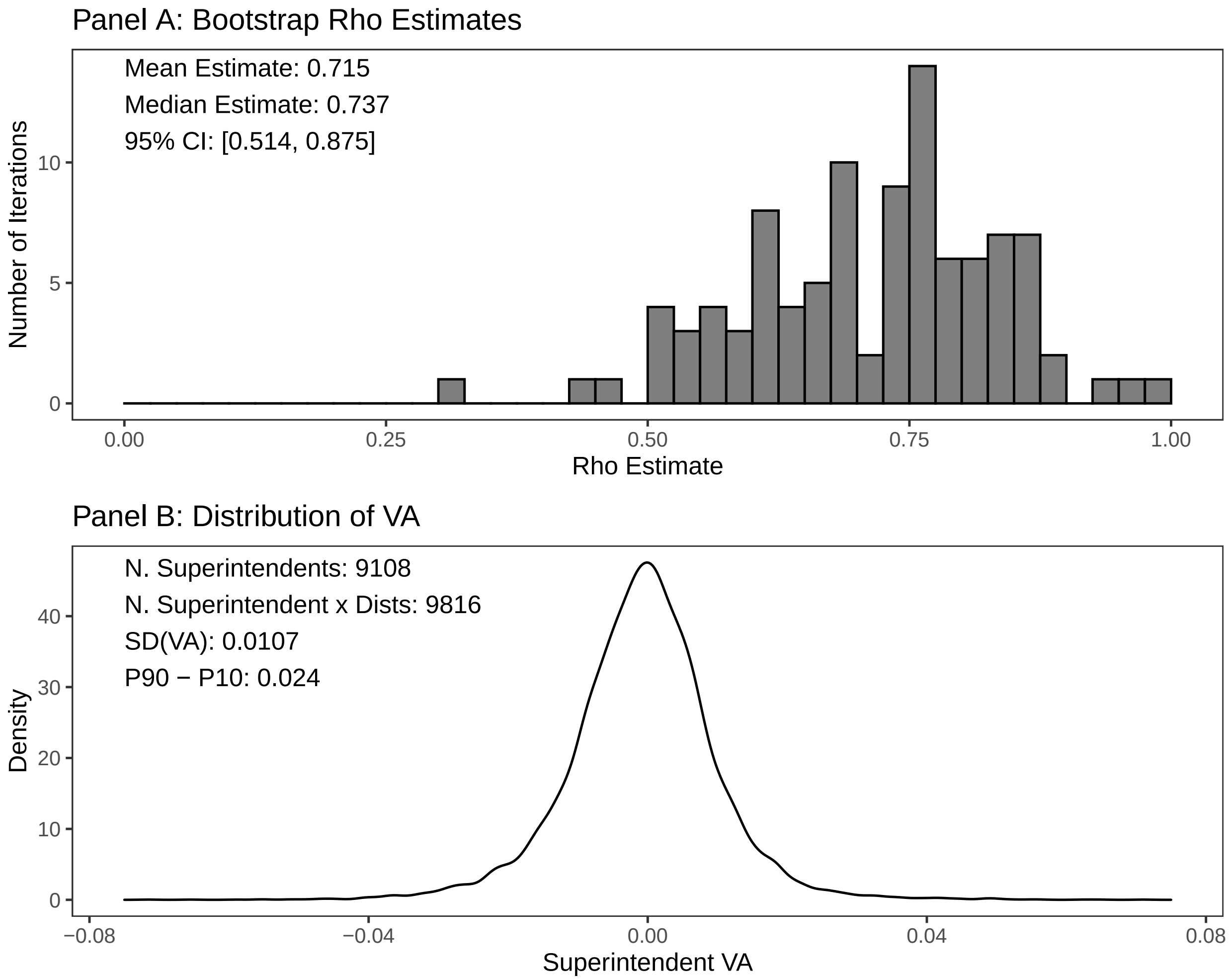
Source: Author Calculations
Notes: Figure displays the predicted effect of a sustained one unit increase in superintendent value-added under different values of ρ.

Figure 3: Illustration of Estimation and Validation Approach



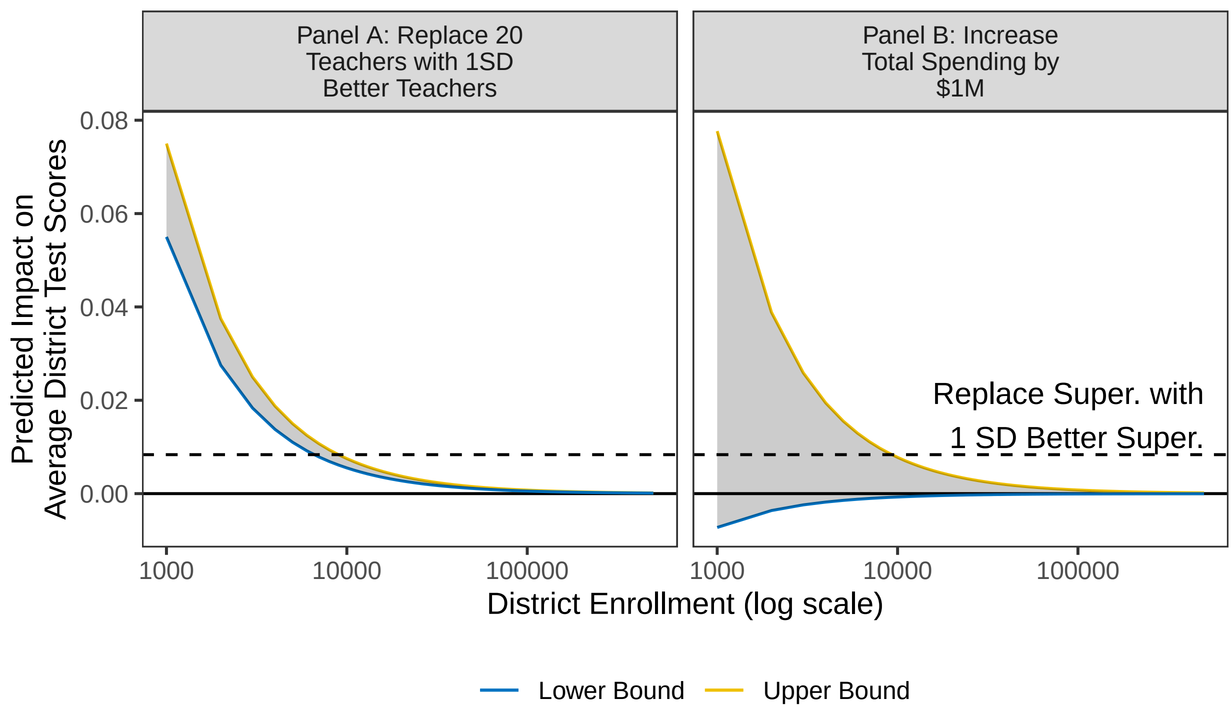
Notes: Figure illustrates the approach used to estimate and validate superintendent value-added.

Figure 4: Distribution of $\hat{\rho}$ and Superintendent Value-Added Estimates



Source: Superintendent Roster Data; SEDA Data
Notes: Figure summarizes the distribution of estimated superintendent value-added parameters. Panel A displays the bootstrap distribution of ρ, based on the bootstrapping procedure described in text. Panel B summarizes the distribution of estimated superintendent value-added.

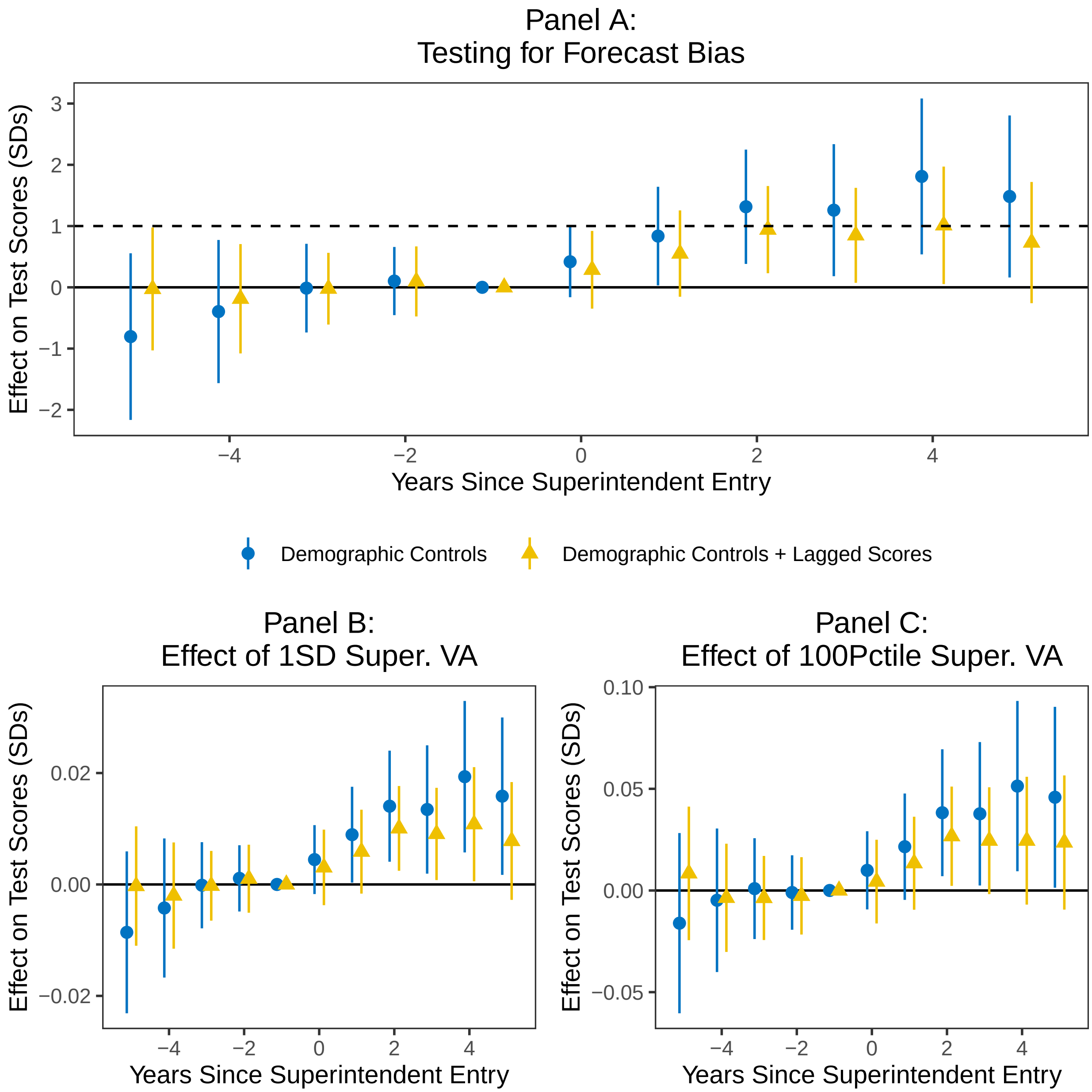
Figure 5: Policy Effect Size Comparisons



Source: Hanushek and Rivkin (2010); Jackson and Mackevicius (2021); Superintendent Roster Data; SEDA Data

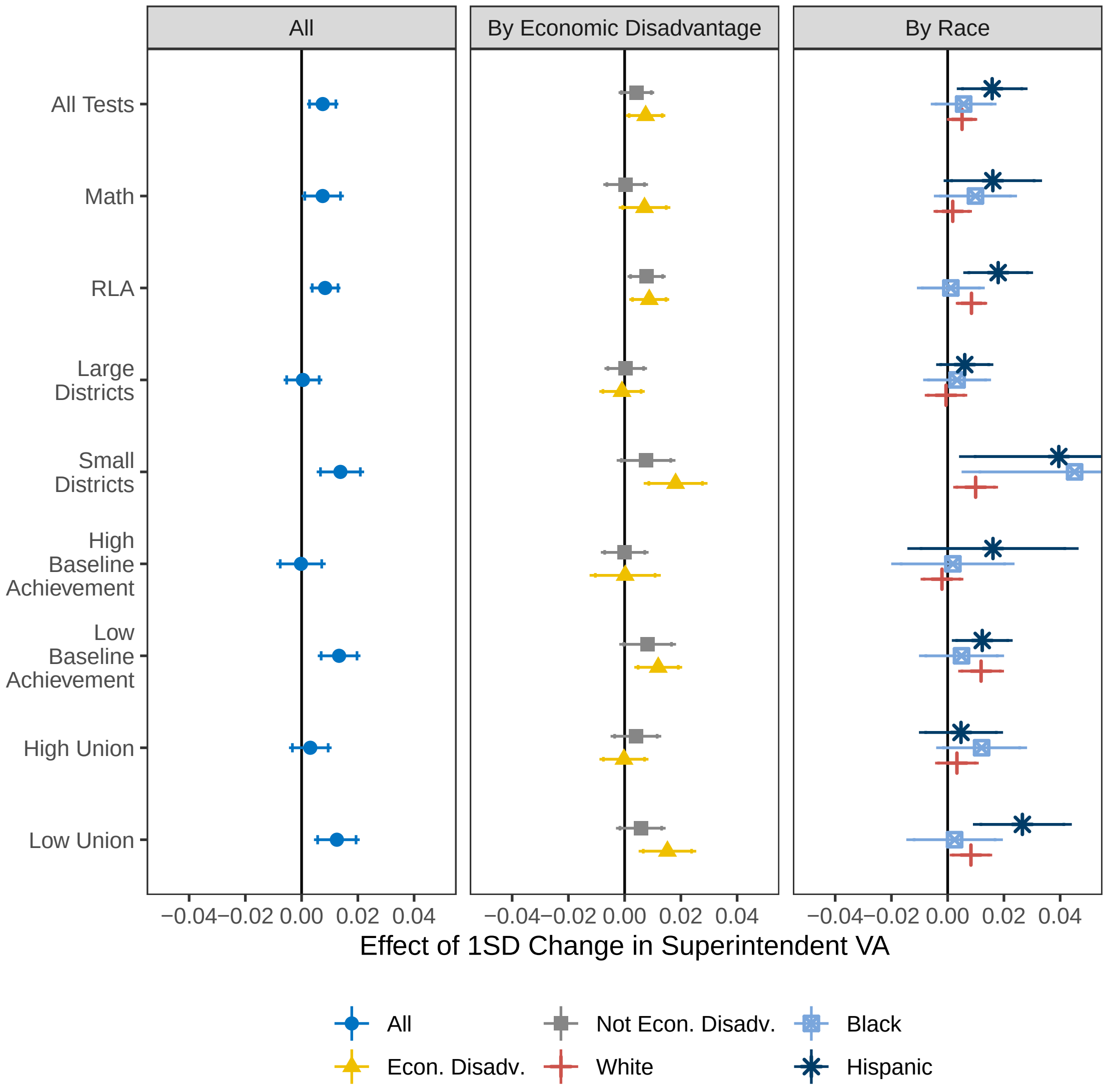
Notes: Figure compares the effects of hypothetical policy changes on average district test scores as a function of district enrollment. In Panel A displays the predicted effect of replacing 20 teachers with teachers with one standard deviation higher value added. Lower and upper bound estimates of the standard deviation of teacher value added are 0.11 and 0.15, based on estimates in Hanushek and Rivkin (2010). Panel B displays the predicted effect of a $1 million spending increase. Upper and lower bounds of the effect of $1,000 in district spending per pupil are -0.00725 to 0.07767, based on estimates in Jackson and Mackevicius (2021). The dashed horizontal line displays the predicted effect of a one standard deviation increase in superintendent value-added, based on estimates described in text.

Figure 6: Evolution of District Test Scores Surrounding Superintendent Entry



Source: Superintendent Roster Data; SEDA Data
Notes: Figure displays event study estimates of the effect of (out-of-district) superintendent value-added estimates on test scores. Displayed coefficients correspond to β_k in Equation 13 and are produced using the stacked difference-in-difference procedure described in text. Panel A uses raw value-added estimates to estimate dynamic effects, where forecast-unbiased estimates will converge to 1. Panel B uses value-added estimates scaled to be mean zero and standard deviation one to estimate dynamic effects. Panel C uses percentiles of value-added estimates to estimate dynamic effects. Units are student test score standard deviations. Standard errors are clustered by district. Bands display 95 percent confidence intervals.

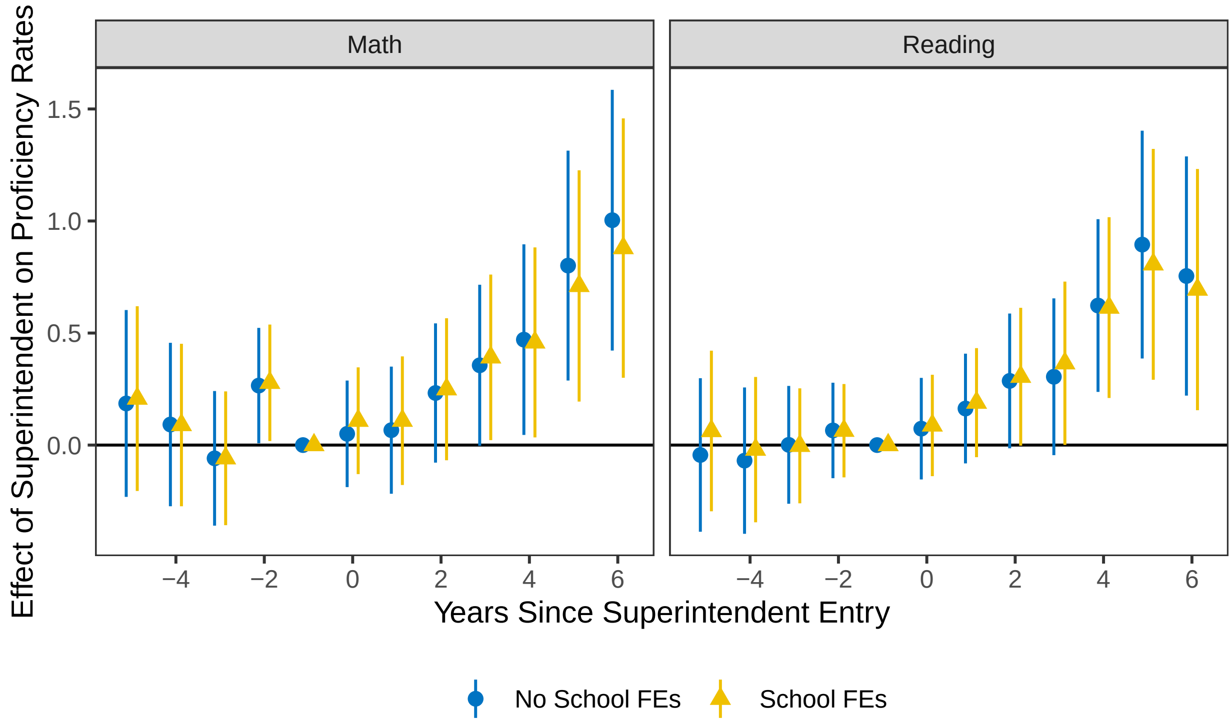
Figure 7: Heterogeneous Effects of Superintendent Value-Added



Source: Superintendent Roster Data; SEDA Data

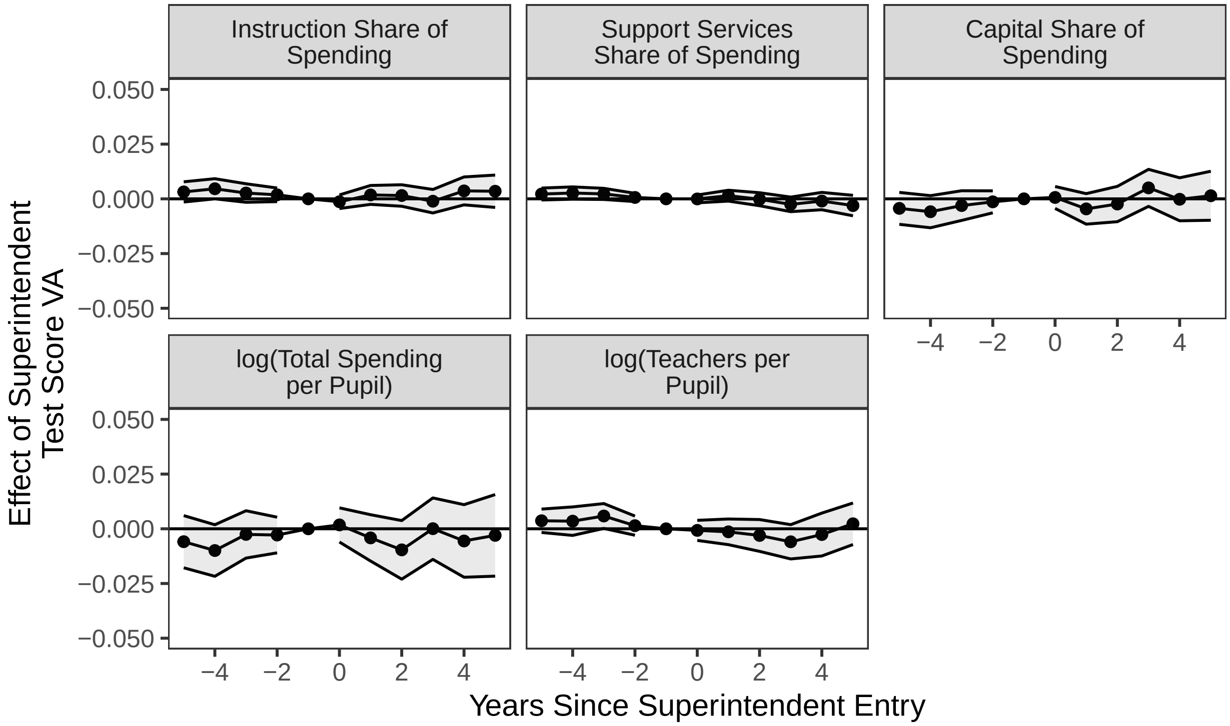
Notes: Figure displays difference-in-difference estimates of the effect of (out-of-district) superintendent value-added estimates on test scores. Displayed coefficients correspond to β in Equation 14 and are produced using the stacked difference-in-difference procedure described in text. Rows indicate subsets of districts in test score data and point estimates reflect estimates for different student subgroups within these districts. Large districts are defined as districts with above-median average district-by-grade enrollment (122) in test score data. District with high baseline achievement are defined as districts with 2009 average district scores above nationally-normed means. High union districts are identified as districts in the 10 sample states with the highest union strength index in Brunner et al. (2020). These states are (in order of highest to lowest union power): Oregon, Pennsylvania, California, New Jersey, Illinois, New York, Ohio, Michigan, Connecticut, and Wisconsin. The remaining (low union power) sample states are, in order: Nebraska, Iowa, Indiana, Kansas, Missouri, Oklahoma, Texas, Virginia, Arkansas, and Georgia. Standard errors are clustered by district. Horizontal bands and vertical ticks display 95 and 90 percent confidence intervals, respectively.

Figure 8: School-Level Effects of Superintendent Value-Added



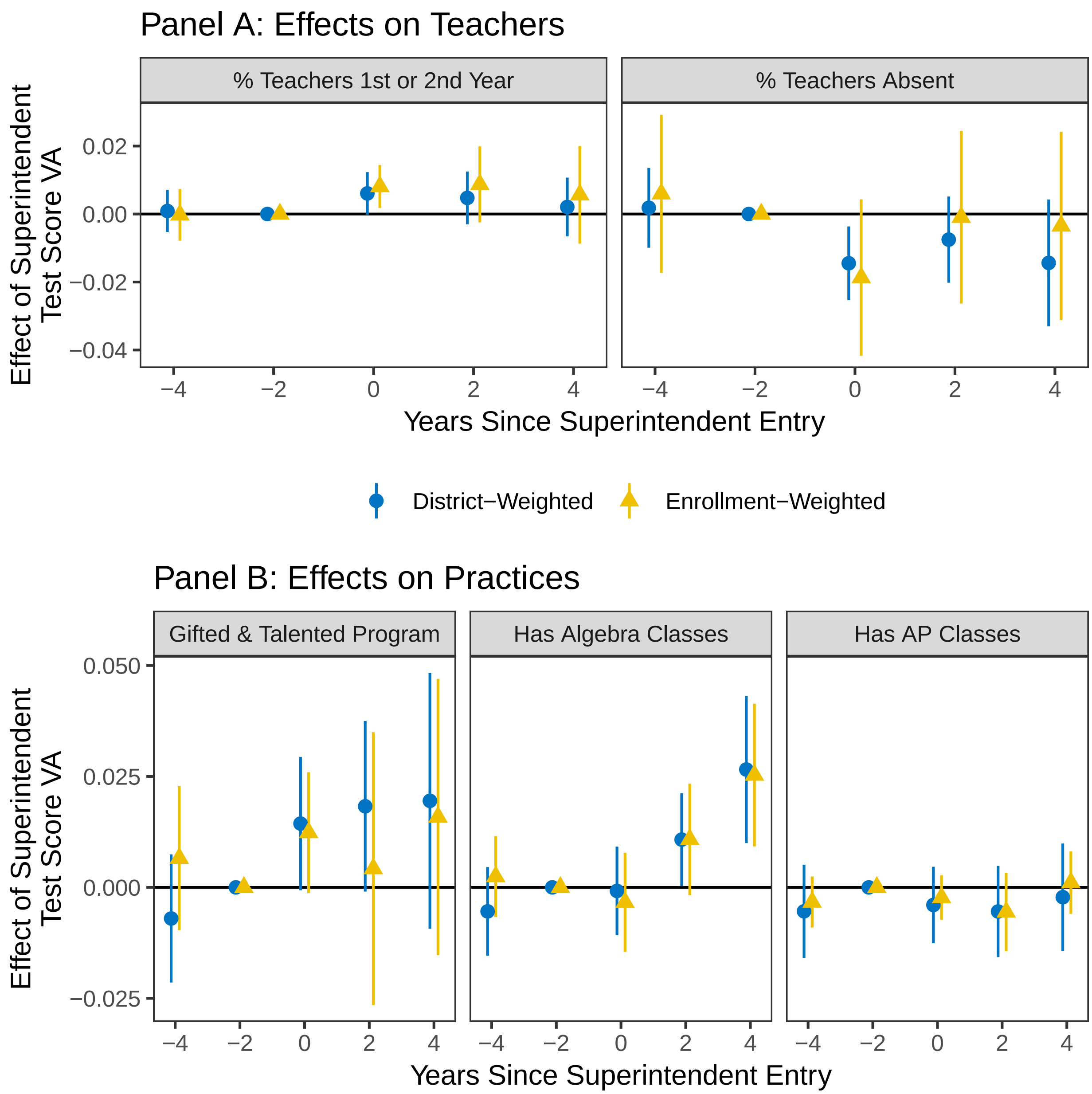
Source: Superintendent Roster Data; SEDA Data; EdFacts Test Score Data
Notes: Figure displays event study estimates of the effect of (out-of-district) superintendent value-added estimates on school level proficiency rates, with and without school fixed effects. Superintendent value-added is scaled to be mean 0 and standard deviation 1. Standard errors are clustered by district. Bands display 95 percent confidence intervals.

Figure 9: Dynamic Effect of Superintendent VA on District Inputs



Source: Superintendent Roster Data; SEDA Data; NCES Common Core of Data
Notes: Figure displays event study estimates of the effect of (out-of-district) superintendent value-added estimates on district inputs. Displayed coefficients correspond to β_k in Equation 13 and are produced using the stacked difference-in-difference procedure described in text. Superintendent value-added is scaled to be mean 0 and standard deviation 1. Total spending per pupil is in 2021 dollars. Standard errors are clustered by district. Bands display 95 percent confidence intervals.

Figure 10: Dynamic Effect of Superintendent VA on District Practices



Source: Superintendent Roster Data; SEDA Data; CRDC Data

Notes: Figure displays event study estimates of the effect of (out-of-district) superintendent value-added estimates on district practices. Displayed coefficients correspond to β_k in Equation 13 and are produced using the stacked difference-in-difference procedure described in text. Superintendent value-added is scaled to be mean 0 and standard deviation 1. District-weighted estimates weight each school level observation by it proportion of district enrollments in each year. Enrollment-weighted estimates weight each school level observation by total enrollment. Chronically absent teachers are teachers who are absent more than 10 days of the school year. Standard errors are clustered by district. Bands display 95 percent confidence intervals.

Table 1: State Superintendent Coverage in 2014

State	Enrollments		Districts	
	Total (000s)	Covered (%)	Total	Covered (%)
Arkansas	490.9	95.5	254	88.6
California	6235.5	88.5	923	80.2
Connecticut	542.6	89.8	193	64.8
Georgia	1744.4	86.4	199	75.9
Illinois	2050.2	71	837	59
Indiana	1046.2	95.8	375	76.5
Iowa	505.3	98.9	335	94.6
Kansas	497.3	96.9	282	84.8
Michigan	1537.9	83.8	831	59.8
Missouri	919	93.1	521	79.5
Nebraska	312.6	96.3	240	78.3
New Jersey	1400.5	94.7	602	78.1
New York	2741.2	58.4	949	68
Ohio	1724.8	91.7	915	65.9
Oklahoma	688.5	95.5	509	83.9
Oregon	585.5	96.8	173	89
Pennsylvania	1743.2	84.8	701	65
Texas	5233.8	95.3	1191	75.9
Virginia	1280.4	96.3	133	93.2
Wisconsin	871.4	90	440	81.1
United States	50631	55.5	16644	47

Source: Superintendent Roster Data; NCES Common Core of Data
Notes: Table summarizes the coverage of superintendent roster data. All data as of 2014. "Covered" districts are those for which I observe a superintendent in 2014. "Total" columns reflect data for all districts with positive enrollment in 2014. Appendix Figure C1 displays counts by state and year.

Table 2: District Summary Statistics

	District-Weighted		Enrollment-Weighted	
	All Dists.	Sample Dists.	All Dists.	Sample Dists.
Panel A: District Characteristics Data (as of 2014)				
Enrollment	3041.99	3592.18	46877.02	42691.79
	(11547.9)	(11852.03)	(99956.14)	(108435.15)
Number of Schools	6.06	6.73	77.48	63.64
	(20.23)	(17.63)	(190.53)	(168.27)
Total Staff	349.2	406.7	5029.89	4489.62
	(1279.91)	(1282.04)	(10495.4)	(10994.29)
Total Teachers	173.53	196.38	2617.82	2208.51
	(672.04)	(633.33)	(5432.94)	(5140.94)
Unique Districts	16644	7818	16644	7818
Panel B: Test Scores Data (as of 2014)				
Math	0.479	0.469	0.473	0.458
	(0.5)	(0.499)	(0.499)	(0.498)
Grade	5.989	5.999	6.022	6.028
	(1.344)	(1.334)	(1.347)	(1.329)
Mean Score	-0.002	0.012	-0.01	-0.012
	(0.388)	(0.383)	(0.369)	(0.399)
% Asian	0.019	0.021	0.048	0.056
	(0.048)	(0.052)	(0.081)	(0.087)
% Black	0.089	0.066	0.162	0.133
	(0.178)	(0.141)	(0.196)	(0.178)
% Hispanic	0.145	0.159	0.254	0.299
	(0.213)	(0.226)	(0.256)	(0.281)
% Reduced Lunch	0.083	0.083	0.073	0.073
	(0.048)	(0.039)	(0.042)	(0.034)
% Free Lunch	0.44	0.43	0.469	0.468
	(0.208)	(0.203)	(0.212)	(0.228)
No. Students Tested	345.641	296.972	3259.034	2934.033
	(1003.496)	(884.958)	(6504.892)	(7095.524)
Unique Districts	9116	6530	9116	6530
Panel C: Finance Data (as of 2014)				
Total Spending PP	16752.625	16422.296	15424.189	15690.677
	(8721.248)	(7376.872)	(5992.726)	(5660.998)
Capital Share of Spending	0.073	0.071	0.09	0.078
	(0.096)	(0.087)	(0.083)	(0.074)
Inst. Share of Spending	0.519	0.523	0.517	0.52
	(0.086)	(0.077)	(0.071)	(0.069)
Sup. Services Share of Spending	0.308	0.302	0.297	0.296
	(0.066)	(0.053)	(0.049)	(0.044)
Unique Districts	13333	7776	13333	7776

Source: Superintendent Roster Data; NCES Common Core of Data; SEDA Data
Notes: Table displays summary statistics as of 2014 for all districts in the US as well as districts that appear in superintendent roster data.

Table 3: Average District Practices by Year (Sample Districts)

	2009	2011	2013	2015	2017
Panel A: Primary Schools					
Has Algebra Classes	0.061	0.090	0.084	0.083	0.060
Gifted & Talented Program	0.728	0.672	0.661	0.646	0.654
Has AP Classes	0.007	0.013	0.014	0.016	0.006
% Teachers Absent	0.349	0.280	0.264	0.277	0.298
% Teachers 1st or 2nd Year	0.087	0.079	0.103	0.117	0.107
Unique Schools	24,140	30,665	30,246	30,752	30,029
Unique Districts	3,813	7,701	7,718	7,771	7,798
Panel B: Middle Schools					
Has Algebra Classes	0.772	0.761	0.740	0.720	0.748
Gifted & Talented Program	0.801	0.746	0.741	0.747	0.742
Has AP Classes	0.002	0.002	0.002	0.001	0.003
% Teachers Absent	0.359	0.286	0.267	0.280	0.301
% Teachers 1st or 2nd Year	0.098	0.085	0.106	0.119	0.112
Unique Schools	6,707	9,359	9,281	9,301	9,202
Unique Districts	3,066	5,360	5,324	5,325	5,395
Panel C: High Schools					
Has Algebra Classes	0.937	0.948	0.955	0.945	0.949
Gifted & Talented Program	0.607	0.579	0.584	0.598	0.604
Has AP Classes	0.723	0.682	0.679	0.682	0.670
% Teachers Absent	0.312	0.257	0.238	0.252	0.269
% Teachers 1st or 2nd Year	0.097	0.086	0.103	0.109	0.103
Unique Schools	6,087	9,497	9,468	9,537	10,157
Unique Districts	3,344	6,278	6,236	6,249	6,534

Source: Superintendent Roster Data; CRDC Data
Notes: Table displays averages of school practices among schools in districts that appear in superintendent roster data. Chronically absent teachers are teachers who are absent more than 10 days of the school year.

Table 4: Superintendent Turnover

Statistic	N	Mean	St. Dev.	Min	Median	Max
Panel A: District-Level						
Superintendent Tenures	7,822	2.747	1.486	1	2	11
Panel B: Superintendent-Level						
Unique Districts	18,734	1.131	0.385	1	1	5
Fraction > 1 District	18,734	0.117	0.321	0	0	1
Panel C: Superintendent-Level (Superintendents Employed in 2008)						
Unique Districts	4,900	1.205	0.479	1	1	5
Fraction > 1 District	4,900	0.176	0.381	0	0	1
Panel D: Superintendent-by-District Level						
Tenure Duration (Years)	21,197	3.48	2.51	1	3	11
Mean Salary	2,604	159,166.10	48,067.02	17,393.72	151,327.20	584,945.80

Source: Superintendent Roster Data
Notes: Table summarizes superintendent turnover in my superintendent roster data. Data is restricted to years 2008 to 2018. Unique superintendents are identified based on the record linking algorithm described briefly in text and in detail in Appendix D. Superintendent salaries are in 2021 dollars.

Table 5: Characterizing Superintendent Transitions

	Exiting District	Entering District	Difference	N
log(Enrollment)	7.187	7.476	0.289^{***}	2080
	(1.085)	(1.141)	(0.023)	
Enrollment	2711.896	3752.272	1040.376^{***}	2080
	(5431.689)	(7293.537)	(139.636)	
No. of Schools	5.592	6.910	1.318^{***}	2080
	(7.991)	(9.696)	(0.187)	
% URM Students	0.204	0.226	0.022^{***}	2080
	(0.253)	(0.263)	(0.005)	
Mean Test Score	0.007	0.024	0.017^{**}	1571
	(0.309)	(0.342)	(0.008)	
Mean Cohort Change in Test Score	0.001	0.000	-0.001	1571
	(0.114)	(0.109)	(0.004)	

Source: Superintendent Roster Data; NCES Common Core of Data; SEDA Data
Notes: Table compares characteristics of exiting and entering districts involved in superintendent transfers. Difference column displays the coefficients and standard errors from a linear regression on a stacked dataset in which each transition appears in two rows, each row representing either the exiting or entering district. Coefficients reflect estimated differences between exiting districts and entering districts after controlling for transition fixed effects and standard errors are clustered by district. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 6: Superintendent Tenure and Changes in School Practices

	(1)	(2)	(3)
Panel A: $\lvert Algebra_{i,t} - Algebra_{i,t-2}\rvert$			
Superintendent Tenure	-0.00168^{***}	-0.00128^{**}	-0.00129^{**}
	(0.00048)	(0.00054)	(0.00054)
Observations	162,961	162,961	162,961
R^2	0.08392	0.31662	0.53771
Panel B: $\lvert AP\ Classes_{i,t} - AP\ Classes_{i,t-2}\rvert$			
Superintendent Tenure	-0.00052	-0.00006	-0.00010
	(0.00032)	(0.00041)	(0.00041)
Observations	162,961	162,961	162,961
R^2	0.10261	0.29246	0.54513
Panel C: $\lvert Gifted\ \&\ Talented_{i,t} - Gifted\ \&\ Talented_{i,t-2}\rvert$			
Superintendent Tenure	-0.00150^{**}	-0.00120	-0.00124
	(0.00063)	(0.00083)	(0.00084)
Observations	162,961	162,961	162,961
R^2	0.05558	0.30232	0.49316
Panel D: $\lvert \%\ Absent\ Teachers_{i,t} - \%\ Absent\ Teachers_{i,t-2}\rvert$			
Superintendent Tenure	-0.00174^{***}	-0.00089^{*}	-0.00096^{**}
	(0.00037)	(0.00048)	(0.00048)
Observations	162,961	162,961	162,961
R^2	0.03299	0.32794	0.47425
Panel E: % Inexperienced Teachers			
Superintendent Tenure	-0.00155^{***}	-0.00035	-0.00041
	(0.00024)	(0.00026)	(0.00026)
Observations	217,896	217,896	217,896
R^2	0.06417	0.30078	0.42869
State-by-Year-by-Level FEs	Y	Y	Y
District FEs	N	Y	Y
School FEs	N	N	Y

Source: Superintendent Roster Data; CRDC Data
Notes: Table displays estimates of the relationship between superintendent tenure and the likelihood of school practices changing between consecutive CRDC surveys. Displayed coefficients are β_1 in Equation 1. Chronically absent teachers are teachers who are absent more than 10 days of the school year. Standard errors are clustered by district. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 7: VA Validation

	Assuming $\mu^{+} = 0$			Assuming $\mu^{+} = \mu_j$		
	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta\theta_{dt}$	1.375*	1.413*	1.027*	2.056***	1.940***	1.329***
	(0.824)	(0.800)	(0.545)	(0.719)	(0.702)	(0.456)
Lagged Score			0.462***			0.462***
			(0.006)			(0.006)
p-value: $\Delta\theta_{dt} = 0$	0.095	0.077	0.059	0.004	0.006	0.004
p-value: $\Delta\theta_{dt} = 1$	0.649	0.605	0.960	0.142	0.180	0.471
Base Controls	N	Y	Y	N	Y	Y
Lagged Score Controls	N	N	Y	N	N	Y
No. Entry Events	1497	1497	1497	1497	1497	1497
Observations	108,585	108,585	108,585	108,585	108,585	108,585
R^2	0.714	0.717	0.778	0.715	0.717	0.778

Source: Superintendent Roster Data; SEDA Data
Notes: Table displays estimates validating superintendent value-added using Equation 12. $\Delta\hat{\theta}$ is the predicted change in test scores (based on out-of-district value-added estimates). μ^{+} denotes the assumed value-added of superintendents in the years following a superintendent's exit. Standard errors are clustered by district. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 8: VA Difference-in-Differences Estimates (Discrete VA)

	(1)	(2)	(3)
Post	-0.021^{**}	-0.020^{**}	-0.014^{***}
	(0.008)	(0.008)	(0.005)
VA Quartile 2 × Post	0.012	0.013	0.011
	(0.011)	(0.011)	(0.007)
VA Quartile 3 × Post	0.018^{*}	0.017	0.009
	(0.010)	(0.010)	(0.007)
VA Quartile 4 × Post	0.024^{**}	0.024^{**}	0.015^{**}
	(0.011)	(0.010)	(0.007)
Lagged Score			0.494^{***}
			(0.004)
Base Controls	N	Y	Y
Lagged Score Controls	N	N	Y
No. Entry Events	1497	1497	1497
Observations	5,426,782	5,426,782	5,426,782

Source: Superintendent Roster Data; SEDA Data

Notes: Table displays difference-in-differences estimates of the effect of (out-of-district) superintendent value-added estimates on test scores. Discrete groups reflect quartiles of the distribution of superintendent value added. Displayed coefficients are produced using the stacked difference-in-difference procedure described in text. Units are student test score standard deviations. Standard errors are clustered by district. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 9: School Fixed Effects Validation Estimates

	Math % Proficient		Reading % Proficient	
Panel A: Baseline Estimates				
	(1)	(2)	(3)	(4)
Post × VA	0.337***	0.319**	0.182	0.190
	(0.129)	(0.133)	(0.153)	(0.154)
District FEs	Y	Y	Y	Y
School FEs	N	Y	N	Y
Observations	1,267,522	1,267,522	1,267,522	1,267,522
R^2	0.820	0.903	0.840	0.908
Panel B: Heterogeneous Estimates				
	(1)	(2)	(3)	(4)
Post × VA	0.493**	0.873***	0.420*	0.532**
	(0.222)	(0.201)	(0.220)	(0.216)
Post × VA × Base Pctile	−0.174	−0.925***	−0.476	−0.739*
	(0.408)	(0.347)	(0.411)	(0.401)
District FEs	Y	Y	Y	Y
School FEs	N	Y	N	Y
Observations	1,147,185	1,147,185	1,147,185	1,147,185
R^2	0.830	0.914	0.841	0.911

Source: Superintendent Roster Data; SEDA Data; EdFacts Test Score Data
Notes: Table displays difference-in-differences estimates of the effect of (out-of-district) superintendent value-added estimates on school level proficiency rates, with and without school fixed effects. Superintendent value-added is scaled to be mean 0 and standard deviation 1. “Base Pctile" refers to the within-district percentile of each school in 2009, ranging from 0 to 1. Standard errors are clustered by district. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 10: Effect of Superintendent VA on District Inputs

	Share of Spending on			log(Total	log(
	Inst.	Sup. Services	Capital	Spend. PP)	Teachers PP)
	(1)	(2)	(3)	(4)	(5)
VA × Post	−0.001	−0.002*	0.001	−0.0003	−0.005*
	(0.002)	(0.001)	(0.002)	(0.004)	(0.003)
No. Entry Events	1497	1497	1497	1497	1497
Observations	1,394,207	1,394,207	1,394,207	1,385,173	1,385,207
R^2	0.588	0.652	0.237	0.855	0.842

Source: Superintendent Roster Data; SEDA Data; NCES Common Core of Data
Notes: Table displays difference-in-differences estimates of the effect of (out-of-district) superintendent value-added estimates on district inputs. Displayed coefficients correspond to β in Equation 14 and are produced using the stacked difference-in-difference procedure described in text. Superintendent value-added is scaled to be mean 0 and standard deviation 1. Total spending per pupil is in 2021 dollars. Standard errors are clustered by district. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 11: Effect of Superintendent VA on District Practices

	Share of Teachers:		School Has:		
	1st/2nd Year	Absent	Gift./Tal.	AP Class	Algebra Class
	District-Weighted Estimates				
VA × Post	0.004	-0.014^{***}	0.016^{**}	-0.001	0.008^{*}
	(0.003)	(0.004)	(0.007)	(0.004)	(0.004)
No. Entry Events	1389	1389	1389	1389	1389
Observations	2,643,858	2,643,858	2,643,858	2,643,858	2,643,858
R^2	0.609	0.543	0.799	0.935	0.880
	Enrollment-Weighted Estimates				
VA × Post	0.006^{*}	-0.017^{**}	0.004	0.001	0.006
	(0.004)	(0.008)	(0.009)	(0.003)	(0.006)
No. Entry Events	1389	1389	1389	1389	1389
Observations	2,643,858	2,643,858	2,643,858	2,643,858	2,643,858
R^2	0.543	0.453	0.783	0.968	0.910

Source: Superintendent Roster Data; SEDA Data; CRDC Data
Notes: Table displays difference-in-differences estimates of the effect of (out-of-district) superintendent value-added estimates on district practices. Displayed coefficients correspond to β in Equation 14 and are produced using the stacked difference-in-difference procedure described in text. Superintendent value-added is scaled to be mean 0 and standard deviation 1. District-weighted estimates weight each school level observation by it proportion of district enrollments in each year. Enrollment-weighted estimates weight each school level observation by total enrollment. Chronically absent teachers are teachers who are absent more than 10 days of the school year. Standard errors are clustered by district. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 12: Superintendent Careers, VA, and District Characteristics

	(1)	(2)	(3)
Panel A: Tenure and Value-Added			
Sample: All Superintendents Hired Pre-2014			
DV: Tenure			
VA	0.042 (0.034)	0.049 (0.034)	0.013 (0.056)
Nearby Districts Index		0.038 (0.056)	(0.000)
Nearby Districts Index × VA		0.059* (0.033)	0.044 (0.057)
State-by-Entry Year-by-Loc. Type FEs	Y	Y	Y
District FEs	N	N	Y
Observations	6,825	6,825	6,825
R^2	0.172	0.173	0.943
Panel B: Year-0 Pay and Value-Added			
Sample: All Superintendents with Salary Data			
DV: log(Salary) in Year 0			
VA	0.021** (0.008)	0.021*** (0.008)	−0.001 (0.011)
Nearby Districts Index		0.034** (0.015)	(0.000)
Nearby Districts Index × VA		0.014* (0.008)	0.002 (0.010)
State-by-Entry Year-by-Loc. Type FEs	Y	Y	Y
District FEs	N	N	Y
Observations	1,253	1,253	1,253
R^2	0.504	0.509	0.868
Panel C: Year-1 Pay and Value-Added			
Sample: All Superintendents with Salary Data			
DV: log(Salary) in Year 1			
VA	0.020** (0.008)	0.021*** (0.008)	−0.012 (0.014)
Nearby Districts Index		0.034** (0.014)	(0.000)
Nearby Districts Index × VA		0.017* (0.009)	0.013 (0.010)
State-by-Entry Year-by-Loc. Type FEs	Y	Y	Y
District FEs	N	N	Y
Observations	1,253	1,253	1,253
R^2	0.535	0.541	0.919

Source: Superintendent Roster Data; SEDA Data
Notes: Table displays estimates of the relationship between value-added and tenure (in Panel A) and value-added and salary (in Panels B and C). Superintendent value-added is scaled to be mean 0 and standard deviation 1. Nearby districts index is calculated based on the number of districts within a 10 mile radius and is scaled to be mean 0 and standard deviation 1. Superintendent salaries are in 2021 dollars. Standard errors are clustered by district. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Appendix A Additional Value-Added Estimates

A.1 Value-Added without Dynamics

In this subsection, I show the estimates produced by a model of superintendent value-added that does not include tenure dynamics. Specifically, I estimate a model setting ρ equal to zero, which assumes that managerial effects are realized immediately after a superintendent enters a district. In all other respects, this model is identical to the model that produces my main estimates.

I summarize the results of this model by comparing value-added estimates to my main estimates and by validating these estimates using superintendent transitions. In Appendix Figure A1, I show how these value-added estimates compare to the main estimates used in the body of the paper. Panel A of Appendix Figure A1 shows that these estimates are strongly correlated with my main estimates. However, the distribution of these estimates is much smaller; linear regression estimates whos that, on average, they are 2.5 times smaller. In Panel B of Appendix Figure A1, I show the rank-rank plot of these two series. The estimated rank-rank correlation is 0.95.

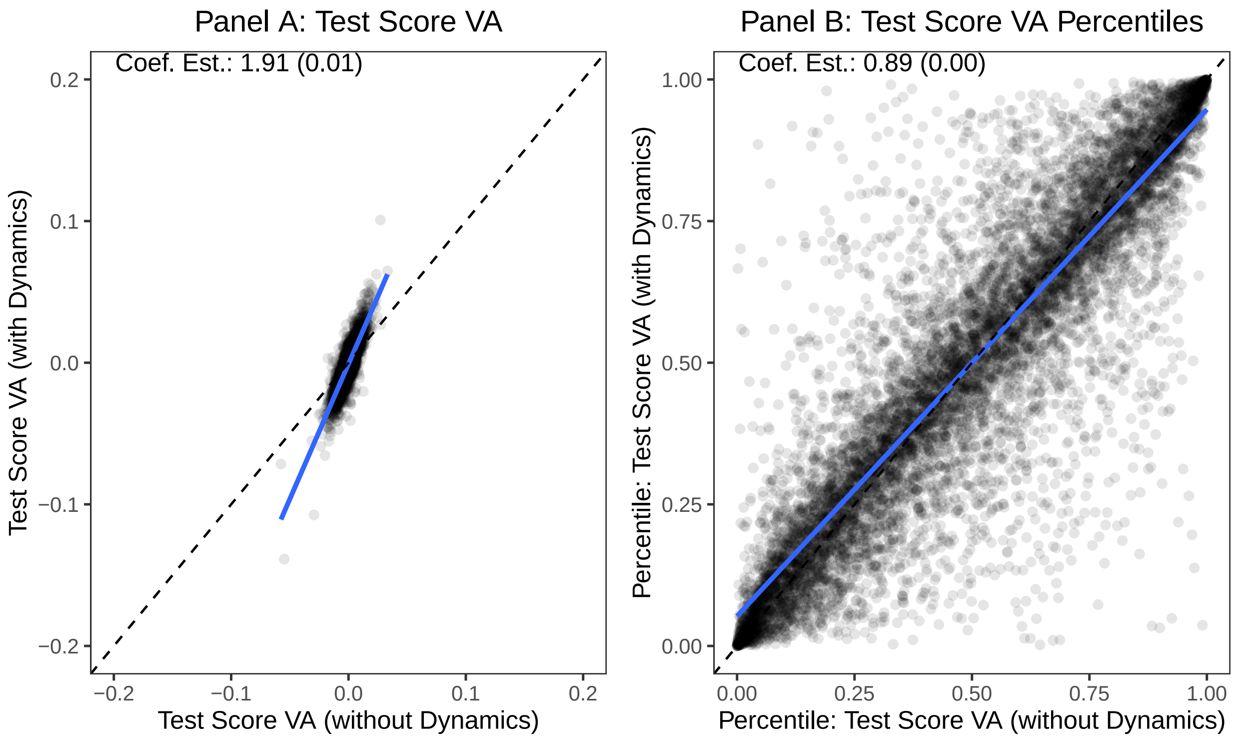
In Appendix Figure A3, I replicate Figure 6 using these estimates of value-added without dynamics. Panels A and B show the dynamic effect of a one standard deviation of 100 percentile points of superintendent value added, respectively. Consistent with the estimates shown in Figure 6, these estimates suggest that a one standard deviation improvement in superintendent value-added increases district test scores by 0.01 to 0.02 standard deviations, and a 100 percentil improvement increases test scores by 0.02 to 0.05 standard deviations.

A.2 Value-Added Excluding Lagged Scores

In this subsection, I show the estimate and validate a model of superintendent value-added that excludes lagged cohort test scores. In all other respects, this model is identical to the model that produces my main estimates.

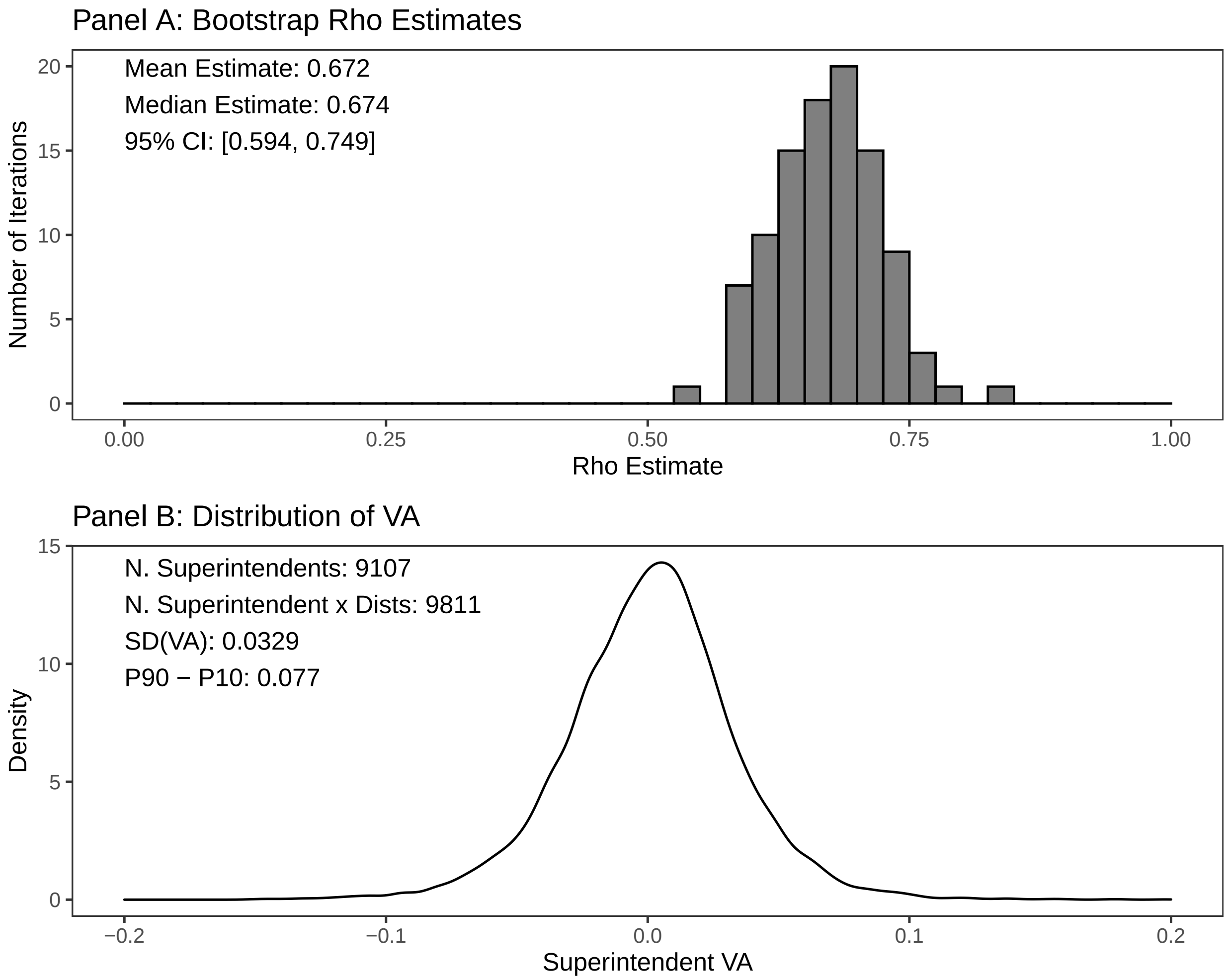
Figure A2 replicates Figure 4, displaying the distribution of boostrap estimates of ρ in Panel A and the estimated distribution of value-added in Panel B. Estimates of ρ are centered around 0.60. The distribution of these value-added estimates is much larger than that of my main estimates; the standard deviation is 0.036. Table A1 displays estimates of Equation 12 with these value-added estimates that exclude lagged cohort test scores. Overall, point estimates are positive, suggesting that the entry of a high value-added superintendent corresponds to increases in test scores. However, the estimates are much less than one, and fail my test for forecast-unbiasedness across a number of different specifications.

Figure A1: Test Score VA with Dynamics vs. Test Score VA without Dynamics



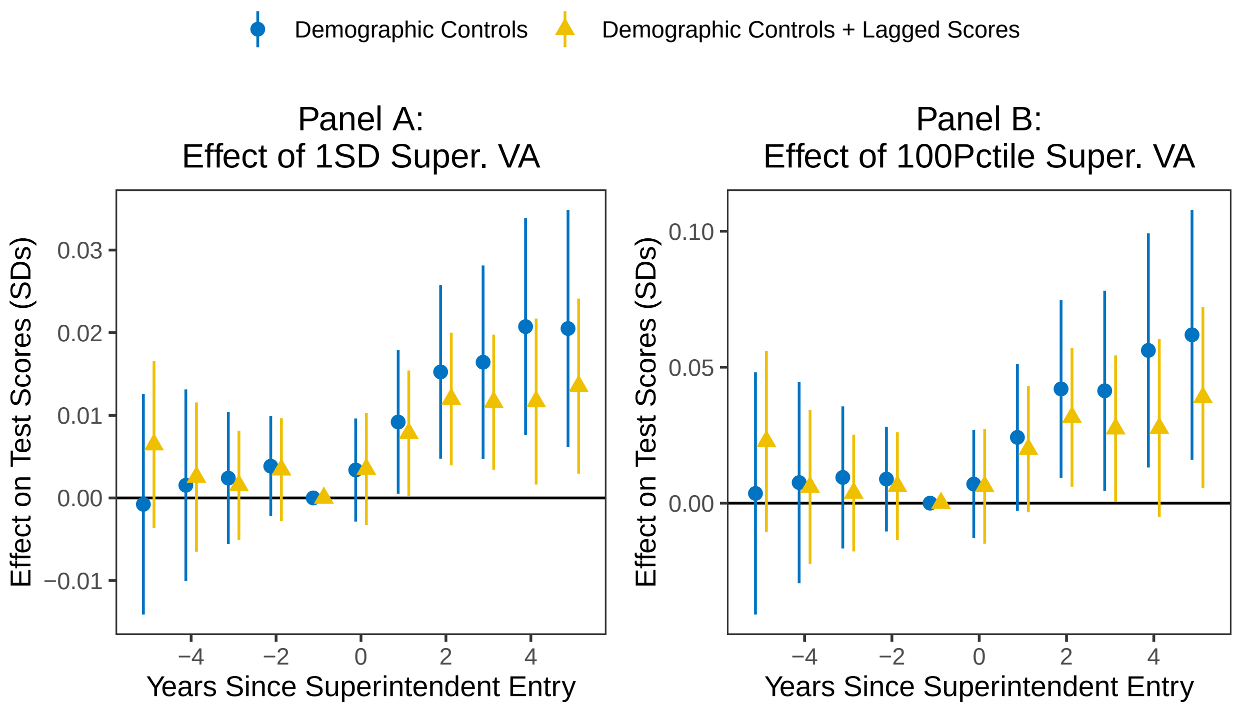
Source: Superintendent Roster Data; SEDA Data
Notes: Figure displays the relationship between superintendent test score value-added estimated without dynamics (on the horizontal axis) and superintendent test score value-added estimated with dynamics (on the vertical axis). Displayed coefficients represent the coefficients of a bivariate regression of the displayed series. Dashed line corresponds to the 45-degree line.

Figure A2: Distribution of $\hat{\rho}$ and Superintendent Value-Added Estimates (No Lagged Score VA)



Source: Superintendent Roster Data; SEDA Data
Notes: Figure summarizes the distribution of estimated superintendent value-added parameters, for value-added estimated without lagged scores. Panel A displays the bootstrap distribution of ρ, based on the bootstrapping procedure described in text. Panel B summarizes the distribution of estimated superintendent value-added

Figure A3: Evolution of District Graduation Rates Surrounding Superintendent Entry (Test Score VA Estimated without Dynamics)



Source: Superintendent Roster Data; SEDA Data
Notes: Figure displays event study estimates of the effect of (out-of-district) superintendent value-added estimates on test scores. Displayed coefficients correspond to β_k in Equation 13 and are produced using the stacked difference-in-difference procedure described in text. Panel A uses value-added estimates scaled to be mean zero and standard deviation one to estimate dynamic effects. Panel B uses percentiles of value-added estimates to estimate dynamic effects. Units are student test score standard deviations. All estimates use value-added estimates produced without dynamics, as described in text. Standard errors are clustered by district. Bands display 95 percent confidence intervals.

Table A1: VA Validation (No Lagged Score VA)

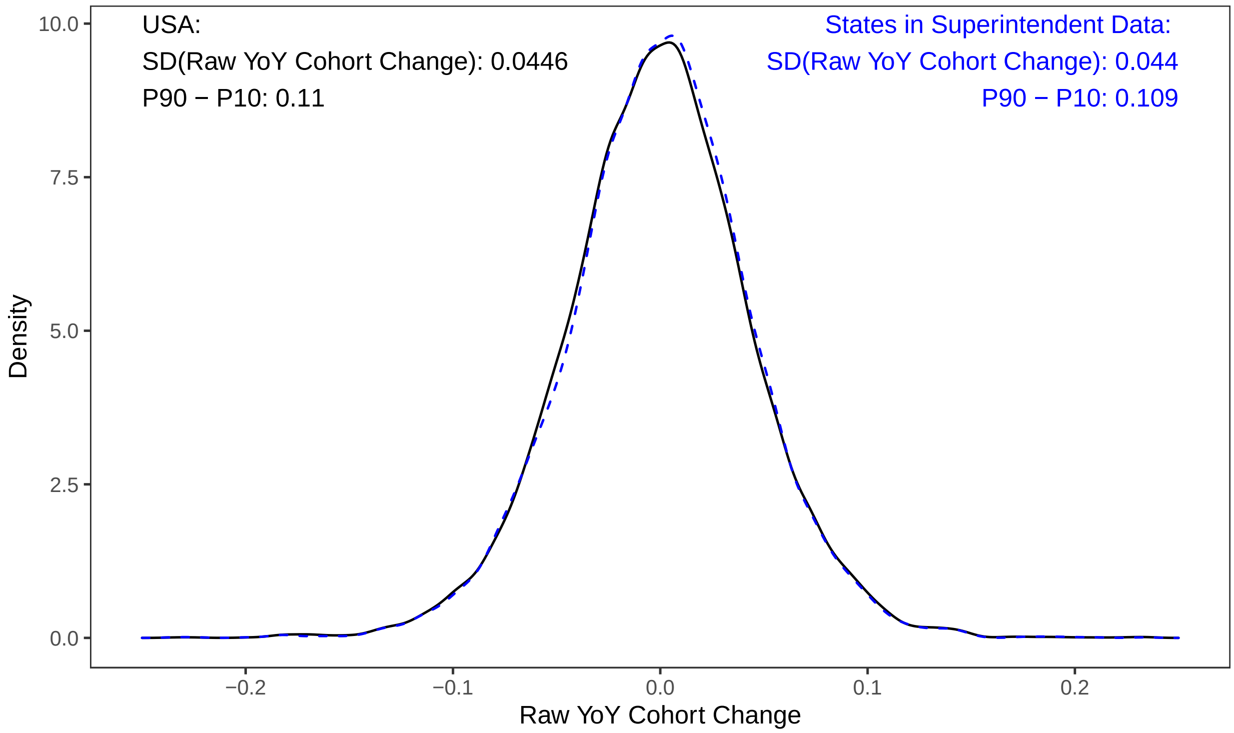
	Assuming $\mu^+ = 0$			Assuming $\mu^+ = \mu_j$		
	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta\theta_{dt}$	0.203	0.221	0.205	0.437*	0.389*	0.288**
	(0.257)	(0.249)	(0.164)	(0.238)	(0.229)	(0.146)
Lagged Score			0.462***			0.462***
			(0.006)			(0.006)
p-value: $\Delta\theta_{dt} = 0$	0.431	0.375	0.210	0.066	0.089	0.049
p-value: $\Delta\theta_{dt} = 1$	0.002	0.002	<0.001	0.018	0.008	<0.001
No. Entry Events	1497	1497	1497	1497	1497	1497
Base Controls	N	Y	Y	N	Y	Y
Lagged Score Controls	N	N	Y	N	N	Y
Observations	108,585	108,585	108,585	108,585	108,585	108,585
R^2	0.714	0.717	0.778	0.714	0.717	0.778

Source: Superintendent Roster Data; SEDA Data

Notes: Table displays estimates validating superintendent value-added using Equation 12. $\Delta\hat{\theta}$ is the predicted change in test scores (based on out-of-district value-added estimates). Value-added is calculated excluding lagged cohort test scores, but is otherwise identical to the model described in text. μ^+ denotes the assumed value-added of superintendents in the years following a superintendent's exit. Standard errors are clustered by district. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

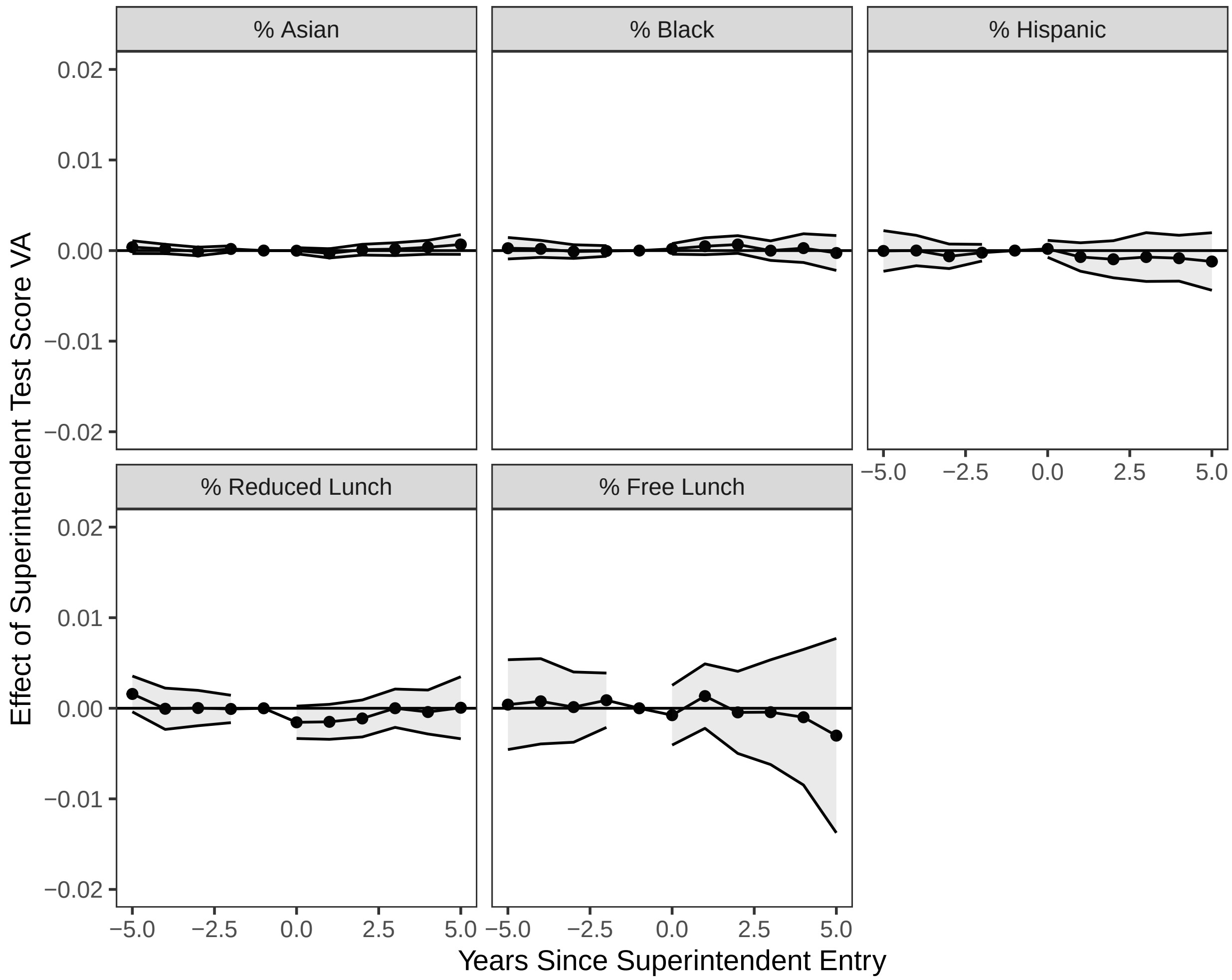
Appendix B Supplemental Figures and Tables

Figure B1: Distribution of District Level Year-over-Year Cohort Changes in Test Scores



Source: SEDA Data
Notes: Figure displays the distribution of raw year-over-year cohort changes in district tests scores, separately for all districts in the US and for states in superintendent data. Districts with fewer than 1,000 students in test score data between 2009 and 2018 are excluded. Units are student test score standard deviations.

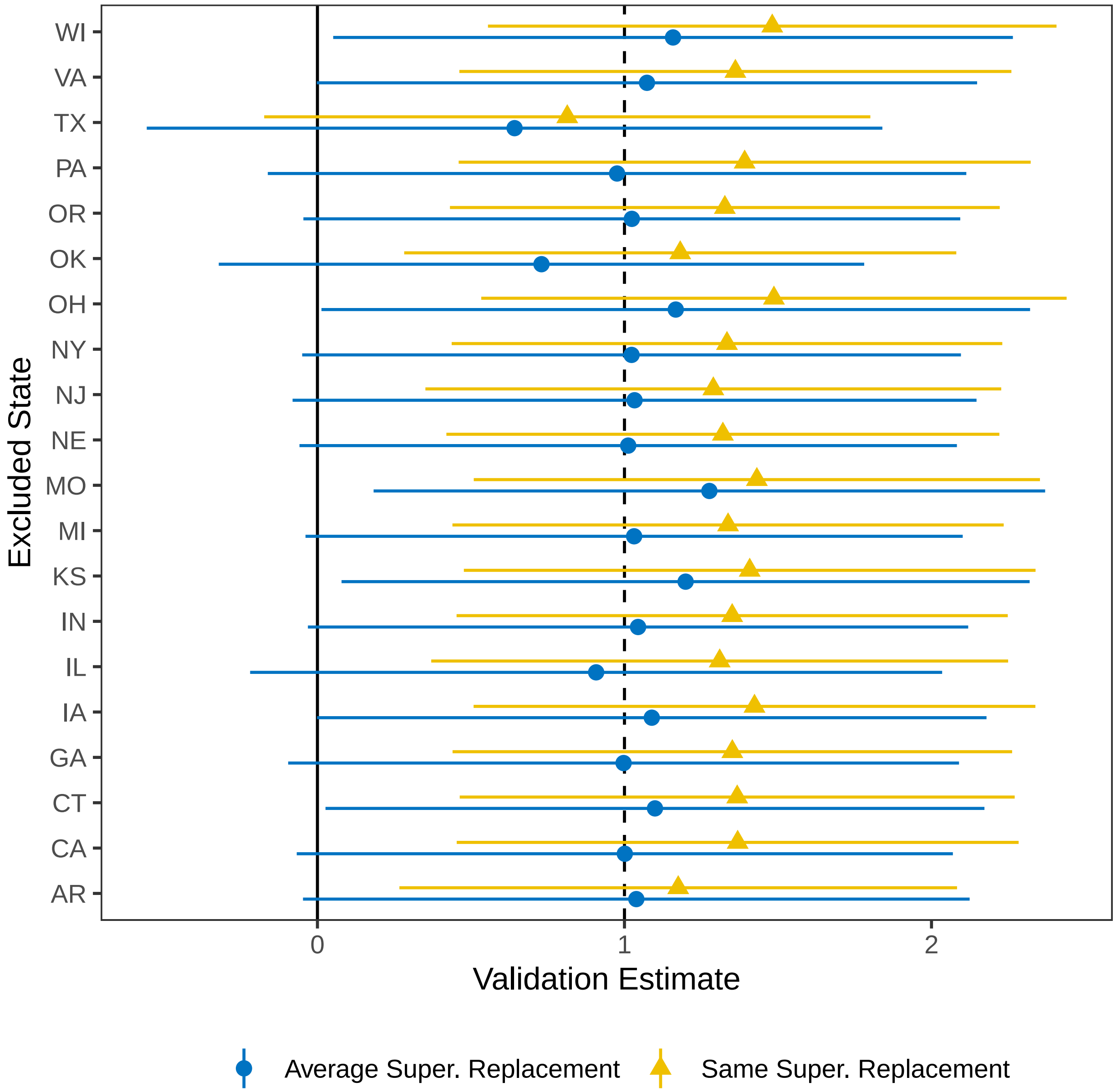
Figure B2: Effect of Superintendent VA on Demographics of District Test-Takers



Source: Superintendent Roster Data; SEDA Data

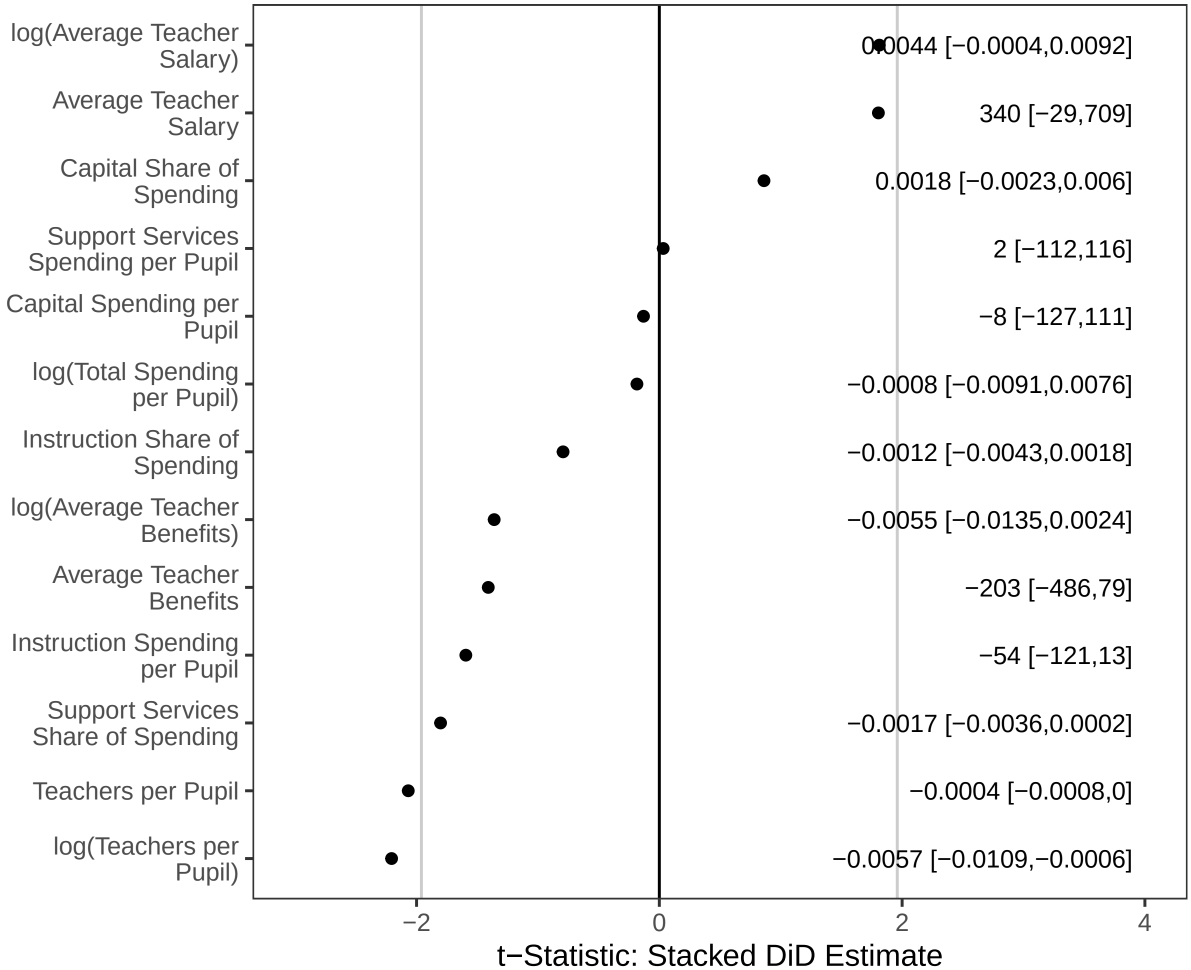
Notes: Figure displays event study estimates of the effect of (out-of-district) superintendent value-added estimates on demographics of test-takers in SEDA data. Displayed coefficients correspond to β_k in Equation 13 and are produced using the stacked difference-in-difference procedure described in text. Superintendent value-added is scaled to be mean 0 and standard deviation 1. Standard errors are clustered by district. Bands display 95 percent confidence intervals.

Figure B3: VA Validation: Sensitivity to State Omission



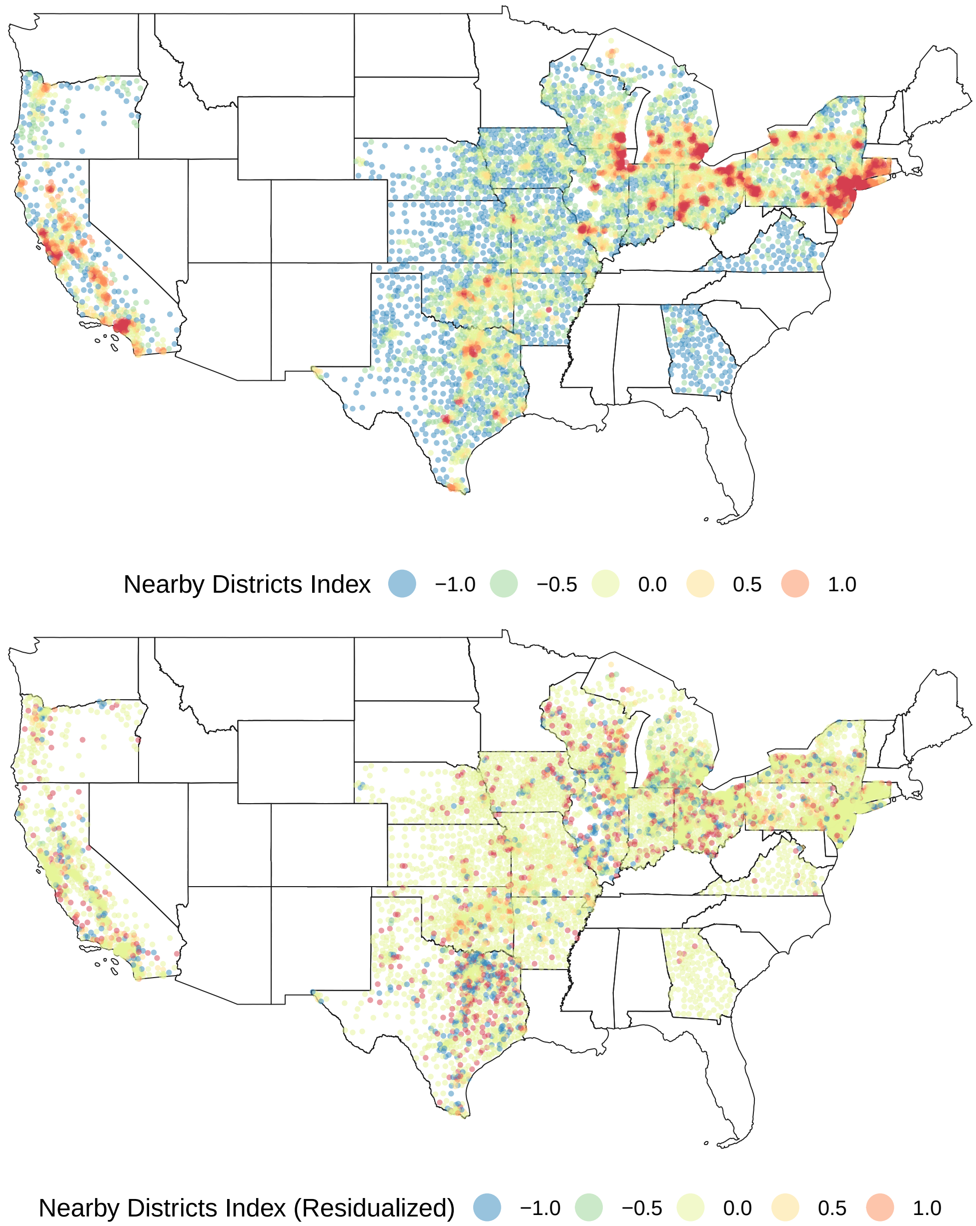
Source: Superintendent Roster Data; SEDA Data
Notes: Figure displays estimates validating superintendent value-added using Equation 12. Each row corresponds to estimates that exclude observations from the identified state. For each state, separate estimates place different assumptions on the assumed value-added of superintendents in the years following a superintendent's exit. Standard errors are clustered by district. Bands display 95 percent confidence intervals.

Figure B4: Effect of Superintendent VA on District Inputs



Source: Superintendent Roster Data; SEDA Data; NCES Common Core of Data
Notes: Figure displays difference-in-differences estimates of the effect of (out-of-district) superintendent value-added estimates on district inputs. Displayed t-statistics correspond to t-statistics on β in Equation 14 and are produced using the stacked difference-in-difference procedure described in text. Point estimates and confidence 95 percent confidence intervals are shown on the right. Superintendent value-added is scaled to be mean 0 and standard deviation 1. Standard errors are clustered by district. All per pupil spending figures are in 2021 dollars.

Figure B5: Geographic Distribution of Nearby Districts Index



Source: Superintendent Roster Data; NCES Common Core of Data
Notes: Figure displays the geographic distribution of my nearby districts index, which is calculated based on the number of districts within a 10 mile radius and is scaled to be mean 0 and standard deviation 1. Residualized values in the bottom map are residualized based on state-by-locale type fixed effects, and are rescaled to be mean 0 and standard deviation 1. Districts that do not appear in my superintendents roster data are not displayed.

Table B1: Superintendent Careers, VA, and District Characteristics (Sensitivities to Fixed-Effects)

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Tenure and Value-Added						
Sample: All Superintendents Hired Pre-2014						
DV: Tenure						
VA	0.030 (0.032)	0.032 (0.032)	0.033 (0.032)	0.049 (0.034)	0.013 (0.056)	−0.006 (0.061)
Nearby Districts Index	0.032 (0.035)	0.035 (0.035)	0.006 (0.050)	0.038 (0.056)	(0.000)	(0.000)
Nearby Districts Index × VA	0.044 (0.031)	0.048 (0.031)	0.050 (0.031)	0.059* (0.033)	0.044 (0.057)	0.067 (0.061)
Observations	6,825	6,825	6,822	6,825	6,825	6,825
R^2	0.060	0.078	0.081	0.173	0.943	0.960
Panel B: Year-0 Pay and Value-Added						
Sample: All Superintendents with Salary Data						
DV: log(Salary) in Year 0						
VA	0.010 (0.010)	0.010 (0.010)	0.021** (0.008)	0.021*** (0.008)	−0.001 (0.011)	−0.002 (0.012)
Nearby Districts Index	0.134*** (0.012)	0.134*** (0.012)	0.020 (0.013)	0.034** (0.015)	(0.000)	(0.000)
Nearby Districts Index × VA	0.015 (0.009)	0.014 (0.009)	0.012 (0.008)	0.014* (0.008)	0.002 (0.010)	0.001 (0.011)
Observations	1,253	1,253	1,253	1,253	1,253	1,253
R^2	0.241	0.254	0.410	0.509	0.868	0.903
Panel C: Year-1 Pay and Value-Added						
Sample: All Superintendents with Salary Data						
DV: log(Salary) in Year 1						
VA	0.009 (0.009)	0.008 (0.009)	0.019** (0.008)	0.021*** (0.008)	−0.015 (0.011)	−0.012 (0.014)
Nearby Districts Index	0.133*** (0.011)	0.133*** (0.011)	0.017 (0.012)	0.034** (0.014)	(0.000)	(0.000)
Nearby Districts Index × VA	0.021** (0.009)	0.021** (0.009)	0.019** (0.008)	0.017* (0.009)	0.016* (0.008)	0.013 (0.010)
State FEs	Y	N	N	N	N	N
Entry Year FEs	Y	N	N	N	N	N
State-by-Entry Year FEs	N	Y	Y	N	Y	N
Loc. Type FEs	N	N	Y	N	N	N
State-by-Entry Year-by-Loc. Type FEs	N	N	N	Y	N	Y
District FEs	N	N	N	N	Y	Y
Observations	1,253	1,253	1,253	1,253	1,253	1,253
R^2	0.258	0.263	0.432	0.541	0.890	0.919

Source: Superintendent Roster Data; SEDA Data
Notes: Table displays estimates of the relationship between value-added and tenure (in Panel A) and value-added and salary (in Panels B and C). Superintendent value-added is scaled to be mean 0 and standard deviation 1. Nearby districts index is calculated based on the number of districts within a 10 mile radius and is scaled to be mean 0 and standard deviation 1. Superintendent salaries are in 2021 dollars. Standard errors are clustered

by district. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table B2: Superintendent Careers, VA, and District Characteristics (Sensitivities to Nearby Distance Radius)

Distance Radius:	(1) 2 Miles	(2) 10 Miles	(3) 20 Miles
Panel A: Tenure and Value-Added			
Sample: All Superintendents Hired Pre-2014			
DV: Tenure			
VA	0.042 (0.034)	0.049 (0.034)	0.053 (0.034)
Nearby Districts Index	0.012 (0.035)	0.038 (0.056)	−0.017 (0.056)
Nearby Districts Index × VA	0.013 (0.029)	0.059* (0.033)	0.079** (0.034)
Observations	6,825	6,825	6,825
R^2	0.172	0.173	0.173
Panel B: Year-0 Pay and Value-Added			
Sample: All Superintendents with Salary Data			
DV: log(Salary) in Year 0			
VA	0.024*** (0.008)	0.021*** (0.008)	0.021*** (0.008)
Nearby Districts Index	0.016 (0.011)	0.034** (0.015)	0.050*** (0.014)
Nearby Districts Index × VA	0.028*** (0.010)	0.014* (0.008)	0.010 (0.009)
Observations	1,253	1,253	1,253
R^2	0.508	0.509	0.514
Panel C: Year-1 Pay and Value-Added			
Sample: All Superintendents with Salary Data			
DV: log(Salary) in Year 1			
VA	0.024*** (0.008)	0.021*** (0.008)	0.021*** (0.008)
Nearby Districts Index	0.019* (0.010)	0.034** (0.014)	0.050*** (0.013)
Nearby Districts Index × VA	0.027** (0.010)	0.017* (0.009)	0.013 (0.009)
State-by-Entry Year-by-Loc. Type FEs	Y	Y	Y
Observations	1,253	1,253	1,253
R^2	0.539	0.541	0.546

Source: Superintendent Roster Data; SEDA Data

Notes: Table displays estimates of the relationship between value-added and tenure (in Panel A) and value-added and salary (in Panels B and C). Superintendent value-added is scaled to be mean 0 and standard deviation 1. Nearby districts index is calculated based on the number of districts within a 2, 10, or 20 mile radius and is scaled to be mean 0 and standard deviation 1. Superintendent salaries are in 2021 dollars. Standard errors are clustered by district. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Appendix C Sources of State Superintendent Data

I collect data on district superintendent assignments and teacher payrolls individually from state agencies and other state-specific sources. Appendix Figure C1 shows the number of districts in my data for each state and year. I describe the source of superintendent roster data for each state below.

C.1 Arkansas

I collect data on Arkansas superintendent assignments from superintendent contact list on the Arkansas Department of Education Data Center.

C.2 California

I collect data on California superintendent assignments and from "Contact Information" files from archived versions of School Accountability Report Card Data Files.

C.3 Connecticut

I collect data on Connecticut superintendent assignments from district-level "Profile and Performance Reports" available from the Connecticut Department of Education's EdSight Portal.

C.4 Georgia

I collect data on Georgia superintendent assignments and teacher payrolls from Open Georgia's "Salary & Travel Reimbursements" database.

C.5 Illinois

I collect data on Illinois superintendent assignments from the llinois State Board of Education's "Archived Directory of Education Entities."

C.6 Indiana

I collect data on Indiana superintendent assignments from archived files of the Indiana Department of Education's School Directory from web.archive.org.

C.7 Iowa

I collect data on Iowa superintendent assignments from annual "Public School District Director[ies]" published by the Iowa Department of Education.

C.8 Kansas

I collect data on Kansas superintendent assignments from current and archived copies of the Kansas State Department of Education's "Directory of Superintendents for Kansas Unified School Districts" files.

C.9 Michigan

I collect data on Michigan superintendent assignments from archived versions of "Legacy DBF Files" from the state of Michigan's Center for Education Performance and Information's Educational Entity Master.

C.10 Missouri

I collect data on Missouri superintendent assignments from copies of "Index of School Administrators" or "Alphabetical List of School Administrators" files from archived versions of the Missouri Department of Education's School Directory website.

C.11 Nebraska

I collect data on Nebraska superintendent assignments from directory searches from the Nebraska Department of Education Education Directory Search.

C.12 New Jersey

I collect data on New Jersey superintendent assignments from archived copies of the State of New Jersey Department of Education's New Jersey School Directory.[20]

C.13 New York

I collect data on New York superintendent assignments from "Archive" files on the New York State Department of Eduaction's Directory of Public and Nonpublic Schools and Administrators in New York State.

C.14 Ohio

I collect data on Ohio superintendent assignments from annual "District Ratings" files from the Ohio Department of Education Report Card Website.

[20]Directories for more recent years can be found at the New Jersey Department of Education's more recent website.

C.15 Oklahoma

I collect data on Oklahoma superintendent assignments from archived School District Directory files on the Oklahoma Department of Education "State Public School and District Directories" website.

C.16 Oregon

I collect data on Oregon superintendent assignments from the Oregon Department of Education's Report Card Download "Media Page."

C.17 Pennsylvania

I collect data on Pennsylvania superintendent assignments from the Pennsylvania Department of Education's Professional Staff Summary data files.

C.18 Texas

I collect data on Texas superintendent assignments from the Texas Education Agency's Texas School Directories.

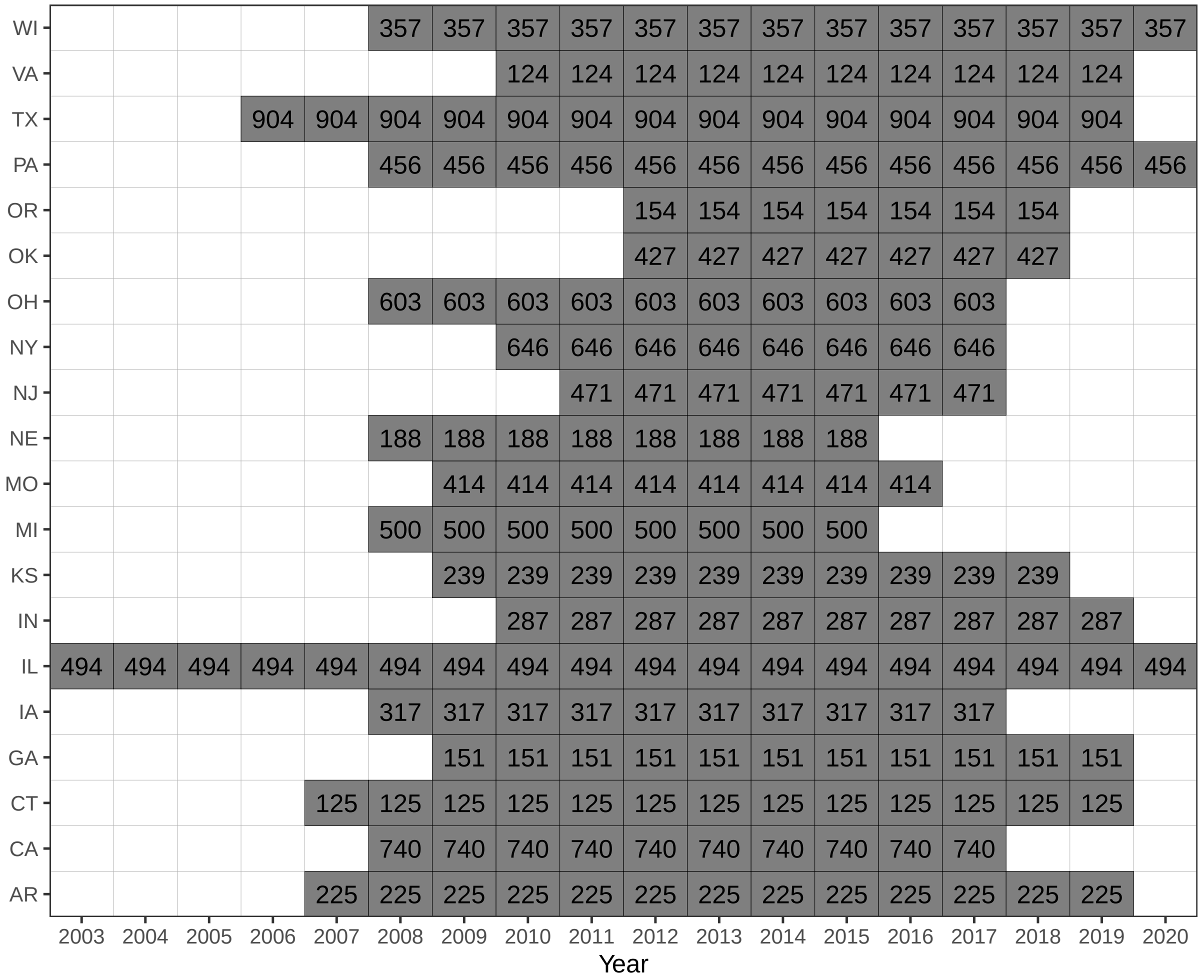
C.19 Virginia

I collect data on Virginia superintendent assignments from archived copies of the Virgina Department of Education's Education Directories.

C.20 Wisconsin

I collect data on Wisconsin superintendent assignments from the Wisconsin Department of Public Instruction Public Staff Reports and archived All Staff Reports.

Figure C1: Superintendent Data: District Counts by State and Year



Source: Superintendent Roster Data
Notes: Figure displays the number of districts represented by each state-year combination in my superintendent roster data.

Appendix D Linking Superintendents Across Rosters

This appendix describes my process for identifying superintendents across annual superintendent rosters. This linkage procedure is different than the typical record linkage case—for example, linking individuals across adjacent censuses—because some of my links will be across non-adjacent years. For example, consider a superintendent who worked in one district from 2009 to 2011 and another district from 2013 to 2015. A linking procedure that is restricted to adjacent years only will fail to recognize that these two tenures belong to the same superintendent. My process involves three steps, which I describe individually below.

D.1 Standardizing Names

To begin, I standardize raw superintendent names to remove inconsistencies in names across superintendent rosters. First, I change any names that appear as Last, First to appear as First Last. Next, I remove the following prefixes from the start of reported names: "mrs", "ms", "dr", "mr", and "miss." Finally, I remove middle initials from all names.

D.2 Identifying Potential Matches

Next, separately for each state, I create a cross product of all unique names that appear in superintendent roster data. To identify potential matches, I calculate a measure of string similarity for all pairs of names, which is higher for names that are more textually similar. I use the normalized Optimal String Alignment (OSA) distance as my metric of similarity. OSA distance is "the minimum cost of operations (insertions, deletions, substitutions or transpositions) required to transform into one string into another."[21]

Appendix Table D1 provides examples of OSA distances for a set of hypothetical names. I define a set of potential matches by linking any names with an OSA score above 0.95. Name pairs meeting this criteria are bolded in Table D1. Note that this threshold is remarkably strict, and allows only small deviations of spelling for two names to be linked.

Using these links, I define sets of connected names as the sets of unique clusters of linked names. For example, if name A is linked only to name B, and name B is linked to names A and C, and name C is linked only to name B, then A, B, and C form a cluster of linked names.

[21] More detailed descriptions can be found in the documentation for the "stringdist" R package and the documentation for the "comparator" R package.

D.3 Breaking Erroneous Matches

This procedure sets a high threshold for potential matches, requiring names to be nearly identical to be treated as matches. Still, the procedure potentially over-matches names that are very common. To break links between potentially false positive matches, I break links for superintendents that I observe in more than one district in the same year. To allow for a one-year overlap when superintendents switch districts, I exclude from this process superintendents who only have one overlapping year and whose total number of district-years employed is at least five.

Table D1: Example OSA Distance Scores for Names

	Samuel Webster Stemper	Samuel Webster Stemper, PhD	Sammuel Webster Stemper	John Doe
Samuel Webster Stemper	**1.000**	0.814	**0.956**	0.091
Samuel Webster Stemper, PhD	0.814	**1.000**	0.778	0.074
Sammuel Webster Stemper	**0.956**	0.778	**1.000**	0.0870
John Doe	0.091	0.074	0.0870	**1.000**

Source: Author Calculations
Notes: Table displays OSA distance measures for name pairs identified in rows and columns.