Center for Education Policy Analysis
SCHOOL OF PUBLIC AFFAIRS
UNIVERSITY OF COLORADO **DENVER**

The System-Level Effects of Denver's Portfolio District Strategy

Technical Report

Parker Baxter, Anna Nicotera, Erik Fuller, Jakob Panzer, Todd Ely, Paul Teske.

This research is made possible by a multiyear grant from Arnold Ventures.
Research support is provided by Basis Policy Research.

Table of Contents

01 SUMMARY

Between approximately 2008 and 2019, Denver Public Schools (DPS) implemented a coordinated, system-wide reform strategy involving an embrace of school choice for families within a common regulatory framework, annual evaluation of school performance, closure of low-performing schools, creation of replacement and new schools with internal and external partners, and district-led school turnaround. Denver's comprehensive implementation of its strategy over a decade provides a context to test whether it is possible to improve public education at scale through an alternative paradigm with different operating assumptions, namely choice for families among multiple providers within a common market for publicly funded schooling that is governed and regulated for quality and equity.

This is a report from a multiphase empirical research study focused on the system-level impact of Denver's implementation of the portfolio district strategy between 2008 and 2019. In this first phase of research, the question is whether system-level outcomes improved in Denver relative to other districts in Colorado during the study period. To answer this question, we rely on two quasi-experimental study designs that are commonly used in education research: comparative interrupted time series (CITS) and difference-in-differences (DID). The CITS analyses use district-level data from 2004-05 through 2018-19, which includes four years of pre-reform and 11 years of reform data, to estimate the impact of DPS reform efforts on student performance in English Language Arts (ELA) and math. A state-level change in the calculation protocols for high school graduation rates reduced the available data for this outcome measure to 2006-07 through 2018-19, necessitating the shift to DID for the associated analyses.

Prior to the start of DPS reform efforts in 2008-09, DPS was among the bottom 10 districts in the state in ELA and math performance on state standardized assessments, ranking below the 5th percentile of districts. By 2018-19, DPS had risen to the 60th percentile of districts in ELA and the 63rd percentile of districts in math, outperforming more than 100 out of roughly 180 districts in the state. The 4-year high school graduation rate increased dramatically during the reform period as well, climbing from 43% in 2008 to 71% in 2019. Our results indicate that the reforms drove these improvements in student academic and graduation outcomes. Specifically, using other low-performing school districts as a comparison groups, the 11 years of reform produced statistically significant annual effects that translate to an overall improvement of 0.561 standard deviations in ELA and 0.572 standard deviations in math. The effect on graduation rate, which is also statistically significant, amounts to an improvement of 14.6 percentage points, compared to other large districts in the state.

Equally important, our findings indicate that the reforms also benefited student subgroups, as all statistically significant results were positive. In math, again considering the low-performing comparison groups, the 11-year reform improvements were 0.341 standard deviations for Special Education students, 0.484 standard deviations for Black students, and 0.506 standard deviations for White students. White students also showed an 11-year improvement of 0.374 standard deviations in ELA, while the graduation rate for Special Education students improved 12.2 percentage points relative to comparison districts. Because the reform strategy was a system-level intervention and because all student groups benefited from them, existing achievement gaps persist.

02 DENVER PUBLIC SCHOOLS' REFORM STRATEGY

In the fall of 2007, Denver Public Schools (DPS), Colorado's second-largest school district at the time, launched what was and remains one of the most comprehensive efforts to restructure the delivery of public education in American history, matched only by the transformation of the public school system in New Orleans, Louisiana, following Hurricane Katrina in 2005 (Baxter, Teske, & Ely, 2019; Harris, 2020).

Between approximately 2008 and 2019, DPS implemented a coordinated, system-wide reform strategy involving an embrace of school choice for families within a common regulatory framework, empowerment for educators at the school level, annual evaluation of school performance, closure of low-performing schools, creation of replacement and new schools with internal and external partners, and district-led school turnaround (Baxter et al., 2019).

DPS's decade-long reform was not a random assortment of strategies. Rather, the strategy implemented in Denver was explicitly modeled on what has become known as the portfolio district strategy (Baxter, 2012a). Developed and refined by Paul T. Hill at the turn of the 21st century, the portfolio district strategy is itself a combination of the three primary levers for education improvement that constituted the bipartisan consensus: choice for families, autonomy for school providers, and accountability for student outcomes. In contrast to other strategies that employed these elements in isolation, either within the unitary district model or to disrupt it from the outside, the portfolio strategy is an effort to fundamentally reinvent the unitary model itself (Hill, 1996; Hill, Campbell & Harvey, 2000; Gross & Lake, 2011; Hill et al., 2009).

Few other cities in the United States have so thoroughly altered the way they govern and deliver public education. Denver was not alone among large city school districts in implementing the portfolio strategy, but along with Chicago, New York City, and New Orleans, it was one of the first and, next to New Orleans, it is the most comprehensive and most long-lasting. Other large cities that have employed the strategy or elements of it include Cleveland, Los Angeles, Hartford, Indianapolis, Newark, Camden, Philadelphia, Baltimore, San Antonio, and Washington, D.C.

What distinguished Denver's portfolio strategy from nearly every effort before it and since is that it was the first time in American history that an elected school board voluntarily relinquished the exclusive franchise to operate schools within its boundaries while maintaining its authority to govern all schools in the district. In doing so, the district rejected the model of singularity in favor of multiplicity (Baxter et al., 2019; Baxter, 2012a, 2012b).

Components of Denver's portfolio strategy included:

- The governance and oversight of multiple school providers.
- A unified enrollment system with a single application and a coordinated process for applying to all schools in the district.
- A district-developed performance framework used to evaluate all schools in the district.
- An annual request for new school proposals from semi-autonomous innovation schools and charter schools based on district identified needs ("Call for New Quality Schools").
- Co-location of traditional district, innovation, and charter schools in underutilized facilities.
- A student-based budgeting model based on student need (Baxter et al., 2019).

DPS implemented its strategy consistently over more than a decade, and its components were applied to all publicly funded schools in the city (Figure 2: Denver's Portfolio District Strategy Timeline with Components).

Figure 2
Denver's Portfolio District Strategy Timeline with Components

	Give families choice	Give schools autonomy	Assess school performance	Schools improve or get intervention	Expand or replace schools
2008	▪ Office of New Schools	▪ Innovation Schools Act ▪ Charter Schools Act of 1995 - in existence prior to DPS reform efforts	▪ School Performance framework (SPF)		▪ 8 schools closed due to low enrollment, lagging performance, and budget considerations ▪ Office of New Schools ▪ Call for New Quality Schools
2009					
2010		▪ First innovation school open			▪ Locating Quality and Access Report ▪ Strategic Regional Analysis ▪ Co-Location Toolkit
2011	▪ District charter collaboration compact signed			▪ School Improvement Grant Turnaround ▪ Denver Summit Schools Network	
2012	▪ Enrollment zones			▪ West Denver Network	
2013	▪ Unified enrollment system			▪ Tiered Support Framework	
2014					

	Give families choice	Give schools autonomy	Assess school performance	Schools improve or get intervention	Expand or replace schools
2015			▪ School Performance Compact (SPC)		▪ School Performance Compact
2016			▪ No SPF Because of Change in Assesment		
2017					
2018			▪ Board votes to pause SPC		▪ Board votes to pause SPC
2019					

Unlike other large city systems that have implemented the strategy in pieces and at various times, and unlike those cities that were unable to sustain the strategy, Denver's implementation of the portfolio strategy was comprehensive across its entire system and from the beginning employed the three core elements—choice, autonomy, and accountability—in an interconnected way and sustained its reforms for a decade, including four board elections. During the study period, the district opened 65 new schools, and closed, replaced, and restarted over 35 others. When the strategy was launched in 2007, Denver authorized fewer than 20 charter schools and innovation schools did not yet exist in Colorado. By 2019, the district authorized over 50 of each—half of all schools in the system—serving almost half of all DPS students (Baxter et al., 2019).

Nationwide variation in context and implementation limits the generalizability of the effects of the strategy found in New Orleans, and DPS provides a logical complement to this work for two reasons: First, the changes in DPS were not the result of a natural catastrophe but rather an intentional reconsideration of a long-standing system that was once considered an exemplar of the unitary district. Second, DPS did not relinquish operation of all schools, instead maintaining a strong centralized administrative presence while working to support accountable and equitable autonomy and promote collaboration among a variety of school types and service providers (Bulkley, Marsh, Strunk, Harris, & Hashim, 2021).

Denver's unified regulatory and market structure and its decade of coordinated reform creates a rare opportunity to evaluate the system-level impact of the reform strategy and eventually its constituent parts. Denver provides a context to test whether it is possible to improve public education at scale through an alternative paradigm with different operating assumptions, namely choice for families among multiple providers within a common market for publicly funded schooling that is governed and regulated for quality and equity (Baxter et al., 2019; Baxter, 2012a, 2012b).

03 EVALUATING THE SYSTEM-LEVEL EFFECTS OF DENVER'S REFORMS

This is a report from a multiphase empirical research study focused on the system-level impact of Denver's implementation of the portfolio district strategy between 2008 and 2019. In this first phase of research, the question is whether system-level outcomes improved in Denver relative to other districts in Colorado during the study period.

There is an extensive body of empirical evidence about the three categories of effects in isolation and in environments where multiple markets for public schooling overlap, but there is little empirical evidence of system-level impacts on student achievement within a common market for public schooling or about how such a system might work in practice (Angrist, Abdulkadiroğlu, Dynarski, Kane & Pathak, 2011a; Angrist, Cohodes, Dynarski, Pathak, & Walters, 2016; Angrist, Pathak, & Walters, 2011b; Harris, 2019).

The focus on the system-level impact of Denver's strategy combined with its governance and regulatory context allows us to capture participant effects, competitive effects, and accountability effects of the reforms over a decade. There is an extensive body of empirical evidence about the three categories of effects in isolation and in environments where multiple markets for public schooling overlap, but there is little empirical evidence of system-level impacts on student achievement within a common market for public schooling or about how such a system might work in practice (Harris, 2019; Berends, Waddington, & Schoenig, 2018; Chen & Harris, 2021). The one exception, again, is New Orleans (Harris & Larsen, 2016, 2018, 2022). In both places, the portfolio strategy was used not as an alternative to the unitary system but to replace it. Unlike New Orleans, Denver did not replace its entire school system with charter schools. For this and related reasons, the experience of Denver in implementing the portfolio strategy is more universally applicable to large school districts across the country.

State-level policies also played a key enabling and sustaining role for Denver's reforms, including:

- Statewide intra and inter-district open enrollment as a parent entitlement (1991).
- A charter school law that requires districts to authorize charters that are in "the best interests of students and the community" or face losing their exclusive authority to govern public schools within their boundaries (1993).
- One of the first statewide longitudinal education data systems in the nation (2002).
- An accountability system established by the State Board of Education that uses a student growth percentile model (The Colorado Growth Model) (2008).
- The Innovation Schools Act of 2008 which gives school-level employees the authority to vote to waive elements of the collective bargaining agreement and certain district and state policies.

In contrast to other jurisdictions where alternatives to the unitary model operate and are regulated by states alongside and overlapping with districts, Denver provides an opportunity to study student achievement over time and under a variety of different conditions within a common market for public schooling that includes all schools in the district—district-operated, innovation, and charter—under a common enrollment and expulsion system, a common set of performance metrics, and a common regulatory structure. These and related policies are intended to mitigate potential and existing market failures arising from discoordination among providers and disincentives to serve the neediest students. These policies also require cross-sector collaboration among providers, even as they compete for students (Baxter et al., 2019; Baxter, 2012a, 2012b; Ladd, 2018).

The following sections describe the analytic strategy we use to examine the impact of DPS's reform efforts on system-level student outcomes and the results of our analysis.

04 ANALYTIC STRATEGY

The purpose of this report is to examine whether DPS's reform efforts resulted in improvements in system-level student outcomes. As described above, DPS's portfolio reform strategy included a variety of mechanisms, including different school governance types, school choice, a common enrollment system, accountability, school turnarounds and closures, and shared facilities. In this report, we will not attempt to disentangle the impact of discrete components of DPS's overall portfolio reform strategy. Following Chen and Harris (2022), we use district-level data and focus on system-level outcomes, which are the average of all outcomes in the district, including traditional public schools, innovation schools, and charter schools. This approach will allow us to examine the overall impact for DPS's portfolio reform for all students regardless of the type of school students attended within DPS.

We use comparative interrupted time series (CITS) for math and ELA achievement and difference in differences (DID) for changes in graduation rates to estimate whether DPS's system-level performance trends and means during reform years differed from pre-reform conditions by an amount significantly greater than the change observed in comparison districts. DID is a long-standing staple of quasi-experimental comparative analysis (Blazar & Schueler, 2022). Researchers have now also assessed the validity of CITS as a quasi-experimental research design to examine system-level outcomes and determined that the approach can eliminate estimation bias with the use of multiple years of preintervention outcome data and the selection of appropriate comparison groups (Hallberg, Williams, Swanlund, & Eno, 2018; Hallberg, Williams, & Swanlund, 2020; Jacob, Somers, Zhu, & Bloom, 2016; St. Clair, Hallberg, & Cook, 2016).

The combination of approaches allows us to address the following threats to estimating the impact of DPS reform efforts:

- First, DPS experienced performance and graduation rate gains in the years before reform efforts began. Our CITS and DID models compare the statistical significance of changes in outcomes before and after the reform efforts. Prior to running the respective analyses, we test whether DPS and the comparison groups had parallel pre-reform performance trajectories. If DPS and the comparison groups were similar before reform, we can — accounting for other changes in the environment — attribute differences in growth trajectories and/or means during the reform years to DPS's reform efforts. In other words, even though DPS may have had positive gains in performance prior to reform, comparison to similarly improving control groups allows us to attribute post-reform increases in performance to the reform efforts.

- Second, academic performance and graduation rates may have improved across the state during the span of years in our study due to other education initiatives to which all districts were subject, such as federal or state accountability initiatives that began prior to the DPS reform effort. The use of several comparison groups allows us to determine whether the growth in DPS performance was reflective of overall growth in the state or due to the DPS reform efforts.

- Third, over the course of the pre-reform and reform years in this study, DPS was a growing and demographically shifting school district. Enrollment grew by over 25 percent and there were changes to the student demographic composition. We conduct sensitivity tests that examine whether changes in the district residential demographics as measured by census data predict improvements in student performance and we subsequently include significant predictors in our models as control variables.

- Fourth, DPS demonstrated large improvements in academic performance when the state standardized assessment changed from CSAP/TCAP to CMAS in 2015 amid DPS reform. We conduct sensitivity tests by running the CITS models through 2014 instead of 2019 to examine whether the impact of DPS reform efforts was driven by the jump in test scores with the new assessment.

05 DESCRIPTIVE STATISTICS

We use district-level administrative data for this study from the Colorado Department of Education (CDE) for the 2004-05 through 2018-19 school years. We do not analyze data for this study beyond 2018-19 because CDE did not administer state assessments in spring 2020 due to the COVID-19 pandemic. With DPS reform efforts beginning in the 2008-09 school year, the data provide four years of pre-reform data and 11 years of reform data.

We downloaded publicly available district-level data from CDE, including demographics, state standardized assessments, high school assessments, graduation rates, and mobility rates. See Appendix A for more information about the data. The data are district-level averages for all publicly funded schools in each district.

Overall, the descriptive data for DPS show a growing, demographically shifting, and academically improving school district. Total enrollment in DPS grew from roughly 72,000 students in 2004-05 to over 91,000 students in 2018-19, a 27 percent enrollment growth rate. Over that same period the district demographics shifted to higher percentages of White and Other students and lower percentages of Black and Hispanic students, while at the same time serving higher percentages of students qualifying for free or reduced lunch (FRL) and English learner (EL) students. Perhaps the most dramatic change for DPS was the academic improvement. Over 15 years, DPS's academic performance on state standardized and high school assessments improved from significantly underperforming the state average to an urban district that outperformed the state average in both ELA and math.

06 DPS ENROLLMENT

Table 6 summarizes DPS enrollment by demographics for the first year of data for this study (2005), the first year of DPS reform efforts (2009), and the final year of data for this study (2019). The data include the number and percent of students by demographic category, as well as the average annual change in enrollment for the four years of pre-reform data and the 11 years of reform data. The descriptive enrollment data show that while DPS was growing in the pre-reform years (an average increase of 441 students per year), enrollment growth was higher in the reform years (an average increase of 1,782 students per year).

Table 6
DPS Enrollment, Grades PK-12

Grades PK-12	2004-05		2008-09		2018-19		Average Annual Change	
							Pre-Reform	Reform
	Num.	Pct.	Num.	Pct.	Num.	Pct.	Num.	Num.
Total	72,412		74,176		91,998		441	1,782
EL	21,688	29.95	24,763	33.38	30,168	32.79	769	541
FRL	44,551	61.52	48,904	65.93	59,702	64.89	1,088	1,080
Race								
Black	13,744	18.98	12,758	17.20	11,918	12.95	-247	-84
Hispanic	41,524	57.34	41,176	55.51	49,629	53.95	-87	845
Other	3,115	4.30	3,364	4.54	7,603	8.26	62	424
White	14,029	19.37	16,878	22.75	22,848	24.84	712	597
Special Educ.	8,550	11.81	8,557	11.54	10,499	11.41	2	194

In the pre-reform years, growth in enrollment districtwide was driven by increases in White, FRL, and EL students. In the reform years, growth in enrollment was driven by increases in White, Hispanic, Other, FRL, EL, and Special Education students. The number of Black students declined in both the pre-reform and reform years.

07 DPS STATE STANDARDIZED ASSESSMENTS

Table 7.1 summarizes the number of DPS test-takers and weighted mean results from the state standardized assessments for the first year of data for this study (2005) the first year of DPS reform efforts (2009), and the final year of data for this study (2019). The test results are presented as z-scores where the state average (by year, subject area, and subgroup) has a mean of zero and standard deviation of 0.35 (see Appendix A for a discussion of the district z-scores). Subgroups are standardized by the performance of those students in each district across the state. Because the district-level test results are weighted by test-takers, the overall z-score is not the average of subgroups within a district. Z-scores below zero indicate that DPS students performed below the state average in the given year, subject area, and subgroup, whereas z-scores above zero indicate that DPS students outperformed the state average. The last two columns in the table present the average annual change in z-scores for the four years of pre-reform data and the 11 years of reform data.

Table 7.1
DPS State Standardized Assessment Z-Scores, Grades 3-10

Z-Scores Grades 3-10	2004-05		2008-09		2018-19		Average Annual Change	
ELA	Num.	Mean	Num.	Mean	Num.	Mean	Pre-Reform	Reform
Overall	37,943	-0.831	39,687	-0.731	39,070	0.067	0.025	0.080
EL	10,579	-0.276	16,488	-0.284	15,822	0.231	-0.002	0.052
FRL	23,640	-0.743	27,045	-0.685	25,261	-0.105	0.015	0.058
Race								
Black	7,785	-0.567	7,217	-0.394	4,819	-0.140	0.043	0.025
Hispanic	20,930	-0.677	22,222	-0.507	20,821	-0.051	0.043	0.046
White	7,581	0.090	8,288	0.099	9,483	0.856	0.002	0.076
Special Educ.	4,821	-0.578	5,535	-0.444	4,460	-0.119	0.033	0.033

Z-Scores Grades 3-10	2004-05		2008-09		2018-19		Average Annual Change	
Math	Num.	Mean	Num.	Mean	Num.	Mean	Pre-Reform	Reform
Overall	38,843	-0.845	40,522	-0.687	39,506	0.071	0.039	0.076
EL	11,499	-0.286	17,274	-0.239	16,127	0.213	0.012	0.045
FRL	24,476	-0.808	27,819	-0.646	25,473	-0.082	0.041	0.056
Race								
Black	7,782	-0.690	7,248	-0.417	4,843	-0.084	0.068	0.033
Hispanic	21,850	-0.686	22,994	-0.438	20,990	-0.007	0.062	0.043
White	7,564	0.028	8,310	0.106	9,498	0.869	0.019	0.076
Special Educ.	4,917	-0.698	5,607	-0.493	4,477	-0.165	0.051	0.033

Before the DPS reform efforts began and in the first year of reform, all DPS student groups other than White students performed below the state average in both ELA and math. The overall mean z-score for DPS students was 0.831 and 0.845 standard deviations below the state average in ELA and math, respectively. By the last year of this study, the overall performance of students in DPS, as well as White and EL students, exceeded the state average in ELA and math. All the remaining DPS subgroups closed the gap in ELA and math performance.

The average annual change columns show that nearly all DPS subgroups were making improvements in performance compared with the state average before the start of the DPS reform efforts (as demonstrated by positive values). However, the average annual change in performance in the reform years was larger for overall students in ELA and math, as well as for most subgroups. Black students in ELA and math, Hispanic students in math, and Special Education students in math had lower average annual changes in z-scores in the reform years compared with pre-reform years.

Prior to the start of DPS reform efforts, the district performed approximately 0.80 standard deviations below the state average in both ELA and math, but the district was making improvements of roughly 0.03 standard deviations per year. If DPS had stayed on that performance trajectory for the next 11 years, the district would have improved to roughly half of a standard deviation below the state average in 2018-19. Instead, after 11 years of reform efforts, DPS outperformed the state average by 2016-17 in ELA and by 2017-18 in math.

Figures 7.2 and 7.3 present performance trajectories using z-scores from the state standardized assessments for DPS students compared to the state for ELA and math between 2004-05 and 2018-19, with 2008-09 indicated as the start of the reform efforts. The state average is a z-score of zero in every year because the test scores are standardized by all districts in the state annually by subject area and subgroup. The overall DPS trajectory is not a simple average of the subgroups (see description of z-scores above). The graphs show the increases in performance that DPS students overall and all DPS subgroups experienced over the span of the study years, including a closing of the gap between subgroups and the state average for those subgroups.

Figure 7.2
DPS State Standardized Assessment Z-Scores, Grades 3-10: ELA

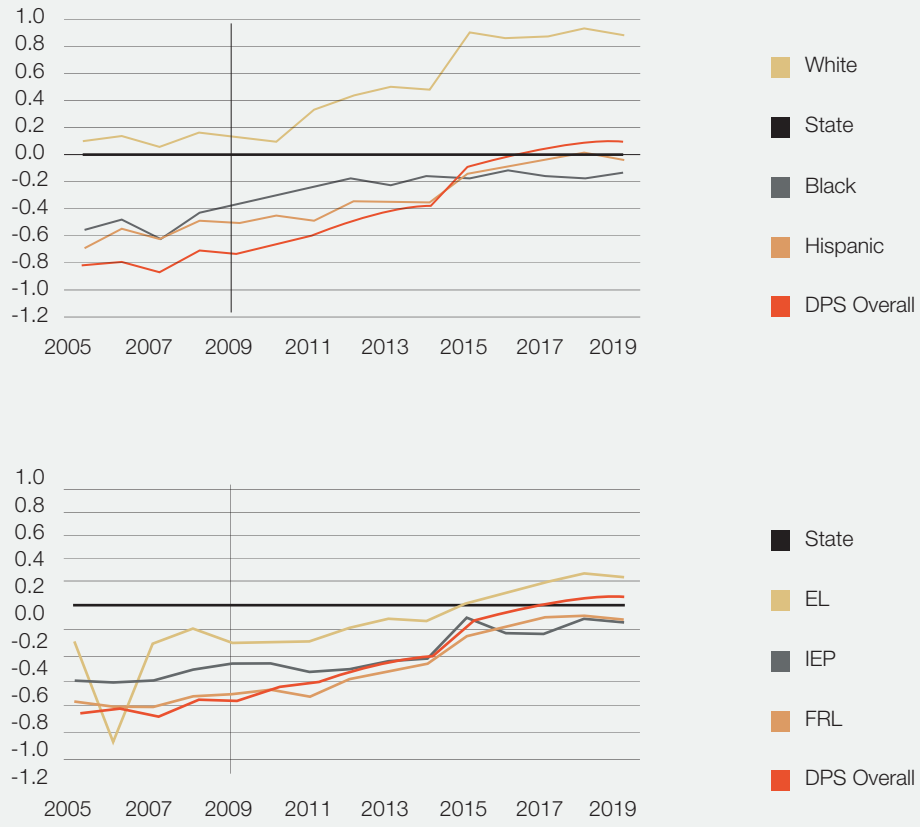
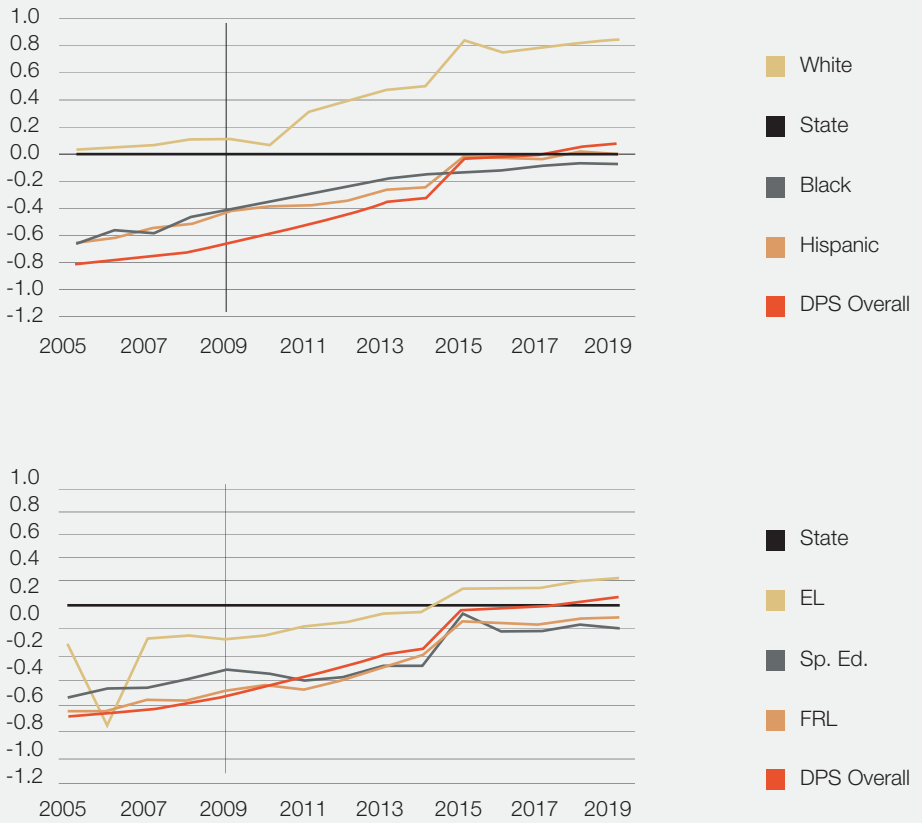


Figure 7.3
DPS State Standardized Assessment Z-Scores, Grades 3-10: Math



Note: ELA and math z-scores are averages from the sample that do not control for the district-level covariates in subsequent analyses.

08 DPS HIGH SCHOOL DATA

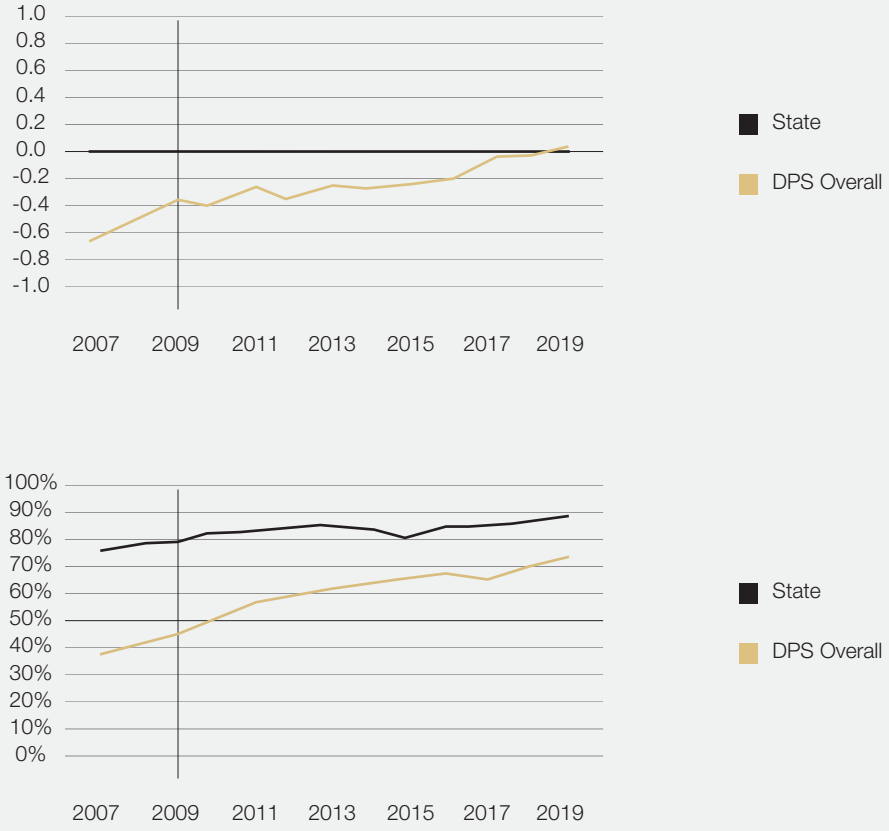
Table 8.1 summarizes two measures of high school student outcomes for DPS in the first year of high school data for this study (2007), the first year of DPS reform efforts (2009), and the final year of data for this study (2019)—standardized ACT/SAT scores for 11th grade students and four-year adjusted cohort graduation rates. Similar to the state standardized assessment results, DPS high school students improved their performance over the years in this study, eventually outperforming the state ACT/SAT average by the final year. However, DPS's graduation rate did not close the gap with the state average and the average annual change for both standardized SAT/ACT and graduation rates was higher in the two pre-reform years compared with the annual average over the 11 years of reform.

Table 8.1
DPS High School Data

	2006-07		2008-09		2018-19		Average Annual Change	
	Num.	Mean	Num.	Mean	Num.	Mean	Pre-Reform	Reform
ACT/SAT z-scores, grade 11	3,245	-0.610	3,367	-0.634	5,049	0.016	0.123	0.038
4-year adjusted cohort grad. rate	2,451%	38.70 %	2,414 %	46.40 %	4,381 %	70.90 %	3.85 %	2.45 %

Figure 8.2 display the performance trajectories for DPS's average ACT/SAT standardized scores for 11th graders and the four-year adjusted cohort graduation rate, with 2008-09 indicated as the start of the reform efforts. The graphs show how DPS closed the ACT/SAT performance gap with the state and made improvements in the graduation rate over the study years.

Figure 8.2
DPS High School Data



Note: ACT/SAT z-scores are averages from the sample that do not control for the district-level covariates in subsequent analyses.

09 CENSUS DATA

Given the performance improvements and demographic shifts for DPS shown in the descriptive data, we use U.S. Census Bureau data to examine whether district-level population changes account for a significant portion of the growth in DPS academic performance. The census data are 1-year estimates from the American Community Survey (ACS) for 2005 through 2019 at the unified district geography level, including total adult population, race/ethnicity, median income, educational attainment, and the count of school-aged children attending public and private schools. The ACS 1-year estimates are available for areas with a total population of 65,000 or more, which includes all of the largest Colorado school districts with public school enrollments of at least 20,000 students.

We use stepwise regression models where the dependent variables are district-level standardized ELA and math scores and the independent variables include one year lagged performance, percent Black, percent Hispanic, percent White, median income, percent attending public school, percent population with an associate's degree or higher, and district and year fixed effects. With a backward selection procedure, we drop insignificant independent variables that do not improve the fit of the model with a significance level of $p > 0.30$. For the ELA model, none of the independent variables are significant predictors of changes in performance. For the math model, however, the change in median income at the district level was a significant predictor of changes in student performance, so we include the census median income variable in our models as a control variable. It is worth noting that models with all of the census variables, regardless of significance in the stepwise regression model, produced similar results.

10 ESTIMATION STRATEGY

We use the following CITS model to estimate the effects of DPS reform (1):

$$Y_{jt} = \beta_0 + \beta_1 T_t + \beta_2 \text{Reform}_t + \beta_3 (\text{Reform}_t * T_t) + \beta_4 \text{DPS}_j + \beta_5 (\text{DPS}_j * T_t) + \beta_6 (\text{DPS}_j * \text{Reform}_t) + \beta_7 (\text{DPS}_j * T_t * \text{Reform}_t) + \beta_8 X_{jt} + \mu_j + \pi_t + \epsilon_{jt}$$

where Y_{jt}, in equation 1, is the district j's standardized average score in ELA or math, normed by subject and year in time t, T_t is the time since the start of the study, Reform_t is a binary variable indicating the DPS reform intervention with pre-reform years equal to 0 and reform years equal to 1, and DPS_j is a binary variable with DPS equal to 1 and comparison districts equal to 0. X_{jt} is a vector of district-level covariates including the natural log of enrollment, the median income of households, and the share of students by race/ethnicity (Black, Hispanic, White), students qualifying for free or reduced-price lunch, English Learners, and Special Education. μ_j is district fixed effects, π_t is year fixed effects, and ϵ_{jt} is the error term. We utilize the Newey-West variance estimator to manage autocorrelation in the residuals and the standard errors are clustered by district.

In the model specification, β_5 indicates the difference in growth trajectories between DPS and the control districts in the pre-reform years. β_5 should be insignificantly different from zero to provide a test of parallel trends in the pre-reform years for DPS and the control districts. β_7 indicates the difference between DPS and the control districts in their growth trajectories in the reform years compared with the pre-reform years. If the parallel trends test on β_5 holds, the treatment effect of DPS reform is $\beta_5 + \beta_7$.

While CITS is our preferred analytic strategy, losing two years of pre-reform data means shifting to the following DID model to estimate the effects of DPS reforms on high school graduation rates (2):

$$Y_{jt} = \beta_0 + \beta_1 \text{DPS}_j + \beta_2 \text{Reform}_t + \beta_3 (\text{Reform}_t * \text{DPS}_j) + \beta_4 X_{jt} + \mu_j + \pi_t + \epsilon_{jt}$$

where Y_{jt}, in equation 2, is district j's graduation rate in year t and the remaining variables are as defined above. Assuming the DPS and comparison group graduation rates would have moved in parallel absent the reforms (i.e., they pass the parallel trends test), β_3 provides an unbiased estimate of the effect of the reforms.

11 COMPARISON GROUPS

To identify a comparison group for DPS, we considered a variety of district-level factors from the pre-reform years, including size, demographic composition, and academic performance. There is not a perfect district match for DPS in Colorado. DPS is one of the largest and most demographically diverse school districts in the state, located in a metropolitan area where the surrounding districts are either large and demographically and academically different or significantly smaller and demographically and academically more similar. Districts in other locations in the state are not similar by size, demographics, or performance.

Rather than identify one matching district for our study, we analyze the impact of DPS's reform efforts compared with several comparison groups. We identify three comparison groups for an overall analysis of system-level student performance. The first comparison group is composed of 10 districts that had total enrollment of 20,000 or higher in the 2004-05 school year. The second comparison group is composed of 30 districts that performed in the bottom 20th percentile in ELA and math in at least three out of four pre-reform years (DPS was in the bottom 20th percentile in all four pre-reform years). The final comparison group is the Aurora Public Schools (APS) district, perhaps the most similar district to DPS when considering both pre-reform demographic composition and academic performance[1].

Table 11.1 shows that each of the comparison groups for our overall analyses have strengths and weaknesses as matches for DPS based on pre-reform characteristics. The group of largest districts were most similar in terms of size, but had higher percentages of White students, lower percentages of FRL and EL students, and outperformed the state test averages in 2005. The districts in the lowest 20th percentile of academic performance had more comparable percentages of Hispanic and FRL students as DPS, as well as lower performance, but they were smaller districts on average. APS was roughly half the size of DPS in 2005 and had low performance levels, but higher percentages of White, Black, and EL students and lower percentages of FRL students.

(1) One issue with using APS as a comparison group for DPS is that APS began implementing district portfolio management reform strategies in 2015-16. The reform efforts included the creation of an office of autonomous schools and using the innovation law to restart schools in one "Action Zone." Given this development, when we compare DPS to APS, we only include data through the 2014-15 school year prior to APS's reform efforts beginning.

Table 11.1
Comparison Groups for Overall Analyses, Pre-Reform Characteristics, 2004-05

2004-05	DPS	Largest Districts	Lowest-Performing Districts	APS
Number of Districts	1	10	30	1
Mean Enrollment	72,412	37,477	3,992	32,251
% EL	29.95	10.69	17.05	42.31
% FRL	61.52	24.97	56.86	48.26
% Race				
% Black	18.98	5.68	2.61	21.79
% Hispanic	57.34	19.64	49.45	44.85
% White	19.37	69.74	43.22	28.41
% Special Education	11.81	9.88	10.63	10.64
Mean ELA z-score	-0.831	0.105	-0.512	-0.772
Mean Math z-score	-0.845	0.116	-0.527	-0.724

2006-07	DPS	Largest Districts	Lowest-Performing Districts	APS
Number of Districts	1	10	30	1
Mean ACT/SAT z-score	-0.610	0.146	-0.375	-0.615
4-year Adjusted Cohort Graduation Rate	38.70	71.01	61.62	44.20

Table 11.2 presents results from tests of parallel trends between the comparison groups and DPS for our overall analyses. The test of parallel trends examines whether the difference in pre-reform growth trajectories was insignificantly different from zero, controlling for district-level demographics. The reported value is coefficient β_5 from the CITS model described above. The available ACT/SAT and graduation data did not allow this test for the APS comparison group. Statistical insignificance indicates that the growth trajectories for DPS and the comparison group were similar in the pre-reform years. The growth trajectories were parallel for DPS and all three comparison groups for ELA and math from the state standardized assessments. For the high school ACT/SAT assessment, none of the district comparison groups met the pre-reform test of parallel trends, so we are not able to analyze the ACT/SAT high school assessment data for this report. The test of parallel trends was statistically insignificant for high school graduation rates, but only for the large district comparison group. Although the standardization built into the nationally normed assessments would have been preferable to performance against graduation requirements that may vary between districts, we recognize the importance in evaluating differences between DPS and the comparison groups beyond 10th grade and consider an evaluation of graduation rates worthwhile, even for a single comparison group.

Table 11.2
Pre-Reform Tests of Parallel Trends with DPS for Overall Analyses

	Largest Districts	Lowest-Performing Districts	APS
ELA	0.037	0.003	-0.079
Math	0.015	0.016	-0.089
ACT/SAT	0.150***	0.129**	--
Graduation rate	0.027	0.051**	--

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Based on the tests of parallel trends for our overall analyses, we present findings for all three comparison groups for the CITS models using state standardized assessment results as outcome measures. Results for the graduation rate DID analysis reflect the large district comparison group model.

The three comparison groups we identify for our overall analyses do not perform as consistently on the test of parallel trends for DPS's subgroups. For example, when we compare the academic performance of DPS's Black students to Black students in the largest districts, DPS's Black students had a different performance trajectory prior to the reform efforts. With this lack of parallel trends, we cannot attribute changes in growth after reform started to the DPS reform efforts.

In order to report on the impact of DPS reform efforts on subgroups, we identify separate control groups for each subgroup. Specifically, we select a subset of districts from the set of lowest-performing districts comparison group we use for our overall analysis based on the demographic composition of the given subgroup. For example, to examine the impact of DPS reform efforts on FRL students, we selected districts from the lowest-performing districts comparison group with an average districtwide FRL percentage of 60 percent or higher between 2004-05 and 2007-08. This selection strategy resulted in 14 districts for the comparison group when we examine the impact of DPS reform efforts on FRL students.

Tables B1 and B2 in Appendix B provide details on the demographic characteristics of the subgroup comparison groups for the CITS and DID analyses, respectively. Table 11.3 provides results of the tests of parallel trends that indicate we can present findings for all subgroups when compared to their respective comparison group of districts among the lowest-performing districts, except for graduation rates among FRL and White students in DPS and the comparison group of lowest- performing districts.

Table 11.3
Pre-Reform Tests of Parallel Trends with DPS for Subgroup Analyses

	EL	FRL	Race			Special Education
			Black	Hispanic	White	
ELA	0.057	-0.025	-0.014	0.006	0.028	0.002
Math	0.061	-0.005	-0.019	0.031	0.036	0.036
Graduation rate	0.081	0.065*	0.038	0.026	0.114***	0.072

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

12 RESULTS

Overall

Table 12.1 reports the overall impact of DPS reform efforts on district-level ELA and math test scores compared with the three comparison groups. The table presents the differences in pre- and post-reform growth trajectories for DPS and the comparison groups, and then the rows indicating treatment effects report the DPS reform impact. The models in columns (2) and (4) control for district-level demographics, while the models in columns (1) and (3) do not. Models (2) and (4) are our preferred models.

Table 12.1
CITS Estimate of Impact of DPS Reform Efforts on Overall ELA and Math Performance

Largest Districts	ELA		Math	
	(1)	(2)	(3)	(4)
DPS treatment effect	0.078***	0.079***	0.076***	0.070***
	(0.006)	(0.012)	(0.005)	(0.009)
Pre-reform parallel trajectory	Yes	Yes	Yes	Yes
Lowest Performing Districts	ELA		Math	
	(1)	(2)	(3)	(4)
DPS treatment effect	0.064***	0.051***	0.061***	0.052***
	(0.006)	(0.008)	(0.006)	(0.007)
Pre-reform parallel trajectory	Yes	Yes	Yes	Yes
Aurora Public Schools (APS)	ELA		Math	
	(1)	(2)	(3)	(4)
DPS treatment effect	0.049***	0.141	0.092***	0.195***
	(0.010)	(0.017)	(0.008)	(0.022)
Pre-reform parallel trajectory	No	Yes	No	Yes
District & year fixed effects	Yes	Yes	Yes	Yes
District controls	No	Yes	No	Yes

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

When we compare DPS with the largest districts and the lowest-performing districts comparison groups, the DPS reform efforts resulted in positive and statistically significant average annual improvements for student performance in both ELA and math. The CITS treatment effect results show that the DPS reform efforts improved overall student performance by 0.079 and 0.070 standard deviations per year in ELA and math, respectively, when compared with the performance of students in the largest districts in the state. Over the course of 11 years of reform, this translates to improvements of 0.869 and 0.770 standard deviations in ELA and math, respectively, for DPS students—effectively closing the gap with the state averages.

The improvements on state assessment were slightly smaller when we look at DPS compared with the lowest-performing districts in the state, where DPS started before the reform efforts. The CITS treatment effect results show that the effect of DPS reform efforts, when compared with the lowest- performing districts, was an average of 0.051 and 0.052 standard deviations annually in ELA and math, respectively. Over the course of 11 years of reform, this translates to an improvement of 0.561 standard deviations in ELA and 0.572 standard deviations in math for DPS students.

When we compare DPS results with APS, the impact of DPS reform efforts was positive in ELA and math but not statistically significant in the preferred models.

To determine whether the increase in DPS performance that occurred when the assessment changed in 2015 is responsible for the overall findings, we conduct a sensitivity test and run the overall CITS models through only 2014. The results are presented in Table C1 in Appendix C. For both the largest and lowest-performing comparison groups, the findings remain positive and statistically significant, though smaller in magnitude through 2014 before the state assessment changed. This suggests that the improvements DPS students made on CMAS are not responsible for the overall results.

We also examine whether the results differed by the grades served by the district's schools (see Table C2 in Appendix C). The impact of DPS reform efforts do not appear to be limited to one type of school (elementary or middle) over another. Compared with similar grade configurations in the largest and lowest-performing districts, the DPS reform efforts resulted in positive and statistically significant results for both elementary and middle schools.

Finally, we compare our CITS analyses to DID, our graduation rate modeling approach. DID models including the same district-level covariates and district and year fixed effects as the CITS models produce the results presented in Table C3 in Appendix C. They do not give us reason to question our CITS results.

Figures 12.2 and 12.3 show how the annual positive results from DPS's reform efforts compounded over time such that DPS closed the academic performance gap with the largest school districts and exceeded the performance of the districts that, along with DPS, performed in the lowest 20th percentile in the pre-reform years.

Table 12.4 reports the overall impact of DPS reform efforts on district-level graduation rates. The results include only the large district comparison group. The table presents the treatment effect as percentage points expressed in decimal form, along with summary information about each model. As with the CITS analyses, our preferred model, in column (2), controls for district-level demographics, while the model in column (1) does not.

Table 12.4
DID Estimate of Impact of DPS Reform Efforts on Overall Graduation Rates

Largest Districts	Graduation Rates (1)	(2)
Treated*Post	0.150***	0.146*
	(0.012)	(0.064)
R-squared	0.92	0.93
Observations	121	121
N (districts)	11	11
District & year fixed effects	Yes	Yes
District controls	No	Yes

*** p < 0.001; ** p < 0.01; * p < 0.05

These results support a more intuitive interpretation than the standard deviation effects found for ELA and Math: Graduation rates in DPS would undoubtedly have climbed from the rate of 38.9% without the introduction of the reform effort, but instead of reaching 70.9% by 2019, the graduation rate may have remained below 60%.

Figure 12.2
CITS Estimate of Overall Outcome Trends by DPS and Comparison Group: ELA

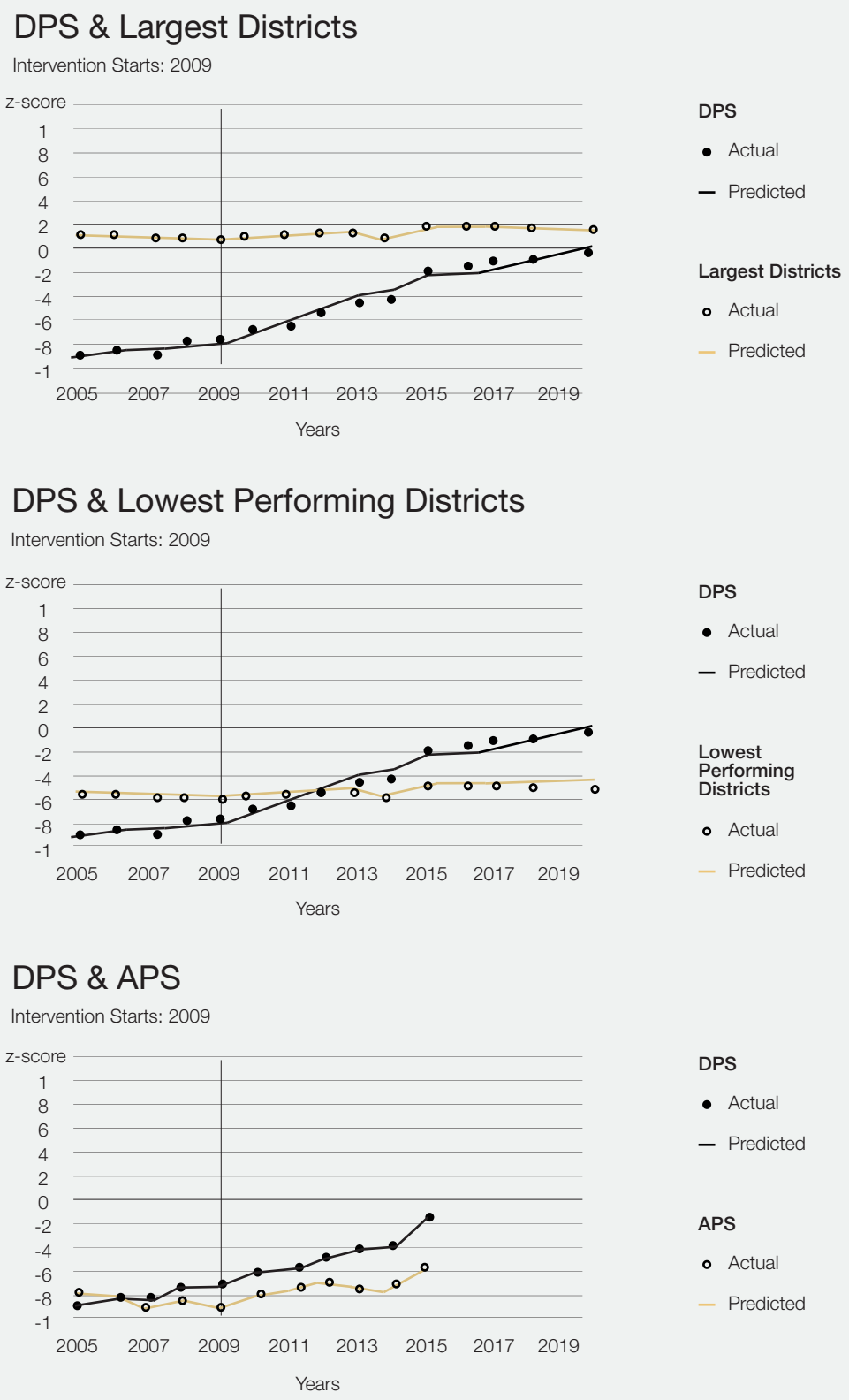
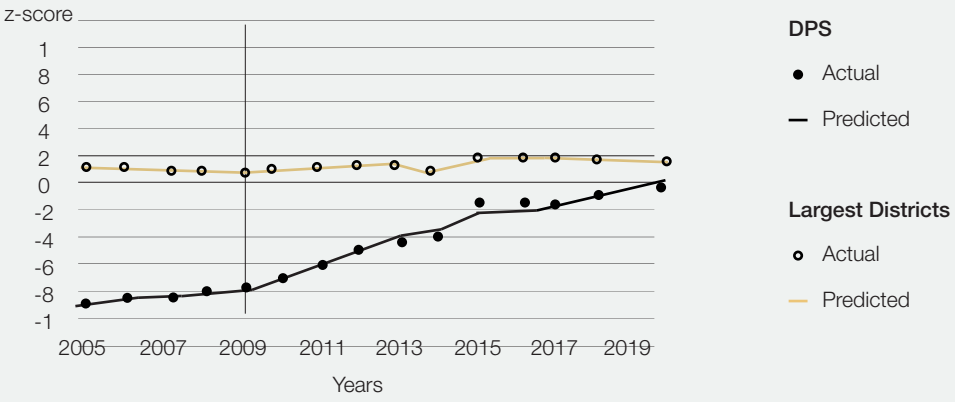


Figure 12.3
CITS Estimate of Overall Outcome Trends by DPS and Comparison Group: Math

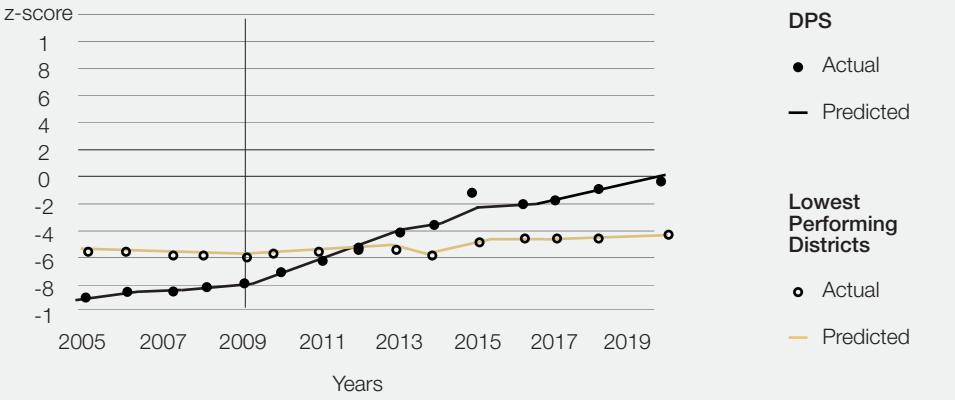
DPS & Largest Districts

Intervention Starts: 2009



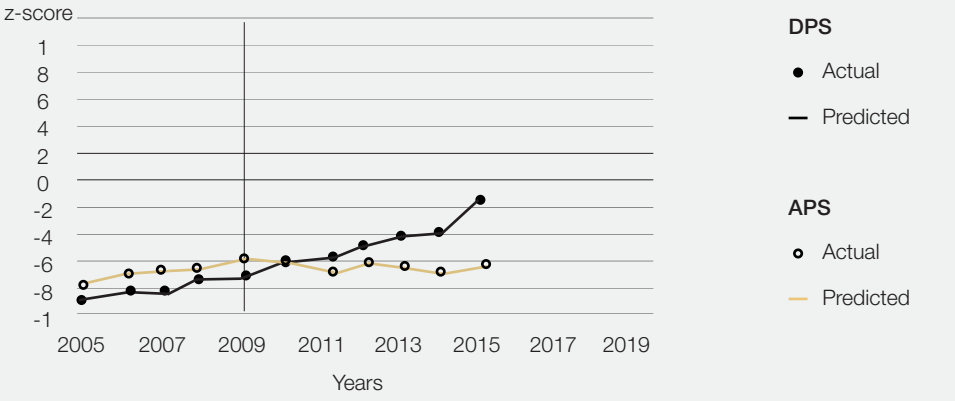
DPS & Lowest Performing Districts

Intervention Starts: 2009



DPS & APS

Intervention Starts: 2009



Subgroups

Table 12.A reports CITS treatment effect results from models that examine the impact of DPS reform efforts on student subgroups defined by program participation (EL, FRL and Special Education). Our preferred models include district-level controls and are presented in columns (2), (4), (6), (8), (10), and (12). The models for each subgroup compare DPS's students in that subgroup to students in the comparison group from the same subgroup. For example, the results for FRL students compare the academic performance of DPS's FRL students to FRL students in the subset of lowest-performing districts included in the FRL comparison group. The subgroup analyses allow us to examine whether subgroups experienced the same or different results from DPS's reform efforts.

Among the preferred models, the only statistically significant effect shows that the impact of DPS reform efforts on Special Education students in DPS compared with Special Education students in the comparison group was an annual average of 0.031 standard deviations in math. Over the course of the reform period, this amounts to an effect of 0.341 standard deviations. Figures 12.C and 12.D show the relative trends, incorporating both actual and model-predicted values for each subgroup, including those that do not demonstrate a statistically significant reform effect.

Table 12.B presents the DID graduation results for the EL and Special Education subgroups. The FRL subgroup is excluded because it failed the parallel trends test for all possible comparison groups. Here, the preferred models with district-level controls are presented in columns (2) and (4). As was the case for ELA and math, the only statistically significant result is for students who receive special education services. In their case, the reforms are associated with a 12.2 percentage point increase in graduation rates.

Table 12.A
CITS Estimate of Impact of DPS Reform Efforts on Student Performance by Subgroups

EL Students	ELA		Math	
	(1)	(2)	(3)	(4)
DPS treatment effect	0.066***	0.016	0.057***	0.023
	(0.014)	(0.020)	(0.012)	(0.014)
Pre-reform parallel trajectory	Yes	Yes	Yes	Yes
District & year fixed effects	Yes	Yes	Yes	Yes
District controls	No	Yes	No	Yes

FRL Students	ELA		Math	
	(5)	(6)	(7)	(8)
DPS treatment effect	0.046***	0.018	0.046***	0.015
	(0.009)	(0.012)	(0.008)	(0.012)
Pre-reform parallel trajectory	Yes	Yes	Yes	Yes
District & year fixed effects	Yes	Yes	Yes	Yes
District controls	No	Yes	No	Yes
Special Education Students	**ELA**		**Math**	
	(9)	(10)	(11)	(12)
DPS treatment effect	0.022***	0.011	0.037***	0.031**
	(0.006)	(0.009)	(0.008)	(0.011)
Pre-reform parallel trajectory	Yes	Yes	Yes	Yes
District & year fixed effects	Yes	Yes	Yes	Yes
District controls	No	Yes	No	Yes

*** p < 0.001; ** p < 0.01; * p < 0.05

Table 12.B
DID Estimate of Impact of DPS Reform Efforts on Graduation Rate

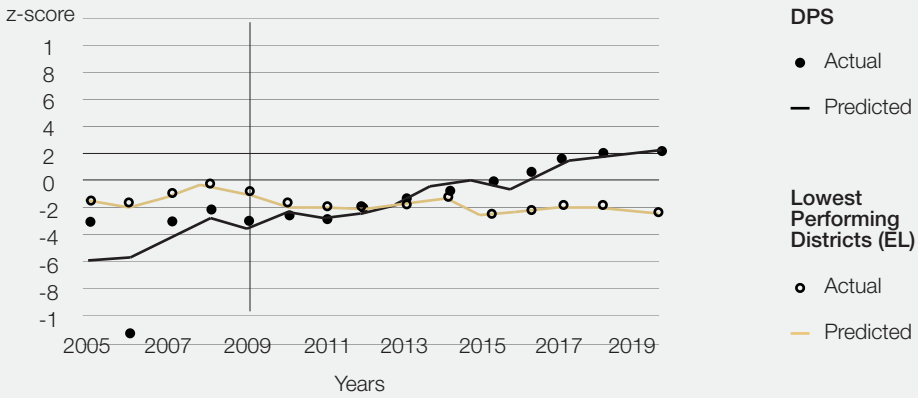
	EL		Special Education	
	(1)	(2)	(3)	(4)
Treated*Post	0.107***	0.123	0.008	0.122**
	(0.032)	(0.083)	(0.031)	(0.034)
R-squared	0.57	0.63	0.48	0.52
Observations	130	130	104	103
N (district)	10	10	8	8
District & year fixed effects	Yes	Yes	Yes	Yes
District controls	No	Yes	No	Yes

*** p < 0.001; ** p < 0.01; * p < 0.05

Figure 12.C
CITS Estimate of Subgroup Outcome Trends by DPS and Comparison Group: ELA

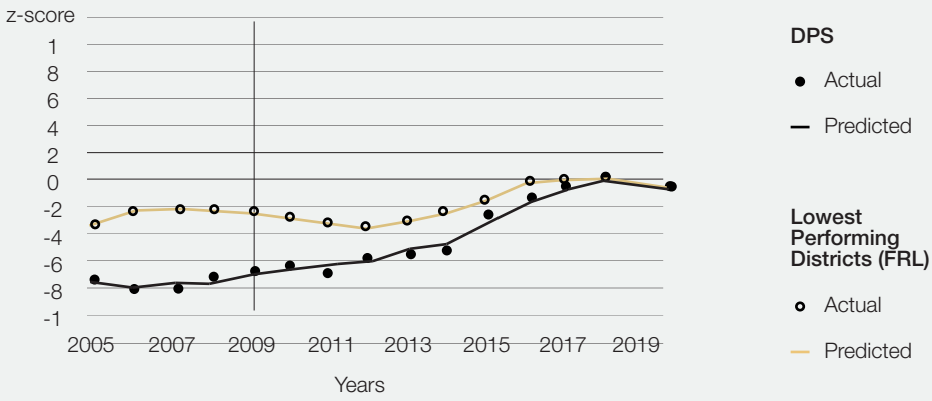
DPS & Lowest Performing Districts (EL)

Intervention Starts: 2009



DPS & Lowest Performing Districts (FRL)

Intervention Starts: 2009



DPS & Lowest Performing Districts (Special Education)

Intervention Starts: 2009

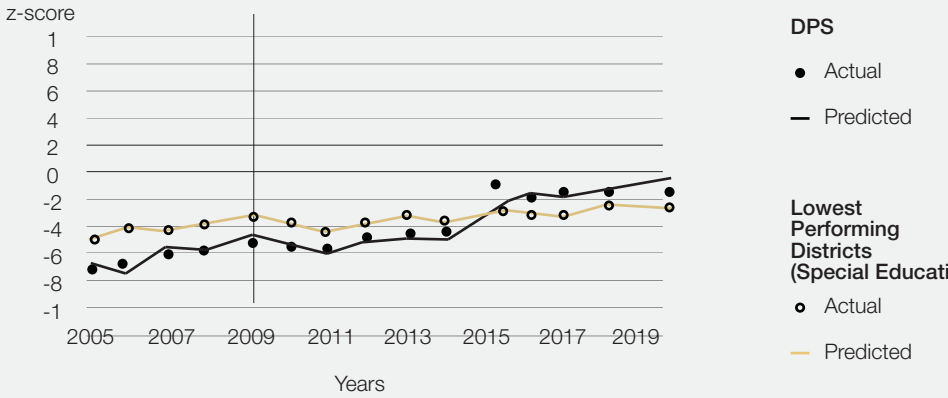
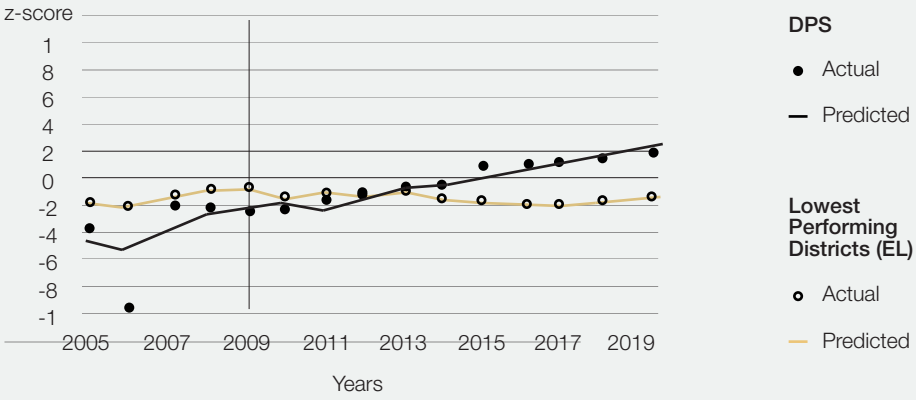


Figure 12.D
CITS Estimate of Subgroup Outcome Trends by DPS and Comparison Group: Math

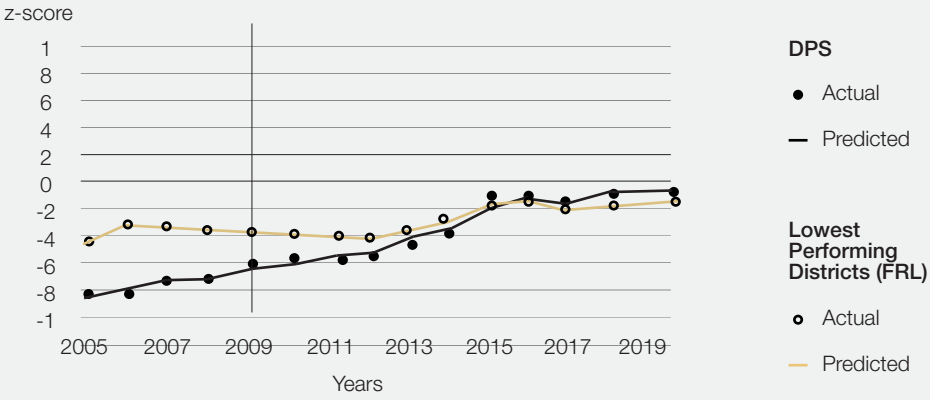
DPS & Lowest Performing Districts (EL)

Intervention Starts: 2009



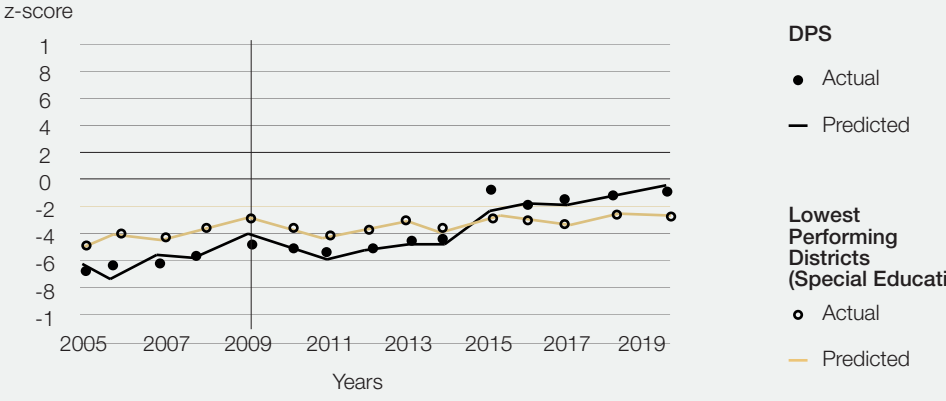
DPS & Lowest Performing Districts (FRL)

Intervention Starts: 2009



DPS & Lowest Performing Districts (Special Education)

Intervention Starts: 2009



When we look at the CITS treatment effect results by race/ethnicity subgroups (see Table 12.E, again with preferred models in even-numbered columns), DPS reform efforts led to statistically significant academic improvements for Black students in math and White students in ELA and math. Over the 11-year reform period, these effects amount to 0.484 standard deviations of improvement for Black students in math and 0.374 and 0.506 standard deviations of improvement for White students in ELA and math, respectively. Figures 12.G and 12.H show the relative trends for this collection of subgroups, demonstrating how the reform treatment effect led Black DPS students to close the gap with the comparison group in math, while White students widened the performance gap with comparison district students in both ELA and math.

Table 12.F presents the results of the DID analyses for race/ethnicity subgroup graduation rates. The White subgroup is excluded because it failed the parallel trends test for all possible comparison groups. Neither of the preferred models, found in columns (2) and (4), returned a statistically significant result.

Table 12.E
CITS Estimate of Impact of DPS Reform Efforts on Student Performance by Race/Ethnicity Subgroups

Black Students	ELA		Math	
	(1)	(2)	(3)	(4)
DPS treatment effect	0.008	0.017	0.027**	0.044**
	(0.010)	(0.019)	(0.008)	(0.012)
Pre-reform parallel trajectory	Yes	Yes	Yes	Yes
District & year fixed effects	Yes	Yes	Yes	Yes
District controls	No	Yes	No	Yes
Hispanic Students	**ELA**		**Math**	
	(5)	(6)	(7)	(8)
DPS treatment effect	0.047***	0.016	0.044***	0.014
	(0.009)	(0.016)	(0.010)	(0.014)
Pre-reform parallel trajectory	Yes	Yes	Yes	Yes
District & year fixed effects	Yes	Yes	Yes	Yes
District controls	No	Yes	No	Yes

White Students	ELA		Math	
	(9)	(10)	(11)	(12)
DPS treatment effect	0.057***	0.034**	0.063***	0.046**
	(0.011)	(0.012)	(0.012)	(0.016)
Pre-reform parallel trajectory	Yes	Yes	Yes	Yes
District & year fixed effects	Yes	Yes	Yes	Yes
District controls	No	Yes	No	Yes

*** p < 0.001; ** p < 0.01; * p < 0.05

Table 12.F
DID Estimate of Impact of DPS Reform Efforts on Graduation Rate by Subgroups

	Black		Hispanic	
	(1)	(2)	(3)	(4)
Treated*Post	0.083	0.074	0.082***	-0.008
	(0.031)	(0.361)	(0.020)	(0.039)
R-squared	0.95	0.98	0.69	0.73
Observations	39	39	156	155
N (district)	3	3	12	12
District & year fixed effects	Yes	Yes	Yes	Yes
District controls	No	Yes	No	Yes

*** p < 0.001; ** p < 0.01; * p < 0.05

Figure 12.G
CITS Estimate of Race/Ethnicity Subgroup Outcome Trends by DPS and Comparison Group: ELA

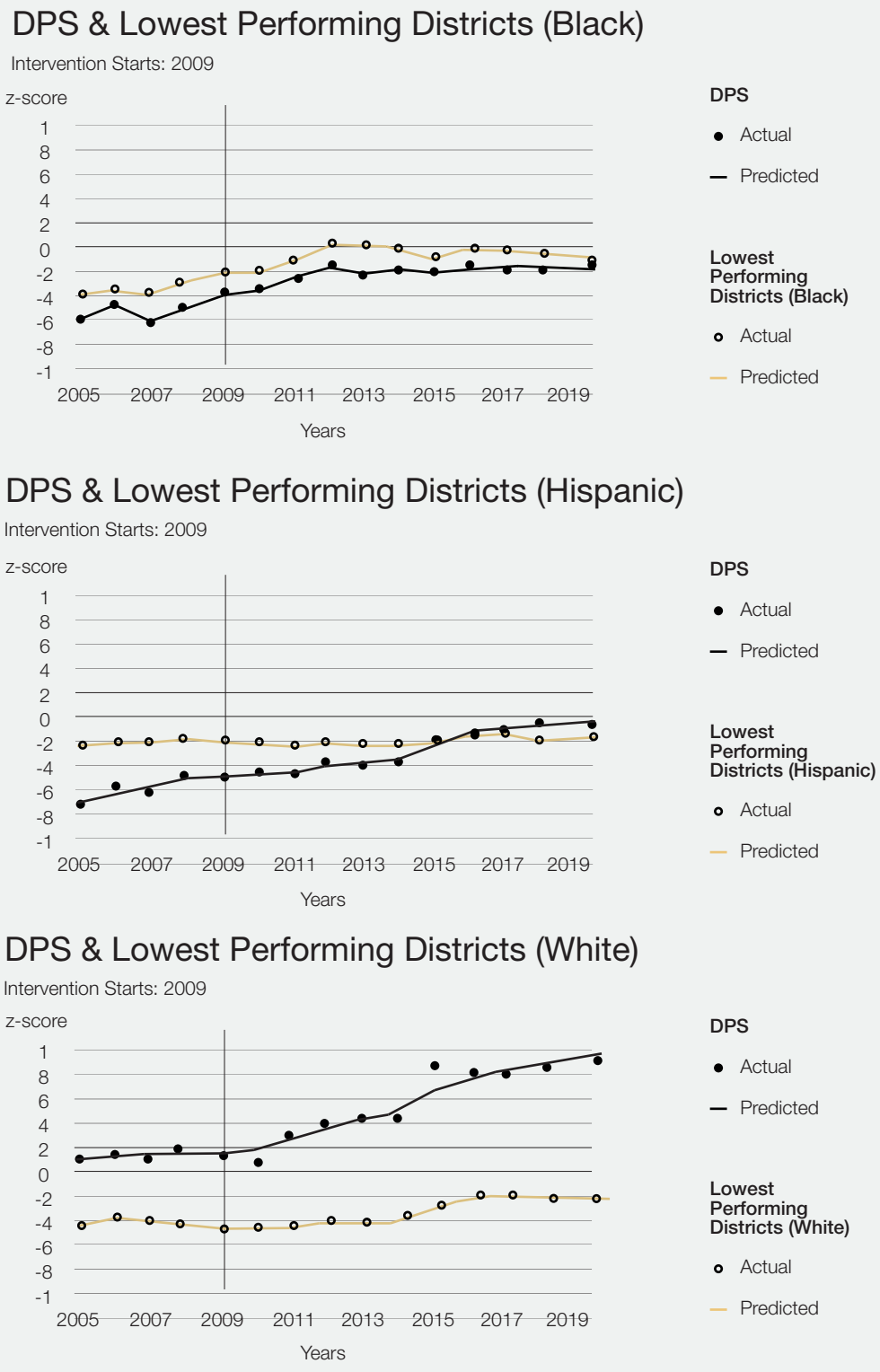
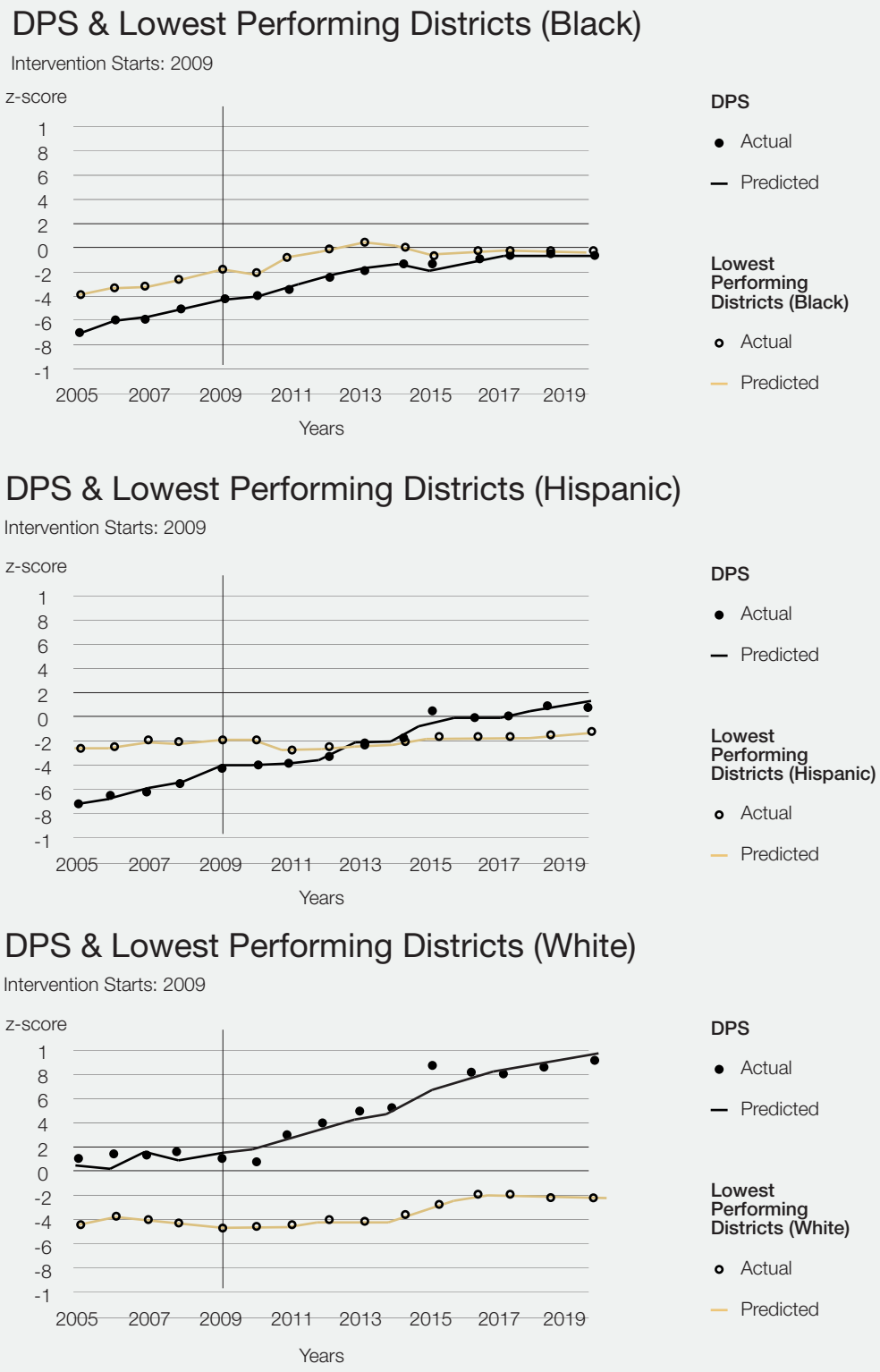


Figure 12.H
CITS Estimate of Race/Ethnicity Subgroup Outcome Trends by DPS and Comparison Group: Math



Conclusion

The reforms implemented by DPS over the last decade have been controversial. Some individuals in Denver have experienced the reforms as a loss, while others have reaped their rewards. In the political arena, "whether the reforms helped or hurt is fiercely debated" (Asmar, 2021).

This study provides the first comprehensive, system-level answer to the question of whether Denver's reform strategy led to improved outcomes for students. We find that the strategy did improve overall outcomes for students in terms of graduation rates and achievement in both ELA and math throughout its implementation. We show that the improvements seen are the effects of the reform strategy and are not due to changes in the demographics of the student population or the city at-large or from other exogenous factors. In addition, we uncover no negative effects for students who identify as Black or Hispanic or participate in EL, FRL, or Special Education programs. These findings are important not only because they speak directly to claims made by proponents and opponents of the reforms in Denver and of similar strategies in other cities, but also because system-wide improvement in student outcomes of this size and duration is rare. In future research, we will explore the components of Denver's reform strategy and the effects of specific interventions on overall improvement and for specific populations of students.

13 REFERENCES

Angrist, J.D., Abdulkadiroğlu, A., Dynarski, S.M., Kane, T.J. & Pathak, P.A. (2011a). Accountability and flexibility in public schools: Evidence from Boston's charters and pilots. The Quarterly Journal of Economics, 126(2), 699–748.

Angrist, J.D., Cohodes, S.R., Dynarski, S.M., Pathak, P.A. & Walters, C.R. (2016). Stand and deliver: effects of Boston's charter high schools on college preparation, entry and choice. Journal of Labor Economics, 23 (2).

Angrist, J.D., Pathak, P. & Walters, C.R. (2011b). Explaining charter school effectiveness (Working Paper No. 17332). Cambridge, MA: National Bureau of Economic Research.

Ahn, S., Myers, N. D. & Jin, Y. (2012). Use of the estimated intraclass correlation for correcting differences in effect size by level. Behavior Research Methods, 44(2), 490-502.

Asmar, M. (2021). What's at stake in this year's Denver school board election. Chalkbeat (2021).

Baxter, P. & Gottlieb, A. (2022). Dismantling Denver. Education Next, 22(2).

Baxter, P., Teske, P. & Ely, T. (2019). Redesigning Denver's schools. Education Next 19(2).

Baxter, P., Ely, T. L. & Teske, P. (2018). A bigger slice of the money pie: Charters in Colorado and Florida win share of local tax dollars. Education Next, 18(2).

Baxter, P. (2012a). For charter schools and school districts: Empty space equals opportunity. In Lake, R. J. & Gross, B. (Eds.), Hopes, fears, & reality: A balanced look at American charter schools in 2011. Center on Reinventing Public Education.

Baxter, P. & Nelson, E.C. (2012b). Mastering change: When charter schools and school districts embrace strategic partnership. In Lake, R. J. & Gross, B. (Eds.), Hopes, fears, & reality: A balanced look at American charter schools in 2011. Center on Reinventing Public Education.

Berends, M., Berends, A. P. & Springer, M. G. (2020). Handbook of research on school choice (Second ed.). Routledge.

Blazar, D. & Schueler, B. (2022). Why do school districts matter? An interdisciplinary framework and empirical review (EdWorkingPaper No. 22-581). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/58m4-fs65.

Bobbitt, F. (Ed.). (1916). Report of the school survey of school district number one in the City and County of Denver. Denver, CO.

Bulkley, K. E. & Hashim, A. K. (2020). Portfolio management models. In J. T. Guthrie, A. Primus, & M. G. Springer (Eds.), Handbook of research on school choice (2nd ed.). Taylor and Francis Group.

Bulkley, K. (2010). Introduction: Portfolio management models in urban school reform. In Bulkley, K., Henig, J. & Levin, H. (Eds.), Between public and private: Politics, governance and the new portfolio models for urban school reform (pp. 3-26). Cambridge, MA: Harvard Education Press.

Bulkley, K. E., Marsh, J. A., Strunk, K. O., Harris, D. N. & Hashim, A. K. (2021). Challenging the one best system: The portfolio management model and urban school governance. Cambridge, MA: Harvard Education Press.

Callahan, R. E. (1962). Education and the cult of efficiency: A study of the social forces that have shaped the administration of the public schools. Chicago, IL: University of Chicago Press.

Chen, F. & Harris, D. (2022). How do charter schools affect system-level test scores and graduation rates? A national analysis (EdWorkingPaper No. 22-525). Retrieved from Annenberg Institute at Brown University: https://www.edworkingpapers.com/ai22-525.

Chin, M. & Kane, T. (2017). School district reform in Newark: Within- and between-school changes in achievement growth (NBER Working Paper No. 23922).

Chubb, J. E. & Moe, T. M. (1990). Politics, markets, and America's schools. Washington, D.C.: Brookings Institution.

Cubberley, E. P. (1916). Supplemental report on the organization and administration of School District Number One in the City and County of Denver. Denver, CO.

DeBray, E. & Jabbar, H. (2013). Review of two presentations on the portfolio school model. Boulder, CO: National Education Policy Center.

Denver Public Schools. (2022a). Authorizing & accountability: Family of schools. Denver Public Schools. Accessed April 29, 2022 from https://portfolio.dpsk12.org/dps-family-of-schools/#1528485642610-2495946f-25b3.

Denver Public Schools. (2022b). Facts & figures: DPS by the numbers. Denver Public Schools. Accessed April 29, 2022 from https://www.dpsk12.org/about/facts-figures/#students.

Fahle, E. M. & Reardon, S. F. (2018). How much do test scores vary among school districts? New estimates using population data, 2009–2015. Educational Researcher, 47(4), 221-234.

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

Gamson, D.A. (2004). "The infusion of corporate values into progressive education: Professional vulnerability or complicity?" Journal of Educational Administration, 42 (2), 137-159.

Glanz, J. (1991). Bureaucracy and professionalism: the evolution of public school supervision. United Kingdom: Fairleigh Dickinson University Press.

Gross, B. & Lake, R. (2011). Reforming districts through choice, autonomy, equity, and accountability: An overview of the voluntary public school choice directors meeting. Center on Reinventing Public Education.

Hallberg, K., Williams, R., Swanlund, A. & Eno, J. (2018). Short comparative interrupted time series using aggregate school-level data in education research. Educational Researcher, 47(5), 295-306.

Hallberg, K., Williams, R. & Swanlund, A. (2020). Improving the use of aggregate longitudinal data on school performance to assess program effectiveness: Evidence from three within study comparisons. Journal of Research on Educational Effectiveness, 13(3), 518-545.

Harris, D. & Witte, J. (2011). The market for education. In D.E. Mitchell, R. Crowson, and D. Shipps (Eds.), Shaping Education Policy: Power and Process. New York: Routledge.

Harris, D. & Larsen, M. (2016). The effects of the New Orleans post-Katrina school reforms on student academic outcomes (Working Paper). Research Alliance for New Orleans.

Harris, D. & Larsen, M. (2018). The effects of Hurricane Katrina on medium term student outcomes in New Orleans. Research Alliance for New Orleans. https://educationresearchalliancenola.org/files/publications/Harris-Larsen-Reform-Effects-2021-05-17.pdf

Harris, D. & Larsen, M. (2022). Taken by Storm: The Effects of Hurricane Katrina on Medium-Term Student Outcomes in New Orleans. J. Human Resources. Published online before print April 5, 2022. doi:10.3368/jhr.58.5.0819-10367R2.

Hill, P. T., Campbell, C. & Harvey, P. (2000). It Takes a City: Getting Serious about Urban School Reform. Brookings Institution Press.

Hill, P. T., Campbell, C., Menefee-Libey, D., Dusseault, B., DeArmond, M. & Gross, B. (2009). Portfolio school districts for big cities: An interim report (Portfolio School Districts Project). Seattle, WA: Center on Reinventing Public Education.

Hill, P. & Campbell, C. (2011). Growing number of districts seek bold change with portfolio strategy. University of Washington: Center for Reinventing Public Education.

Hill, P. & Lake, R. (2004). Charter schools and accountability in public education. Washington, D.C.: Brookings Institution Press.

Hill, P. T., Pierce, L. C. & Guthrie, J. W. (1997). Reinventing public education: How contracting can transform America's schools. Chicago, IL: University of Chicago Press.

Honig, M. I. & Rainey, L. R. (2012). Autonomy and school improvement: What do we know and where do we go from here? Educational Policy, 26(3), 465–495.

Honig, M. I. & DeArmond, M. (2010). Where's the management in portfolio management? In Bulkley, K., Henig, J., Levin, H. (Eds.), Between public and private: Politics, governance and the new portfolio models for urban school reform (pp. 195–216). Cambridge, MA: Harvard Education Press.

Hoxby, C. (2000). Does competition among public schools benefit students and taxpayers? American Economic Review 90(5), 1209-1238.

Hoxby, C. M. (2003). School choice and school productivity: Could school choice be a tide that lifts all boats? In Hoxby, C. M. (Ed.), The Economics of School Choice. National Bureau of Economic Research.

Jacob, R., Somers, M. A., Zhu, P. & Bloom, H. (2016). The validity of the comparative interrupted time series design for evaluating the effect of school-level interventions. Evaluation Review, 40(3), 167-198.

Ladd, H.F. (2018). "Self-Governing schools, school choice, and the public interest." In Berends, M., Waddington, R. J., & Schoenig, J. (2018). School choice at the crossroads: Research perspectives. Routledge.

Lake, R. J. & Hill, P. T. (2009). Performance management in portfolio school districts. Center on Reinventing Public Education.

Lipsey, M. W., Puzio, K., Yun, C., Hebert, M.A., Steinka-Fry, K., Cole, M. W., Roberts, M., Anthony, K. S. & Busick, M. D. (2012). Translating the statistical representation of the effects of education interventions into more readily interpretable forms. (NCSER 2013-3000). Washington, DC: National Center for Special Education Research, Institute of Education Sciences, U.S. Department of Education. Retrieved from the IES website at http://ies.ed.gov/ncser/.

Marsh, J. A., Albright, T. N., Brown, D., Bulkley, K. E., Strunk, K. O. & Harris, D. N. (2020). The process and politics of educational governance change in New Orleans, Los Angeles and Denver. American Educational Research Journal, 58(1), 107–159.

Marsh, J. A., Strunk, K. O., Bush-Mecenas, S. C. & Huguet, A. (2015). Democratic engagement in district reform: The evolving role of parents in the Los Angeles public school choice initiative. Educational Policy, 29(1), 51–84.

Marsh, J. A., Hashim, A. K., Daramola, E. J. & Mulfinger, L. (2020). Chapter 6: Autonomy: Flexibility in response to student needs. In K. E. Bulkley, J. A. Marsh, K. O. Strunk, D. H. Harris, & A. K. Hashim (Eds.), Challenging the one best system: The portfolio management model and urban school governance (pp. 121–148). Cambridge, MA: Harvard Education Press.

McEachin, A. J., Welsh, R. O. & Brewer, D. J. (2016). The variation in student achievement and behavior within a portfolio management model: Early results from New Orleans. Education Finance and Policy, 11, 150–176.

Menefee-Libey, D. (2010). Neoliberal school reform in Chicago? In K. E. Bulkley, J. R. Henig, H. M. Leven (Eds.), Between public and private: Politics, governance, and the new portfolio models for urban school reform (pp. 55–90). Cambridge, MA: Harvard Education Press.

St. Clair, T., Hallberg, K. & Cook, T. D. (2016). The validity and precision of the comparative interrupted time-series design: three within-study comparisons. Journal of Educational and Behavioral Statistics, 41(3), 269-299.

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

Tyack, D. (1974). The one best system: A history of American urban education. Cambridge, MA: Harvard University Press.

Winters, M. (2012). Measuring the effect of charter schools on public school student achievement in an urban environment: Evidence from New York City. Economics of Education Review, 31, 293–301.

14 APPENDIX (A - B - C)

A - DATA

Demographics

District-level demographic data[2] include the total student enrollment by race/ethnicity (White, Black, Hispanic, and Other), those qualifying for free or reduced-price lunch (FRL), English Learners (EL), and Special Education. We calculated the percent of students in each demographic category by dividing the subgroup enrollment by total enrollment.

State Standardized Assessments

During the time frame for this study, CDE administered three state standardized assessments to students in grades 3-10: Colorado Student Assessment Program (CSAP), Transitional Colorado Assessment Program (TCAP), and Colorado Measures of Academic Success (CMAS). CSAP was administered in the spring of 2005 through 2011, TCAP in the spring of 2012 through 2014, and CMAS in the spring of 2015 through 2019. We downloaded district-level reading, writing, and math CSAP and TCAP test score outcomes through CDE's historical SchoolVIEW Data Lab[3] and English Language Arts (ELA) and math CMAS test score outcomes through CDE's SchoolVIEW State Assessment Data Lab[4].

For each subject area, the state standardized assessment data include weighted scale scores for the overall district (across all tested grade levels), as well as by subgroups including grade configuration (elementary, middle, high), race/ethnicity, FRL, EL, and Special Education. The CSAP and TCAP assessments included reading and writing assessments, while the CMAS assesses English Language Arts in one test. In order to standardize the ELA assessments over time, we created a weighted average scale score from the reading and writing scores for CSAP and TCAP. For each year and subgroup we averaged reading and writing scores by test-taker count weights.

(2) https://www.cde.state.co.us/cdereval/pupilcurrent
(3) https://datalab.cde.state.co.us/cognos/cde/datalabreport.htm
(4) https://datalab.cde.state.co.us/cognos/cde/datalabstatereport.htm

Our analyses require that districts have data for every year between 2005 and 2019. As a result, we excluded seven Boards of Cooperative Educational Services (BOCES) entities and five school districts that did not have complete data for the study.

After excluding BOCES and districts without complete data, we created longitudinal district-level academic data for our analyses by standardizing the weighted scale scores by year, subject area, and subgroup within Colorado. The z-scores have a statewide mean of zero and standard deviation of one by year, subject area, and subgroup.

Given that we are using district-level average performance from the state standardized assessment for our outcome variables, the z-scores reflect variance between districts, rather than between schools or students. Because this variance is relatively smaller, the district z-scores are typically larger than estimates standardized on variation between individuals (Blazar & Schueler, 2022; Lipsey et al., 2012). Analyses using unadjusted district-level z-scores would then produce misleadingly large effect sizes. The district interclass correlation (ICC) can be used to address this issue by scaling district z-scores to reflect variation by individuals (Ahn, Myers, & Jin, 2012). Using national data, Fahle and Reardon (2018) estimate district intraclass correlations (ICC) for ELA and math in all states, and their ICC range for Colorado is 0.100 to 0.125 for both ELA and math. We transform district z-scores by multiplying them by the square root (ICC is based on variance, not standard deviation) of the high range of the Colorado ICC ($\sqrt{0.125} = 0.35$). This reduces the magnitude of the district z-scores to reflect z-scores more equivalent to what we would expect if we were using student-level data. It also rescales the z-scores to have a statewide mean of zero and a standard deviation of 0.35 by year, subject, and subgroup. This transformation does not affect the statistical significance of the findings but provides a more conservative magnitude for the estimates.

High School Assessments

CDE administered the ACT[5] to 11th graders between 2007 and 2016 and the SAT[6] to 11th graders between 2017 and 2019. We have two fewer years of high school assessment data compared with the state standardized assessment data, but sufficient data to analyze pre-reform and reform performance trends. We created a longitudinal district-level data file by standardizing the ACT average composite score and the SAT overall mean score by year within Colorado. We also correct the ACT/SAT z-scores by the same district ICC described above. After the ICC correction, the z-scores have a statewide mean of zero and standard deviation of 0.35 by year.

(5) https://www.cde.state.co.us/assessment/coact
(6) https://www.cde.state.co.us/assessment/sat-psat-data

Graduation Rates

CDE reports adjusted cohort graduation rates (ACGR) for four- to seven-year periods. The ACGR is calculated by first identifying cohorts of first-time ninth-grade students. In Colorado, these are called anticipated year of graduation (AYG) cohorts. The ACGR is calculated by dividing the number of students in an AYG cohort who receive regular high school diplomas by the number of students who entered the ninth-grade cohort adjusted by adding students who transfer into the cohort after ninth grade and subtracting students with a verified transfer out of the cohort. CDE officially began calculating the ACGR in 2009-10[7] and retroactively applied the method to data from 2006-07 through 2008-09[8]. For this study, we analyze on-time, four-year ACGR. We have two fewer years of graduation rate data compared with the state standardized assessment data, but sufficient data to analyze pre-reform and reform performance trends.

(7) https://www.cde.state.co.us/cdereval/rvprioryeargraddata
(8) https://www.cde.state.co.us/cdereval/gradhistrates

B - SUBGROUP ANALYSES COMPARISON GROUPS

Table B1

CITS Comparison Groups for Subgroup Analyses, Pre-Reform Characteristics, 2004-05

		Lowest Performing Districts					
	DPS	EL	FRL	Race			Special Educ.
				Black	Hispanic	White	
Number of Districts	1	9	14	2	14	11	13
Mean Enrollment	72,412	7,204	2,750	21,479	2,523	6,782	8,327
% EL	29.95	35.32	16.09	25.50	20.02	25.55	15.40
% FRL	61.52	59.29	69.06	54.54	66.12	58.91	59.50
% Race							
% Black	18.98	3.83	2.93	23.19	1.32	5.52	6.34
% Hispanic	57.34	58.11	62.52	39.26	67.80	55.37	48.51
% White	19.37	31.86	31.93	31.52	28.32	31.73	42.38
% Special Education	11.81	11.02	9.95	10.56	9.93	11.28	11.13
Mean ELA z-score	-0.831	-0.146	-0.327	-0.379	-0.225	-0.412	-0.377
Mean Math z-score	-0.845	-0.184	-0.419	-0.373	-0.274	-0.429	-0.471

Table B2

DID Comparison Groups for Subgroup Analyses, Pre-Reform Characteristics, 2006-07

		Lowest Performing Districts					
				Race			
	DPS	EL	FRL	Black	Hispanic	White	Special Educ.
Number of Districts	1	9	10	2	11	11	7
Mean Enrollment	72,561	7,327	3,817	22,498	3,117	6,964	9,781
% EL	31.63	37.16	21.51	25.72	26.12	27.45	31.63
% FRL	64.47	62.15	68.44	59.99	64.45	60.80	64.47
% Race							
% Black	17.84	3.69	3.80	22.32	1.50	5.46	17.84
% Hispanic	57.46	61.60	66.76	43.05	67.45	58.40	57.46
% White	20.45	28.72	26.33	28.78	28.07	28.90	20.45
% Special Education	12.06	10.10	9.92	10.47	10.38	11.04	12.06
Graduation Rate							
EL	31.20	45.04					
FRL	35.70		58.56				
Black	43.00			60.95			
Hispanic	30.40				50.21		
White	53.10					68.95	
Special Education	41.43						25.60

C - ADDITIONAL RESULTS

Table C1

DID Estimate of Impact of DPS Reform Efforts on Overall ELA and Math Performance

	ELA		Math	
	(1)	(2)	(3)	(4)
Largest Districts				
Treated*Post	0.468***	0.111	0.490***	0.110
	(0.030)	(0.587)	(0.027)	(0.155)
R-squared	0.95	0.96	0.96	0.97
Observations	165	165	165	165
N (district)	11	11	11	11
Lowest-Performing Districts				
Treated*Post	0.357***	0.207***	0.394***	0.278***
	(0.038)	(0.059)	(0.032)	(0.060)
R-squared	0.62	0.69	0.66	0.70
Observations	465	452	465	452
N (district)	30	30	30	30
District & year fixed effects	Yes	Yes	Yes	Yes
District controls	No	Yes	No	Yes

*** p < 0.001; ** p < 0.01; * p < 0.05

Table C2

CITS Estimate of Impact of DPS Reform Efforts on Student Performance, Overall, 2005-2014

	ELA		Math	
	(1)	(2)	(3)	(4)
Largest Districts				
DPS treatment effect	0.059***	0.039***	0.068***	0.061***
	(0.004)	(0.009)	(0.004)	(0.011)
Pre-reform parallel trajectory	Yes	Yes	Yes	Yes

*** p < 0.001; ** p < 0.01; * p < 0.05

	ELA		Math	
	(1)	(2)	(3)	(4)
Lowest-Performing Districts				
DPS treatment effect	0.060***	0.033***	0.057***	0.037***
	(0.008)	(0.010)	(0.009)	(0.011)
Pre-reform parallel trajectory	Yes	Yes	Yes	Yes
Aurora Public Schools (APS)				
DPS treatment effect	0.036**	0.021	0.077***	0.039
	(0.007)	(0.144)	(0.003)	(0.188)
Pre-reform parallel trajectory	No	Yes	No	Yes
District & year fixed effects	Yes	Yes	Yes	Yes
District controls	No	Yes	No	Yes

*** p < 0.001; ** p < 0.01; * p < 0.05

Table C3

CITS Estimate of Impact of DPS Reform Efforts on Student Performance by Grade Configuration

	Elementary Schools				Middle Schools			
	ELA		Math		ELA		Math	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Largest Districts								
DPS treatment effect	0.061***	0.069***	0.061***	0.059***		0.088***		0.058***
	(0.005)	(0.013)	(0.004)	(0.011)		(0.014)		(0.012)
Pre-reform parallel trajectory	Yes	Yes	Yes	Yes	No	Yes	No	Yes

*** p < 0.001; ** p < 0.01; * p < 0.05

	Elementary Schools				Middle Schools			
	ELA		Math		ELA		Math	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Lowest-Performing Districts								
DPS treatment effect	0.067***		0.066***	0.051***				
	(0.008)		(0.004)	(0.011)				
Pre-reform parallel	Yes	No	Yes	Yes	No	No	No	No
Aurora Public Schools (APS)								
DPS treatment effect	0.054***	0.098	0.069***	0.049		0.094*		0.127
	(0.003)	(0.018)	(0.003)	(0.038)		(0.007)		(0.061)
Pre-reform parallel	Yes	Yes	Yes	Yes	No	Yes	No	Yes
District & year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District controls	No	Yes	No	Yes	No	Yes	No	Yes

*** p < 0.001; ** p < 0.01; * p < 0.05