DIFFERENTIATED ASSISTANCE

The Effects of Technical Assistance on Mathematics and English Language Arts Achievement: Evidence from California's Statewide System of Support

REPORT 2021

STRATEGIC DATA PROJECT

san diego county office of EDUCATION
FUTURE WITHOUT BOUNDARIES℠

Full report can be accessed at:
www.sdcoe.net/lls/assessment/Pages/EvalServices.aspx

The Effects of Technical Assistance on Mathematics and English Language Arts Achievement: Evidence from California's Statewide System of Support

A SAN DIEGO COUNTY OFFICE OF EDUCATION PUBLICATION
APRIL 2021

san diego county office of
EDUCATION
FUTURE WITHOUT BOUNDARIES™

The Effects of Technical Assistance on Mathematics and English Language Arts Achievement: Evidence from California's Statewide System of Support

STUDY

PRINCIPAL INVESTIGATOR
Shannon E. Coulter, Ph.D.
Director, Research and Evaluation
San Diego County Office of Education

HARVARD STRATEGIC DATA PROJECT

ADVISORS
Beth Schueler, Ed.D.
Assistant Professor of Education and Public Policy
Curry School of Education and Human Development
University of Virginia

Miriam Greenberg
Director, Strategic Data Project
Center for Education Policy Research
Harvard University

Jared Knowles, Ph.D.
President, Civilytics Consulting, LLC
Center for Education Policy Research
Harvard University

REPORT DESIGN

Albert Hernandez
Graphic Arts Technician
San Diego County Office of Education

TABLE OF **CONTENTS**

FIGURES AND **TABLES**

LIST OF FIGURES

LIST OF TABLES

EXECUTIVE SUMMARY

The Differentiated Assistance model (DA) within the California System of Support aims to build district capacity to improve outcomes in low-performing schools and districts and close the achievement gap for vulnerable student groups. State legislation empowered county offices of education (COEs) to recognize school districts as the drivers of change, and provide professional development in analyzing and using data, systems analysis, and equity-driven continuous improvement and implementation. COEs coach district leaders on strategies for building a data culture, developing goals, implementing change, and monitoring the progress of their actions. These reform efforts maximize district choice and local decision-making and provide resources and financial support in excess of $25 million annually.

This report assesses the progress of the DA model to improve student achievement and reduce the achievement gap at the end of the second year of implementation. We used a combination of econometric strategies and panel data from the California Department of Education (CDE) to answer questions about the causal impact of this reform.

MAIN FINDINGS

- **After two years of policy implementation, little evidence exists to suggest DA improves achievement and reduces achievement gaps in a meaningful way.**
    - o Both assistance and non-assistance districts experienced a shock (substantial decline in achievement) in student achievement the year of DA eligibility and neither group has rebounded from this decline to date.
    - o In the two years following identification, both English language arts (ELA) and math scores of assistance districts improved fractionally, but not meaningfully, compared to similar districts.
    - o The achievement gap between low-income districts and other districts declined slightly in the first year following DA eligibility, and the gap returned to pre-intervention levels in year two..

POLICY IMPLICATIONS

Operationalizing Success

- Two years of evidence suggests that school districts receiving assistance are no better or worse off than districts not receiving assistance. DA is more successful relative to other states' accountability reforms, but DA lacks clear, measurable outcomes that guide the efforts of COEs directing the work.

Avoiding Unintended Consequences

- Limited empirical evidence links the effects of DA to achievement; however, the state's focus on using multiple measures of quality on the California Schools Dashboard along with using DA as a carrot (versus a stick) for low performing districts may avoid many of the unintended consequences that plague accountability reforms.

Consistency in Implementation

- A better understanding of how to address implementation challenges may consolidate DA practices across COEs and create important connections between DA practices and outcomes.

SECTION 1

INTRODUCTION

SECTION 1: INTRODUCTION

With nearly 1,400 local educational agencies (LEAs) comprising more than 10,000 schools, it is a challenging problem to respond to the needs of struggling districts in a state as geographically diverse as California. While California has defined its accountability system based on each iteration of the No Child Left Behind Act (NCLB), a substantial policy shift occurred in 2013 when the California Legislature passed the Local Control Funding Formula (LCFF). LCFF altered K-12 funding across the state by directing more monies to districts that serve larger percentages of historically marginalized student groups. Along with this shift in funding, the state profoundly transformed the roles of county offices of education (COEs) by mandating that COEs approve districts' Local Control and Accountability Plans (LCAPs). This shift placed COEs in an oversight position, approving districts' intended activities and resource allocations to meet the eight state priorities. In 2017, county offices' functions shifted again when California began implementing a new statewide system of support — a response to mandates in the Every Student Succeeds Act (ESSA). Modeled after a multi-tiered system of support (MTSS), districts became eligible for multiple levels of assistance based on results from the California Schools Dashboard. The first level, general assistance (GA), comprises resources and assistance available to all districts and schools. The second level of assistance, known as differentiated assistance (DA), targets support to districts with performance gaps between student groups. The final level of support, intensive intervention (II), provides support to districts identified with persistent performance issues and a lack of improvement over consecutive years. In a four-year span, legislation empowered these historically disinterested COEs to approve each districts' three-year funding plan and directed COEs to support districts in addressing performance gaps across the state's eight priorities. This shift in policy created a dramatically different landscape to explore questions about (a) whether the district level is the right unit of support in a state accountability system and (b) the extent to which COE support influences student outcomes.

THE PURPOSE OF THIS REPORT

In 2018-19, about one-third of California school districts failed to make adequate progress in addressing critical equity issues, resulting in their eligibility for assistance. Once identified, legislation mandates that COEs provide support that addresses performance gaps across student groups. Understanding the effects of this assistance on student performance, therefore, is relevant because DA is a substantial state investment designed to have broad effects on district systems serving marginalized students. The primary purpose of this report is to expose the impact of differentiated assistance on student performance in mathematics and English language arts. Findings from this report also build on existing accountability literature by investigating the role of a district as a driver of school change, the extent to which states can build improvement capacity regionally, and finally how a Tier II statewide support compares to more aggressive school turnaround models.

SECTION 2

ACCOUNTABILITY IN CONTEXT

SECTION 2: ACCOUNTABILITY IN CONTEXT

Since the introduction of the NCLB Act of 2001, the role of accountability and its impact on student outcomes has been a central theme. Early reviews of state accountability systems found accountability positively affected student achievement. However, consequential accountability systems were responsible for most of these effects. While some studies linked accountability policies to improvements in student achievement, others found that accountability failed to improve achievement gaps substantially. Furthermore, an increasing number of accountability studies identified unintended consequences, such as narrowing the curriculum, cheating on standardized tests, along with other problems (Crocco & Costigan, 2007; Hanushek & Raymond, 2005; Kane & Staiger, 2002).

One crucial theme that emerged from research on school level accountability suggested accountability creates differential effects on student achievement. For example, Dee and Jacobs (2009) examined the national effects of the NCLB Act on test-score changes across states with school accountability policies similar to NCLB in place prior to the implementation of NCLB. These researchers determined that NCLB-like accountability policies generated large increases in the math achievement of 4th-graders and modest gains for 8th-grade students, including improvements across Latinx students, students who are eligible for subsidized lunch, and among students at all levels of performance. In reading, however, researchers found no impact of the accountability policies on either 4th- or 8th-grade students' test scores.

Research examining school and teacher behaviors resulting from increased accountability painted another ambiguous picture on the effects of accountability on student achievement. For example, Neal and Schanzenbach (2007) examined the impact of accountability policies on "bubble students" because teachers and schools placed greater emphasis on these students who were at the threshold of achieving proficiency. Neal and Schanzenbach's research suggested sizable effects for students in the middle of the distribution or students who had a chance to achieve proficiency. However, the effects of accountability policies showed little evidence of impact for students who were already proficient or students for whom proficiency was out of reach. Other researchers found similar effects of focusing on "bubble students" (Booher-Jennings, 2005; White & Rosenbaum, 2013), suggesting the effects of accountability depend on where teachers and schools focus.

SCHOOL ACCOUNTABILITY AND SCHOOL REFORM MODELS

Consequential accountability — a form of accountability connecting consequences to school performance — increased during the Obama administration under the 2009 American Recovery and Reinvestment Act (ARRA). ARRA provided $3 billion in funding for School Improvement Grants (SIGs), and these grants required chronically low-performing schools to choose one of four options for improvement: transformation, turnaround, restart, or closure. The transformation model emphasized replacing the principal and evaluating teachers based on student performance. It also focused on instructional strategies, an extended school day, and professional development and support from the district or other providers. The turnaround model replaced the principal and required replacing 50 percent of the staff, whereas under the restart model, the school reopened under the management of a charter organization. Dee examined the effects of SIG grants on the lowest performing schools in California and found gains in achievement equivalent to 34 points on the state's Academic Performance Indicator (2012).

The Institute of Education Sciences (IES) investigation of SIGs involved a national sample. IES compared the performance of schools just low enough to qualify for the grant to otherwise similar schools whose performance was not low enough to receive funding. This study revealed no significant effects on student achievement using a similar methodology as Dee (Dragoset, et al., 2017).

SIG reform efforts have a relatively lengthy history in the U.S. Department of Education, and, like other kinds of school-based accountability, show mixed results. Other improvement models initiated over the last 40 years include School-Wide Program (SWP), Comprehensive School Reform Demonstration, and Comprehensive School Reform models. Similar to other SIG programs, most evaluations of these programs were unable to produce clear evidence of their effectiveness to improve student achievement nor were they able to demonstrate unequivocally that reform efforts could scale nationally (Wong & Meyer, 1998; U.S. Department of Education, 2010).

SCHOOL-LEVEL ACCOUNTABILITY AND FOCUS SCHOOLS

With the growing demands of accountability, schools began focusing on the most effective practices reported in school improvement research. Findings emphasized using context specific instructional strategies, embracing different methods of turnaround, incorporating systemic district-level changes, and investing in capacity building (Knudson, Shamburgh, & O'Day, 2011). Los Angeles Unified adopted prior research findings in the development of the Public School Choice Initiative (PSCI) — a response to improving student achievement at its lowest performing schools. An evaluation of PSCI determined that schools adopting transformational and turnaround models had decreases in student achievement in one cohort and no effect in another cohort. However, significant positive impacts in ELA occurred in schools implementing reconstitution and restart models, which included hiring new leadership, staff, and making programmatic changes (Strunk, Marsh, Hashim, Bush, & Weinstein, 2016). Results confirmed Dee's findings that turnaround models are less effective at improving student achievement. Additionally, PSCI is an important study regarding the impact of a district's attempt to scale school improvement across schools.

Similar to the PCSI model, Bonilla and Dee described efforts to improve Focus Schools in Louisiana by examining other effective practices including (a) comprehensive data analysis and needs assessment and (b) coordinated support within a technical assistance network. Essentially, Focus Schools are low-performing schools targeted for intervention. Bonilla and Dee examined the impact of Louisiana's new school reform policy on school performance and academic achievement using a Regression Discontinuity design (RD). The authors concluded that participating in the first cohort of Focus Schools had no significant impact on the performance of the schools. Also, the authors found a significant negative estimate for the 2015 outcomes, which suggested that schools that were "treated" for three years performed significantly worse than other low-performing schools (2020).

In a follow-up study, Dee and Dizon-Ross examined Kentucky's Comprehensive School Improvement Plan that focused on the strategies of rigorous instructional practice and data-informed decision-making. These authors investigated the impact of Kentucky's school reforms on academic achievement using a similar RD approach. The authors concluded that after the first year of implementation, math and reading proficiency of students on the

RD threshold increased meaningfully. Additionally, 47 Focus Schools exited Focus School status. Results suggested that high-quality professional development was partly responsible for the effects on student achievement.

Given years of mixed results linking school-level accountability to student achievement, many policymakers and researchers refocused their efforts on a different unit of change. Many state-level accountability systems redirected policies from providing school-level accountability and support to examining the effects of support provided by a district to a school.

DISTRICT-LEVEL ACCOUNTABILITY

For the past decade, researchers questioned whether school districts can improve school performance (Johnson, Marietta, Higgins, Mapp, & Grossman, 2015; Zavadsky, 2012). Chingos, Wright, and Gallaher determined that districts explain a share of variation in student achievement over and beyond the variability explained by schools and teachers (2013). These researchers established the rationale for examining patterns in district performance linked to student achievement along with identifying the mediators of school performance within a district.

Ouchi (2006) examined the link between decentralized districts and student performance by comparing three decentralized districts to three districts with more centralized control. Ouchi characterized decentralization as individual districts treated as a self-contained unit. For example, decentralized units are free to hire their own personnel; principals are free to alter the school schedule; and teachers have the autonomy to choose their own teaching methods. Ouchi concluded that decentralized school districts in Edmonton, Houston, and Seattle out-performed centralized districts in New York, Los Angeles, and Chicago. However, Ouchi's conclusions are tenuous at best, given that he failed to provide evidence of a causal effect of decentralization on student performance. In addition, Ouchi's conclusion about differences in achievement and reductions in achievement gaps were isolated to two districts: Los Angeles and Houston. Despite methodological limitations, Ouchi's research raised an important question regarding the district's role in mediating effects in school achievement.

In a more methodologically robust study of schools operating as districts, Abdulkadiroglu and colleagues compared lottery-based estimates of the impact of charter attendance on achievement in Boston to estimates on achievement in Boston's Pilot Schools. These pilot schools operated as individual districts within the Boston Public Schools system, with allowances in autonomy for staffing, budgeting, curriculum, governance, and school calendar. Results suggested significant gains in achievement in ELA and math for charter students, but the comparative analysis found insignificant effects for the pilot schools (2011).

Similarly, Wong and Shen examined district takeovers in 11 states finding no clear trends in achievement. Authors discovered examples of impact in certain grade levels, in certain low-performing schools, or with certain students in the achievement distribution, but no consistent trend within or across districts (Wong & Shen, 2002; 2003).
A study examining the effects of district-level reforms on student achievement in a takeover model in Massachusetts found sizable gains in mathematics and modest gains in reading across the first two years of the takeover (Schueler, Goodman, & Deming, 2017). These researchers found performance gaps diminished in the high poverty area of Lawrence, Massachusetts, and the district sustained the reductions across multiple years. Similar to findings from the Institute of Sciences, the effects in Schueler occurred in a turnaround model (where the district replaced only 10% of teachers).

Harris and Larsen's study of New Orleans schools showed promising results from converting all the city's schools into self-managed charter schools. Researchers found reforms post Hurricane Katrina increased student achievement by 0.2 standard deviations at minimum and more likely 0.3-0.4 standard deviations. Harris and Larsen's study provided evidence that intensive system-wide reform can produce large effects on student learning. However, it is unclear whether districts can achieve large gains at scale without incurring a natural disaster like Hurricane Katrina (Harris & Larsen, 2016).

The creation of Tennessee's Achievement School District (ASD) and Innovation Zones (iZone) provided a competing narrative regarding the impact of accountability on student achievement. The ASD included the state's lowest performing 5% of Title I schools reconstituted into a new state-run school district. Results from a difference-in-differences analysis indicated insignificant or in some cases mixed effects depending on the subject, cohort, and academic year. iZone schools however showed robust improvement. One major difference between the two models: iZone schools remained under district management and included greater autonomy in resource decisions (Zimmer, Kho, Henry, & Viano, 2015). While results suggested that students in the ASD were no better off with the reforms, research on the iZone schools raised important questions about the types of reforms required to improve low-performing schools.

ESSA required states to develop policies to identify and turnaround low-performing schools as part of its accountability system. California has a substantial need for preliminary evidence about the state's efforts to comply with ESSA, and whether its accountability policies change outcomes for students in low-performing districts. The research on accountability, as it applies to low-performing schools, is mixed. Some studies found gains in reading and mathematics, while others found no gains or even negative results. While some evidence exists on the importance of districts as drivers of improvement, other evidence indicates otherwise. Changes in achievement appear more likely where substantial improvement efforts occur, when states invest sufficient resources into district capacity, and when districts and schools implement interventions effectively. Furthermore, several studies suggest that less aggressive reform efforts, where states provide districts with greater autonomy and additional resources results in better student-learning outcomes. This report connects to prior research by examining the impact of technical assistance at a regional level to improve student achievement. California's System of Support uses county offices of education (54 in total) to build the capacity of the state's 1,390 school districts to improve outcomes. In states as diverse as California, regional assistance may better support student achievement and scale improvement work. Furthermore, the California System of Support model focuses on a process for determining the needs of districts and carrying out a plan to address the needs versus more traditional school reform practices of replacing structural components in the system such as principals and teachers. California's accountability policies focus on a process to learn and improve versus a corrective process that punishes schools and districts.

SECTION 3

THE CALIFORNIA SYSTEM OF SUPPORT

SECTION 3: THE CALIFORNIA SYSTEM OF SUPPORT

California Education Code Section 52095.5(b) authorized California's System of Support in 1999. The most recent iteration of the support system is a response to changes in ESSA that encouraged states to adopt a tiered intervention system. California designed the system to support districts through geographic lead agencies using a three-tiered model. Lead agencies, made up of multiple county offices of education, support the 54 California COEs in building the local capacity of the state's 1,390 districts. This support focuses on improving outcomes for California's students in three major areas:

1. **Support the continuous improvement of student performance in each of the eight state priorities**
2. **Address the gaps in achievement between student groups**
3. **Improve outreach and collaboration with stakeholders to ensure that goals, actions, and services described in school district and COEs Local Control and Accountability Plans reflect the needs of students and the community, especially for historically underrepresented or low-achieving groups** (California Department of Education, 2020)

California designed a statewide system of support to assist districts and schools to meet the needs of each student. Modeled conceptually after a Multi-Tiered System of Support framework, California's statewide System of Support aligns state and regional resources to support improvement for all schools and districts using three levels of supports: General support for all districts and schools, Differentiated Assistance, and Intensive Intervention. The first level, general assistance (GA), comprises resources and assistance available to all districts and schools. Resources include curriculum frameworks, professional development, coaching aimed at narrowing disparities among student groups. The second level of assistance, known as differentiated assistance (DA), is targeted assistance offered to districts that meet certain eligibility requirements by California's COEs, the CDE, and the California Collaborative for Educational Excellence (CCEE). These organizations are responsible for supporting districts with the underlying causes that led to eligibility for assistance in addition to strengthening the districts capacity to evaluate the effectiveness of its programs. California offers the final level of support, intensive intervention (II), to schools within districts that have persistent performance issues over consecutive years.

California's current system of support differs from prior systems in several important ways. First, the system emphasizes the district as the unit of change versus the school. Second, it stresses continuous improvement over consequential accountability. Next, the system uses a multi-indicator dashboard aligned to all eight state priorities to assess school quality versus a single number. Finally, the system integrates a data and monitoring system (California Schools Dashboard) with the planning and funding system (LCAP and LCFF) in an effort to develop coherence across all accountability elements (Humphrey & O'Day, 2019). The California System of Support aligns the California Schools Dashboard, Local Control and Accountability Plans, and a tiered assistance model to focus districts on comprehensive student success.

CALIFORNIA SCHOOLS DASHBOARD

The California Schools Dashboard provides information on a district's progress on the state's eight priorities. Every district's Dashboard, depending on the grade spans of the students it serves, illustrates the district's and school's current status and change across multiple indicators, including graduation rate, suspension rate, college/career readiness, and mathematics and ELA achievement, chronic absenteeism, and English learner progress (California Department of Education, 2020). The Dashboard illustrates performance levels using gauges across five levels of performance from red (the lowest performance level) to blue (the highest). In addition to visualizing achievement for all students, the Dashboard provides a comprehensive analysis by each student group in the district or school, including students who are identified as homeless, English learners, foster youth, students with disabilities, and socioeconomically disadvantaged, along with breakdowns of every race and ethnicity category. Combinations of red and orange gauges for multiple student groups determine whether districts are eligible for differentiated assistance.

LOCAL CONTROL AND ACCOUNTABILITY PLAN

California's Local Control and Accountability Plan (LCAP) is a three-year plan (updated annually) that describes the district's or LEA's goals, actions, services, and expenditures to support student outcomes. Essentially, the LCAP is an opportunity for districts to share the how, what, and why programs and services they selected meet local needs. Each plan consists of a process for engaging stakeholders in addition to sections on resource inequities, analysis, identified need, goals, actions, strategies, and measurable outcomes.

DIFFERENTIATED OR TECHNICAL ASSISTANCE

DA is state assistance provided to districts that fail to meet the performance criteria established by the State Board of Education (SBE). When the state operationalized DA in the 2016-17 school year, approximately 218 districts were eligible for assistance based on results from the California Schools Dashboard. FIGURE 1 illustrates each school district in California participating in Differentiated Assistance in 2017.

DA is a multi-stage process that utilizes a district-based team of 4-12 members including superintendents, assistant superintendents, directors, principals, and frontline staff including teachers, counselors, and support personnel. Teams receive individual support from COEs, the CDE, and the CCEE in the form of improvement coaching throughout the DA process, which involves a needs assessment, root cause analysis, and continuous improvement action planning. Figure 2 outlines DA's theory of action.

FIGURE 1. Map of California and School Districts Receiving Assistance.

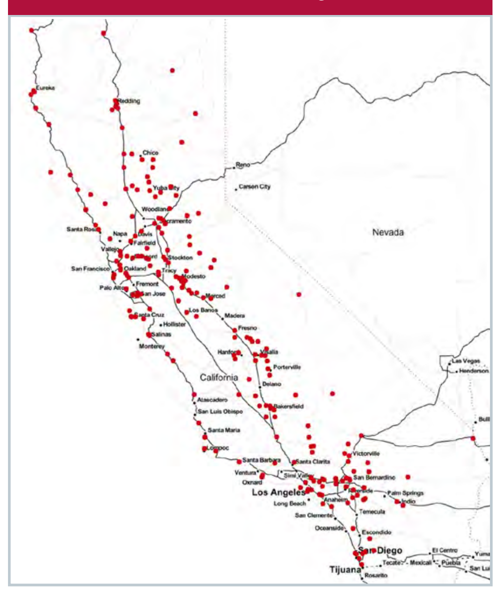
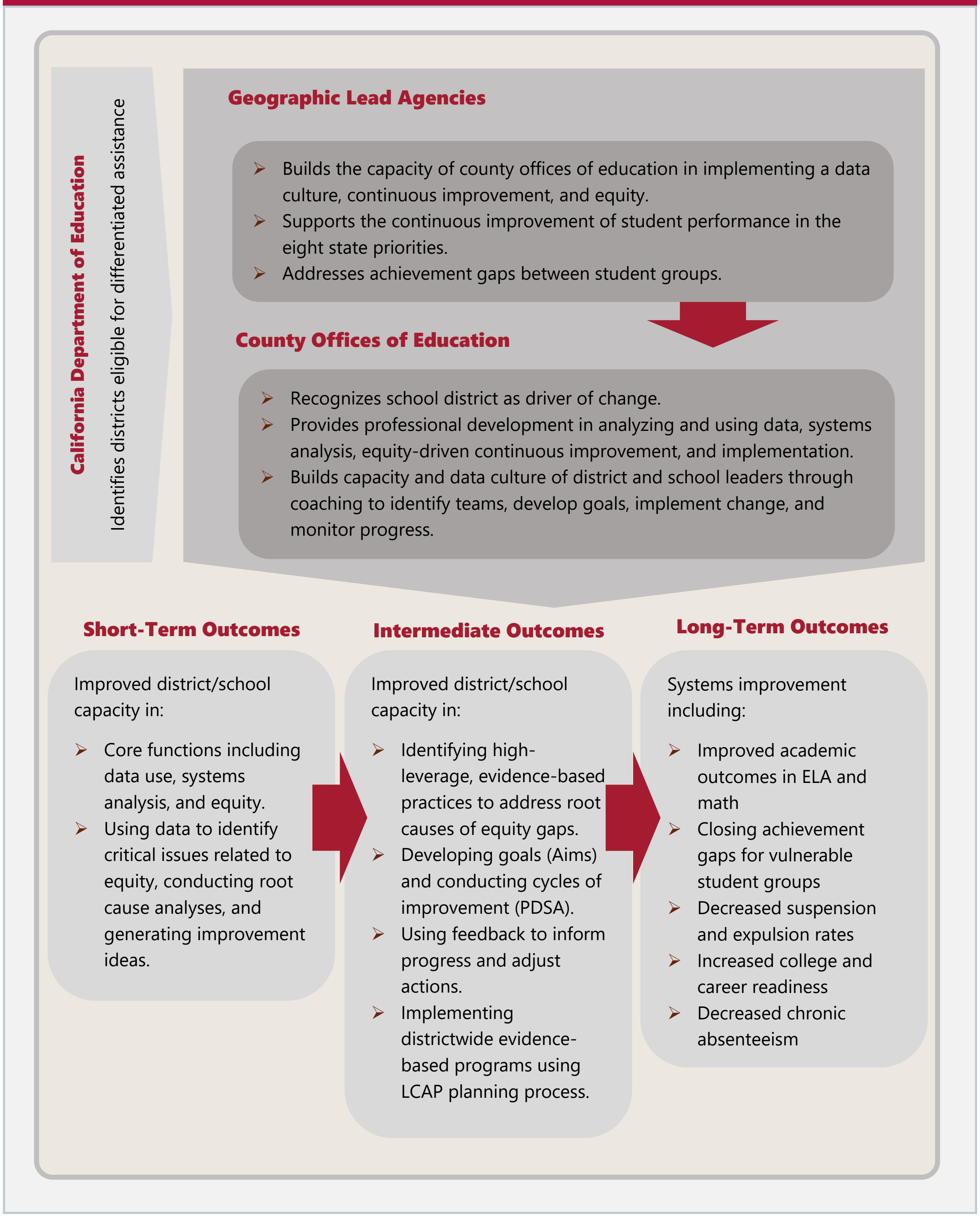


FIGURE 2. Differentiated Assistance Model.



NEEDS ASSESSMENT/ROOT CAUSE ANALYSIS

The first phase of the process consists of a California Schools Dashboard orientation, a systems exploration, a root cause analysis, and a synthesis of findings. District-based leadership teams orient themselves to the California Schools Dashboard and the reasons why the district is eligible for DA by reviewing its performance on the statewide indicators. Districts follow a guided protocol that promotes exploration of the data and supports the team to identify information for continuous improvement efforts. The district team members, armed with data, return to the district and begin examining their system's performance. Many districts engage in continuous improvement processes, such as creating a systems map, conducting empathy interviews, and outlining processes. The systems analysis compliments the Dashboard review and focuses the district team on root causes. The root cause analysis uses an inquiry protocol where district-based teams scrutinize problems by identifying contributing factors to performance gaps and examining the differences between current and desired conditions in student achievement. The team's insights lead to an improvement planning process where teams consider (a) change ideas to improve current processes, and (b) implementing strategies and interventions that have a demonstrated impact on performance gaps. District-based teams summarize and consolidate findings from the root cause analysis and begin the planning for addressing system challenges. Finally, each team integrates the findings into the continuous improvement process.

CONTINUOUS IMPROVEMENT

The next stage of the DA process involves the Charter Institute, which originated from a collaboration with the Carnegie Foundation and their six core principles of improvement (Bryk, Gomez, Grunow, & LeMahieu, 2015). Essentially, district-based teams use the root cause analysis to define the problem and examine the variability in the system that produces the problem. Teams continue to examine the system using mapping tools and interviewing strategies to understand the work people carry out in the system. Teams develop a set of measures that help understand progress made in addressing the problem. Teams use measures as part of the small-scale experiments to test change ideas. These Plan, Do, Study, and Act (PDSA) cycles are critical milestones to refining and scaling ideas successfully. Finally, teams conduct the work embedded in a networked community of other districts struggling with similar problems. The collaborative approach within and across teams unites the groups around a shared purpose and diffuses solutions across a wide community.

Eligibility for DA is complex and requires evidence of a student group failing to meet the criteria for two or more state priorities. Four main priorities contribute to DA eligibility including achievement in English language arts (ELA) and math (priority 4), graduation and chronic absenteeism rates (priority 5), suspension rates (priority 6), and college readiness (priority 8). While the system for determining eligibility involves too many scenarios to describe, one possible scenario might involve a student group, for example, Foster Youth, with red gauges on the California Schools Dashboard in both graduation rate and suspension rate. The gauges suggest that the student group has a low percentage of students graduating on time and a higher percentage of students suspended in the current year.

FIGURE 3. Criteria for DA Eligibility

Red is the lowest of the five performance categories for state indicators, and Not Met for Two or More Years is the lowest rating for local performance indicators.

- Basics (Priority 1)
  - ***Not Met for Two or More Years*** on Local Performance Indicator
- Implementation of State Academic Standards (Priority 2)
  - ***Not Met for Two or More Years*** on Local Performance Indicator
- Parent Engagement (Priority 3)
  - ***Not Met for Two or More Years*** on Local Performance Indicator
- Pupil Achievement (Priority 4)
  - ***Red*** on both English Language Arts (ELA) and Math tests OR
  - ***Red*** on ELA or Math test AND ***Orange*** on the other test OR
  - ***Red*** on the English Learner Indicator (EL student group only)
- Pupil Engagement (Priority 5)
  - ***Red*** on Graduation Rate Indicator OR
  - ***Red*** on Chronic Absence Indicator (when available)
- School Climate (Priority 6)
  - ***Red*** on Suspension Rate Indicator OR
  - ***Not Met for Two or More Years*** on Local Performance Indicator
- Access to & Outcomes in a Broad Course of Study (Priority 7 & 8)
  - ***Red*** on College/Career Indicator
- Coordination of Services for Expelled Pupils – COEs Only (Priority 9)
  - ***Not Met for Two or More Years*** on Local Performance Indicator
- Coordination of Services for Foster Youth – COEs Only (Priority 10)
  - ***Not Met for Two or More Years*** on Local Performance Indicator

DA eligibility indicates prolonged inequities between student groups in local school districts. Each of the inequities present real student-level costs, including limited college and career options, less career earnings power, and a lower quality of life in general. DA eligibility triggers a substantial investment by COEs to support districts to reduce persistent gaps across the eight state priorities. Such an investment begs the question of whether DA is an effective accountability intervention for reducing inequities in school districts.

SECTION 4

DATA AND METHODS

SECTION 4: DATA AND METHODS

DATA

We used administrative datasets from the CDE. The district-level panel data included results from 2014-15 to the 2018-19 school year for district, grade, standardized test scores, district-level student characteristics, and other district characteristics. Our full sample of data included over 49,000 unique records by district, year, grade, and test. We standardized students' scores on ELA and math exams by year, grade, and subject. We also captured data on district enrollment, proportion of students receiving free and reduced-price lunch, proportion of English learners, a diversity index of student race and ethnicity demographics, proportion of staff with master's degrees, average teaching experience, average daily attendance, and district revenue. We transformed enrollment and revenue by taking log counts and dollars of the variables. We constructed a diversity index with a scale of 0-100 to measure district demographic composition. For the diversity index, we calculated the proportions of each of the eight major subgroups. A score near zero represents a homogenous district composition, and a score of 100 reflects equal shares of all major subgroups. Scores on the index ranged from a low of zero to a high of 74. Table 1 illustrates descriptive characteristics for general assistance districts, differentiated assistance districts, and a subset of general assistance districts that experienced both low-status and growth in ELA and math during the year of DA eligibility (2016-2017). We arrayed descriptive statistics for each group both before and after the policy change.

TABLE 1. Descriptive Characteristics of General and Differentiated Assistance Districts, and a Smaller Set of Low-Achievement & Growth General Assistance Districts.

	General Assistance Districts		Differentiated Assistance Districts		Low-Status/ Growth Districts	
Characteristics	Before 2018	After 2018	Before 2018	After 2018	Before 2018	After 2018
Free-and-Reduced-Price Lunch	49%	50%	63%	63%	61%	61%
English Learners	20%	19%	22%	22%	24%	22%
Diversity Index	35	36	38	38	32	32
Master's Degree	9.4%	9.5%	9.4%	9.3%	8.7%	8.9%
Teaching Experience (years)	13	13	13	13	13	13
Enrollment	5,189	5,216	13,368	13,352	4,924	4,883
Average Daily Attendance	88%	88%	87%	87%	88%	87%
Revenue (in millions)	$15,031	$16,296	$17,055	$17,783	$14,622	$15,359
CAASPP 4th grade ELA Score	2458	2468	2431	2442	2437	2447
CAASPP 4th grade Math Score	2465	2475	2439	2449	2447	2455
Districts (count)	664	651	218	217	270	270

MODEL

We estimated the impact of technical assistance by comparing changes over time in districts receiving assistance with changes in districts that never received technical support using an econometric approach called a difference-in-differences (DD) analysis. A DD approach provides robust estimates of the causal effect of a policy change in situations where random assignment is either impractical, impossible, or unethical. To use DD, we observe outcomes in districts who received assistance (treated) and districts not eligible for assistance (control), both before and after the policy change. Armed with our observations, we compute a simple difference between the post policy performance and pre-policy performance across both groups, and then we simply subtract this first difference in the control from the treatment group. In this way, the DD design removes the influence of the pre-policy performance in the first difference and the second difference removes the change that occurred in the control group over the analytic window, which leaves an estimate of the effect of the intervention. To ensure we have unbiased estimates, both groups must have a similar likelihood of receiving the assistance; therefore, the trends in their pre-policy performance must be similar.

Districts received assistance for gaps in different types of outcomes including gaps in academic achievement, discipline, chronic absenteeism, and college readiness rates. In this model, our treatment group includes only the districts receiving assistance for academic gaps and excludes the districts receiving assistance for other reasons. We considered the 2015-17 school years as the pre-assistance control period and used both the 2018 and 2019 school years as the post treatment periods. When the accountability policy change went into effect at the end of the 2017 (eligibility year), the state categorized districts into two groups: Differentiated assistance (districts receiving services from COEs for performance gaps in ELA and mathematics) and general assistance for any district not receiving support for gaps in student outcomes.

We estimated the variation for our DD analysis using equation (1):

$$\Upsilon_{d[g]t} = \alpha + \beta_1\lambda_d + \beta_2\gamma_t + \beta_3\chi_{dt} + \beta_4 Assist_{dt} * \gamma_t + \varepsilon_{d[g]t}$$

where $\Upsilon_{d[g]t}$ represents the dependent variable, standardized test scores, where $d[g]$ indexes grade level by district (1 if eligible for assistance, 0 otherwise) and t indexes years (1 is after the policy, 0 before). λ_d are the district fixed effects, γ_t are year fixed effects, $Assist_{dt} * \gamma_t$ is an interaction for assistance and time (and our estimate of impact), χ_{dt} represents a matrix of covariates historically linked to achievement outcomes, including percent eligible for FRPM, percent of district demographic composition, and enrollment, ε_{dt} is an error term clustered at the district level (Bertrand, Duflo, & Mullainathan, 2002; Angrist & Pischke, 2009) since assistance is a district-level intervention.

We compared districts receiving assistance to general assistance districts who experienced low status and growth on the California Schools Dashboard. Our first table compares differentiated assistance districts to all general assistance districts and our more restrictive sample. We recognize the bias in comparing DA districts to all general assistance districts given the fact that some of the general assistance districts are both high achievement and growth districts. To address the bias, we also compared differentiated assistance districts to a subset of 270 general assistance districts with both low status and growth in 2017, the same year our assistance policy turned on for some districts. We constructed this restricted sample using the California Schools Dashboard, which assigns every district a color based on its performance. The California Department of Education calculates each district's performance based on data from current and prior years, resulting in five color-coded performance levels for each indicator. Red areas on the table represent very low status and change, whereas blue reflects high status and change. Our more restrictive comparison group comprised districts whose performance increased by less than 15 points, had low or very low status, and were not eligible for assistance. Most of these districts were orange on the Dashboard's five-by-five colored table compared to DA districts, which were mostly red.

In addition to the previous analysis examining the impact of assistance on academic achievement, we modeled a separate analysis to examine the effect of assistance on the achievement gap. We estimated the variation in the Difference-in-Difference-in-Differences model using the following equation (2):

$$\begin{aligned}\Upsilon_{d[g]t} = \alpha + \beta_1\lambda_d + \beta_2\gamma_t + \beta_3\chi_{dt} \\ + \beta_4 Assist_d + \beta_5 Assist_d * \gamma_t \\ + \beta_6 LowIncome_d * \gamma_t + \beta_7 LowIncome_d \\ * Assist_d + \beta_8 LowIncome_d * \gamma_t * Assist_d \\ + \varepsilon_{d[g]t}\end{aligned}$$

where $\Upsilon_{d[g]t}$ represents the dependent variable, standardized test scores, where g indexes district by grade level (1 if eligible for assistance, 0 otherwise) and t indexes years (1 is after the law, 0 before). λ_d are the district fixed effects, γ_t are year fixed effects, $Assist_d * \gamma_t$ is an interaction between assistance and time, $LowIncome_d * \gamma_t$ is an interaction between low-income districts and time, $LowIncome_d * Assist_d$ is an interaction between low-income districts and assistance, $LowIncome_d * \gamma_t * \lambda_d$ is an interaction among low-income, assistance, and time, and χ_{dt} represents a matrix of covariates historically linked to achievement outcomes, including percent eligible for FRPM, percent of district demographic composition, and enrollment, $\varepsilon_{d[g]t}$ is an error term clustered at the district level.

A DDD analysis works similarly to our DD analysis with the exception of one more subtraction. In a DDD, we difference the pre-policy performance from the post policy performance (first difference) across all our groups. In this case, we have a low-income and higher income category in both the assistance and non-assistance groups. Next, we difference the change (second difference) in the higher income group receiving assistance from the low-income group receiving assistance. Finally, we difference (third difference) the change from the control group. This final difference provides an estimate of the effect of the intervention on low-income districts relative to other districts and represents the causal impact of assistance on the achievement gap.

ROBUSTNESS CHECKS

To examine potential bias in our models, we conducted several robustness checks. Robustness checks are tests on the assumptions of the models. One assumption is that trends in achievement between treatment and control groups were not changing differentially prior to the policy change. Such an assumption can be particularly challenging for accountability research since assignment to assistance occurs for districts with declines in achievement.

Our first check, therefore, examined the validity of our estimate by examining the effect of assistance on the outcome variable both before and after the change for our assistance and non-assistance districts (Angrist & Pischke, 2009). Researchers refer to this check as the parallel trends test since the effect of assistance on treated districts compared to control districts should be constant prior to the policy change. That is to say, trends in achievement in districts receiving assistance should look similar to trends in districts not eligible for assistance. Another assumption particularly important in this context is whether compositional changes occurred in districts across the analytic window. For example, if certain types of students were leaving districts receiving assistance, then these changes in composition could bias our estimates. To test this assumption, we conducted event studies for each of the demographic variables in the report to determine if compositional changes in districts occurred prior to the policy period, which would represent a source of selection bias. A final assumption examined whether an outcome changed as a result of assistance that should not have changed. This kind of falsification check occurs by replacing the test score outcome with another outcome, which we believed the policy change would not affect. Any effect of assistance on these other outcomes (in addition to our other robustness checks) provides evidence on the extent to which the models offer biased estimates on the causal impacts of assistance on achievement and achievement gaps.

SECTION 5

HOW HAS ASSISTANCE CHANGED STUDENT OUTCOMES?

SECTION 5: HOW HAS ASSISTANCE CHANGED STUDENT OUTCOMES?

DIFFERENCE-IN-DIFFERENCES

We examined trends in achievement for assistance and non-assistance districts with particular attention given to years 2017-18 and 2018-19, the year the state initiated the differentiated assistance policy for some districts. Figure 4 and Table 2 illustrate the effects of assistance on ELA and mathematics achievement using the low status and growth districts as our control group.

Figure 4 depicts the event study analyses using the regression formulas from Section 4. Table 2 contains estimates generated by our regression models. The event study allows us to compare two groups over time, one group under assistance compared to another group that shares many of the same characteristics. Carefully selecting a comparison group whose achievement trends match our assistance group prior to the policy change allows us to attribute any differences in achievement to assistance post policy.

To design the event study, we first developed a set of leading and lagging indicators for each year. For example, Lead-2 represents 2014-15 since 2014-15 reflects two years prior to eligibility. We considered the 2016-17 school year the identification year, 2017-18 is the first year of the Differentiated Assistance model implementation, 2018-19 is the second full year of implementation.

For these event studies, we estimated outcomes using aggregate, district-level data. We assigned treatment status based on the district eligibility for DA in the 2016-17 school year. Our results therefore represent the impact of assistance on districts that experienced the Differentiated Assistance model. Furthermore, event studies use a reference year to compare outcomes between two groups relative to another point in time. In this case, we selected the 2016-17 school year as the reference point since 2016-17 is the year the California Department of Education identified districts as eligible for assistance. As noted in Section 3, the model included a combination of year- and district-fixed effects to control for any shocks within districts and years due to testing along with covariates for demographic and district-level factors to account for any compositional changes over time.

FIGURE 4. Effects of Differentiated Assistance on English Language Arts and Mathematics Achievement.

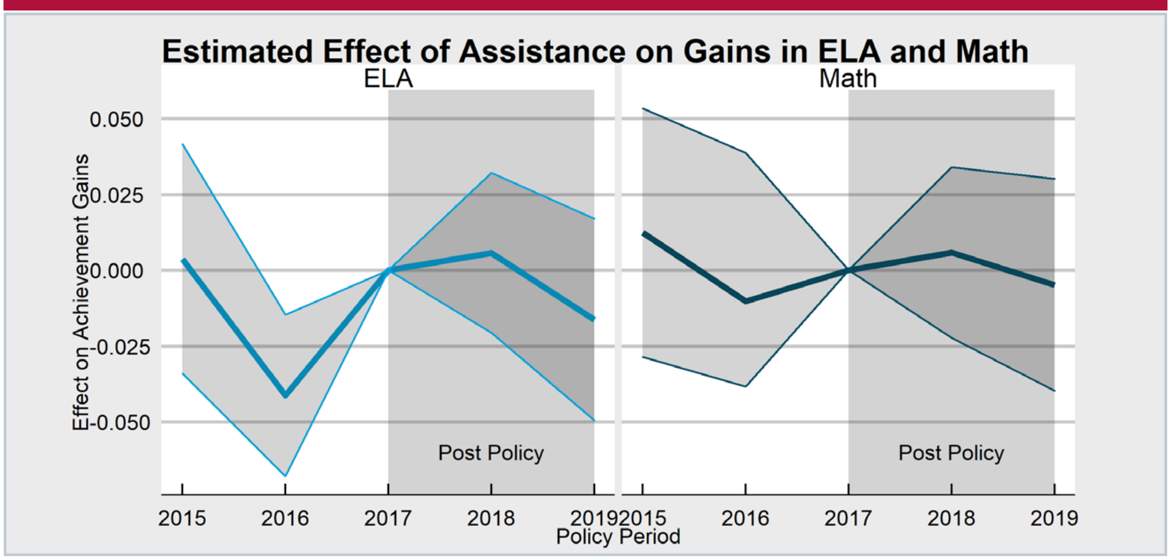


Figure 4 plots the coefficients over time. A coefficient in this case represents the average difference between the performance of the assistance districts to the non-assistance districts relative to the year of eligibility, which is the reason why both the ELA and math graphs zero out in 2017. In the post-policy period of 2018-19, DA districts experienced increases in math and ELA during the first year of implementation and declines during the second. The increase in achievement for assistance districts in math and ELA, however, are not significantly different from the eligibility year, otherwise the shaded ribbon on the graph would not cover zero. Additionally, there appears to be a pre-treatment effect in 2016 in English language arts. Because this effect would call into question our ability to make inferences about the effect of DA from this comparison, we conducted a separate analysis where we matched differentiated assistance districts with a subset of low-growth and status districts to remove this pre-treatment difference. We included the results of the matched comparison group in Table 2.

TABLE 2. Effects of Differentiated Assistance on English Language Arts and Mathematics Achievement.

	Dependent Variables					
	All General Assistance		Low Status & Growth Sample		Matched Comparison Group	
	ELA (1)	Math (2)	ELA (3)	Math (4)	ELA (5)	Math (6)
Year 1 Implementation	-0.049** [0.022]	-0.059*** [0.020]	-0.002 [0.016]	-0.006 [0.014]	0.000 [0.027]	-0.007 [0.025]
Year 2 Implementation	-0.063*** [0.024]	-0.056** [0.023]	-0.029 [0.020]	-0.005 [0.018]	-0.026 [0.090]	-0.017 [0.028]
Observations	24795	24784	14941	14934	14911	14904
R^2	.829	.837	.690	.702	.700	.702

Notes: Standard errors, clustered at the district level, are in brackets. Both models include district and year fixed effects, assistance by year interactions, and covariates: percent of students eligible to receive free or reduced-price lunch, the percent of English learners, the proportion of diversity, and district enrollment. The panel data in this table include 664 districts ineligible for the assistance, 218 districts that were eligible for DA in 2017, and a subset of 270 low status and growth districts. Matched comparison group consisted of a subset of 270 low status and growth districts. Districts observed annually from 2015-2019.

*p<.1, **p<.05, ***p<.01

Table 2 provides point estimates of the effects of assistance on achievement in districts experiencing the DA model compared to all general assistance districts and to a subset of general assistance districts who experienced low status and growth during the year of DA eligibility. We estimated the impact of assistance for DA districts compared to all general assistance districts purely to illustrate the bias that occurs when making an erroneous comparison. Our preferred comparison group is the subset of general assistance districts that experienced low status and growth during the eligibility period, but not to the extent that their performance triggered assistance.

Estimates from our preferred model suggests that ELA scores on average declined .002 standard deviations in year one of assistance and by .029 standard deviations in year two, and suggests no detectable effect of assistance on ELA achievement in both years following the policy change. Estimating changes in standard deviation units is the best way to calculate changes in test scores relative to their average (where the average represents 0). Estimates in mathematics were similar. Mathematics test scores on average declined by .006 standard deviations in year one and by .005 standard deviations in year two relative to the control group. The effects were not significantly different from zero.

Finally, we provided point estimates for the analysis we conducted using the matched comparison group due to the pre-treatment effect in ELA. We used full matching on a propensity score to achieve adequate balance and to account for this pre-treatment difference. To estimate the treatment effect we fit a linear regression using the same outcome and predictors as prior regressions. Our results suggested similar effects as the analysis conducted using the Low Status and Growth districts as a comparison group. The effects were not significantly different from zero for the matched comparison group.

SECTION 6

HOW HAS ASSISTANCE CHANGED THE ACHIEVEMENT GAP?

SECTION 6: HOW HAS ASSISTANCE CHANGED THE ACHIEVEMENT GAP?

DIFFERENCE-IN-DIFFERENCE-IN-DIFFERENCES

The DD analysis examined the effect of assistance on achievement. We were also curious about whether assistance differentially benefited low-income districts receiving assistance more than other districts. We categorized low-income districts as the districts with more than 55% (median) of their students receiving free and reduced-price meals.

FIGURE 5. Average Standardized ELA Test Score in Low Income and Other Districts.

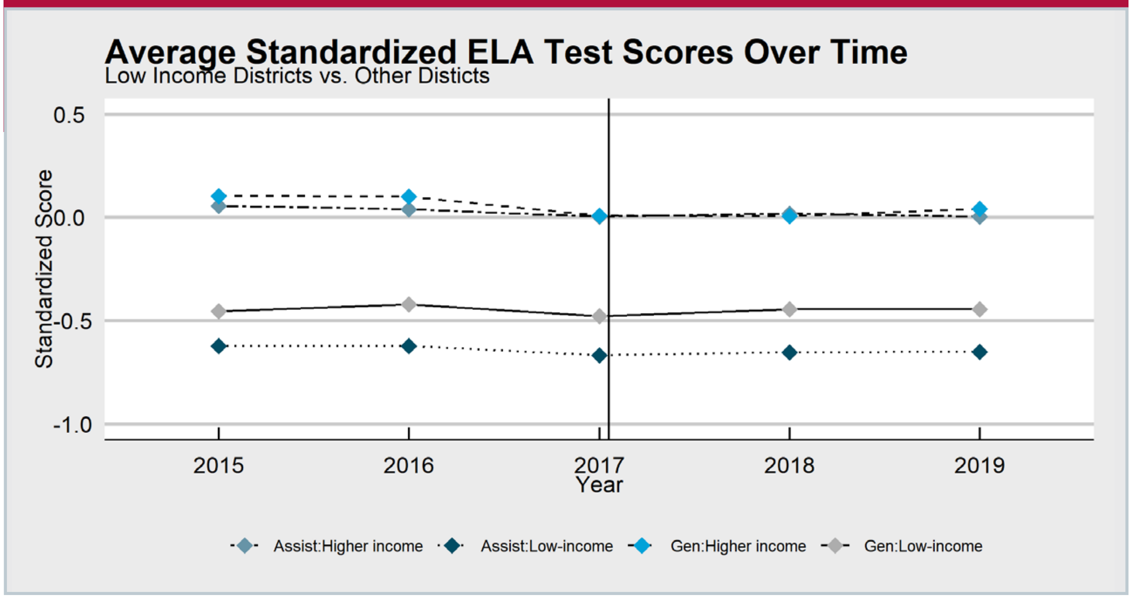


Figure 5 depicts the performance of all four groups. The top dashed line represents higher income districts part of our subset of 270 low status and growth districts not receiving assistance. Their performance is similar in both trend and status to the higher income districts in differentiated assistance. The bottom line represents low-income districts receiving assistance and these districts' trends in status and growth are comparable to the subset of low-income districts not receiving assistance. The solid vertical bar shows the point of eligibility for DA. Patterns in achievement across all four groups appear unchanged in the post policy period.

We first manually calculated the effect of the assistance on standardized test scores in ELA for low-income districts receiving assistance compared to the other groups.

TABLE 3. Effects of Differentiated Assistance on English Language Arts Achievement in Low Income Districts.

Groups	Average Pre Policy	Average Post Policy	Difference
Treatment: Differentiated Assistance			
Low-income	-0.637	-0.652	-.015
	[.634]	[.633]	
Higher income	0.032	0.009	-.023
	[.670]	[.80]	
Difference-in-difference	.008		
Control: General Assistance			
Low-income	-0.450	-0.444	.005
	[.717]	[.683]	
Higher income	0.074	0.022	-.047
	[.896]	[.925]	
Difference-in-difference	.053		
Difference-in-difference-in-difference	-.045		

Notes: Standard errors, clustered at the district level, are in brackets. Each cell contains mean standardized achievement scores for each group. See text for explanation of Pre/Post Policy and Groups. The DDD estimate is the difference-in-differences of the upper panel from the lower panel. *p<.1, **p<.05, ***p<.01

The top panel compares the change in scores for low and high-income districts receiving assistance to the change for other districts not receiving assistance. Each cell contains the average standardized score for the group labeled on the axes along with the standard errors. ELA test score decreased by 0.015 standard deviations in low-income districts receiving assistance in this period compared to a 0.023 decrease in the standardized test scores of higher income districts receiving assistance. Thus, there was a .008 relative increase in standardized test scores in low-income districts receiving assistance. The achievement gap between low-income districts compared to higher income districts receiving assistance decreased slightly during the analytic window.

The bottom panel provides the same comparison for a subset of low status and growth districts not receiving assistance. Low-income districts not receiving assistance experienced a 0.005 increase in standardized test scores in this period compared to a 0.47 decrease in standardized test scores in higher income districts. Essentially, there was a 0.53 relative improvement in standardized test scores in low-income districts not receiving assistance.

By subtracting the assistance effects in higher income districts and the difference-in-differences in non-assistance districts, we recovered the impact of assistance on achievement in low-income districts receiving assistance. There is a 0.045 decrease in relative achievement in low-income districts receiving assistance compared to the change in relative achievement in the other groups. While lower income districts receiving assistance performed marginally better than higher income districts in the post period, after subtracting the difference from the control group, the relative performance of the non-assistance groups was slightly better than our low-income assistance group, suggesting the achievement gap widened during the analytic window. Of course, this analysis fails to consider important variation that exists within low- and higher-income districts, and to address this issue, we expressed the analysis above in a regression model.

DDD REGRESSION

Moving the data into a regression model allows us to consider the influence of fixed effects and covariates on the test scores.

TABLE 4. Effects of Differentiated Assistance on English Language Arts and Mathematics Achievement across Demographic Groups

	Dependent variables			
	ELA Post Policy Period		Math Post Policy Period	
Group	(1)	(2)	(3)	(4)
Low-income districts (with assistance)	.042	-.004	.028	.016
	[.030]	[.040]	[.030]	[.037]
Observations	14923	14923	14916	14916
R^2	.690	.691	.703	.703

Notes: Standard errors, clustered at the district level, are in brackets. Each cell contains mean standardized achievement scores for each group. See text for explanation of Pre/Post Policy and Groups. The DDD estimate is the difference-in-differences of the upper panel from the lower panel.

*p<.1, **p<.05, ***p<.01

The naïve model (no fixed effects or covariates) considered the assistance effects on ELA test scores for low-income districts without accounting for the variation that exists within districts, within years, and due to other district characteristics. Nonetheless, our naïve model suggested a negative effect of assistance on the achievement gap between low-income districts receiving assistance relative to other groups. The first row of Table 4 presents the estimates of the interaction between low-income districts receiving assistance over time relative to the three other groups. The coefficient indicates the achievement gap in ELA decreased by .042 standard deviations in the first year after the policy change and increased by .004 standard deviations after the second year of the policy change. In mathematics, the achievement gap decreased by .028 standard deviations in the first year and by .016 standard deviations in year two. None of the results is significantly different from zero. Furthermore, the coefficients generated from our regression model in ELA were smaller than the coefficients from Table 3, suggesting the importance of fixed effects and covariates.

FIGURE 6. Effects of Differentiated Assistance on English Language Arts and Mathematics Achievement Across Demographic Groups.

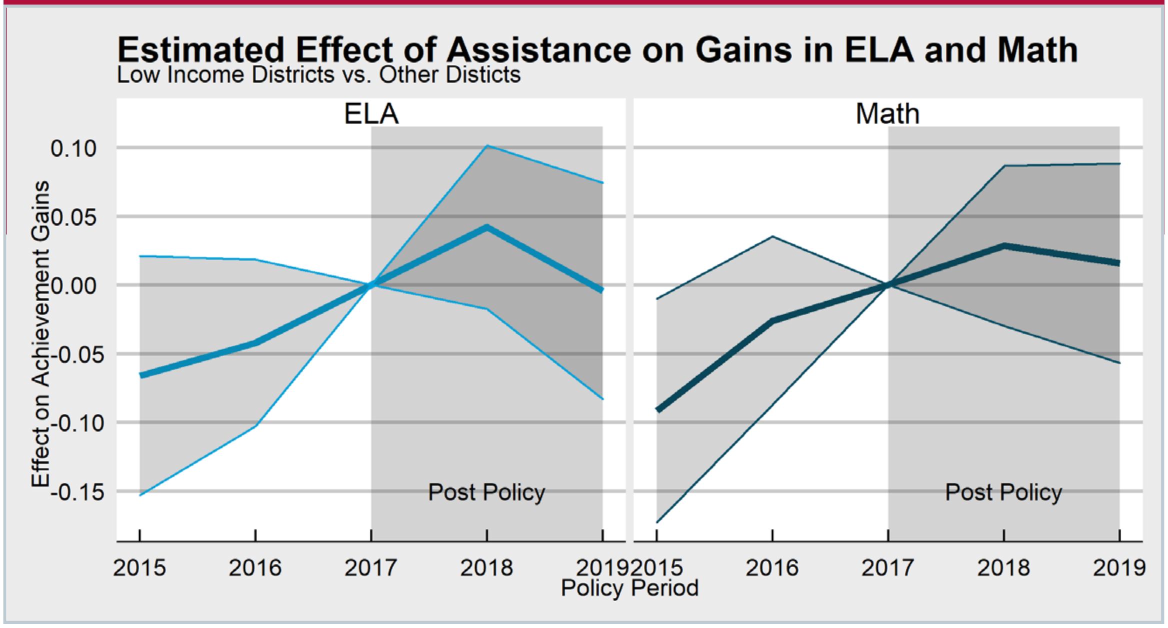


Figure 6 illustrates the difference visually between low-income districts and other districts in both post-policy years of 2017-18 and 2018-19 in both ELA and mathematics. Both panels highlight the pre-policy period and display the average difference between low-income districts and others over time. Whereas the differences started below zero in the pre-policy period, by the first post-policy year point estimates for both ELA and math increase. By the second year, the improvements wane considerably, especially in ELA. Confidence ribbons for both ELA and math encompass zero, suggesting effects both before and after the policy change were no different from zero and the achievement gap was unchanged by the addition of assistance.

SECTION 7

HOW COMPARABLE ARE ASSISTANCE AND NON-ASSISTANCE DISTRICTS?

SECTION 7: HOW COMPARABLE ARE ASSISTANCE AND NON-ASSISTANCE DISTRICTS?

ROBUSTNESS CHECKS

Parallel Trends

Essential to a difference-in-differences analysis is the robustness of the parallel trends assumption. To examine whether trends were parallel, we plotted the means and confidence bands for both ELA and math in Figures 7 and 8.

FIGURE 7. Average Standardized ELA Scores Over Time.

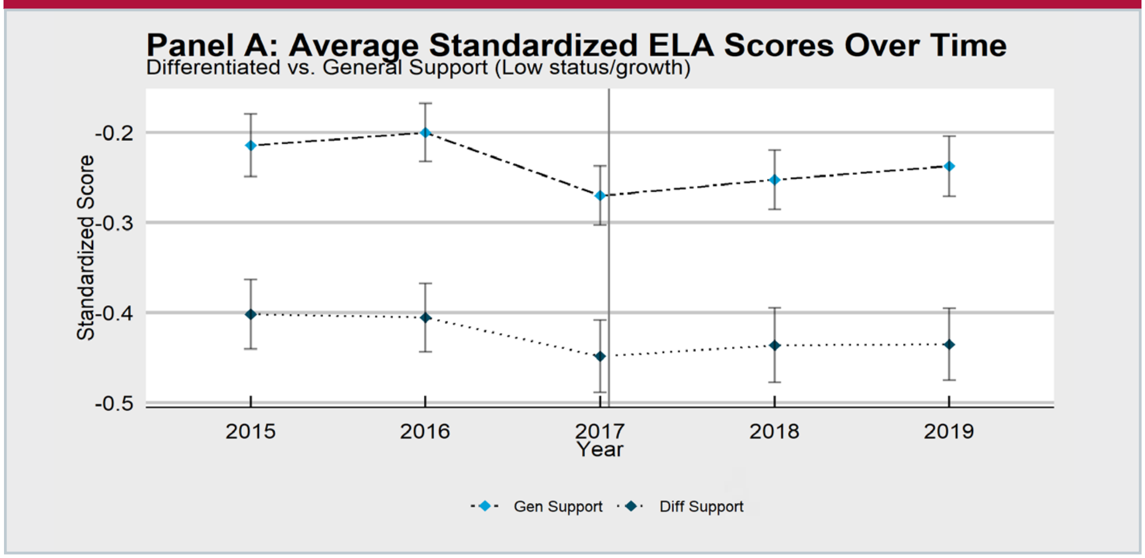
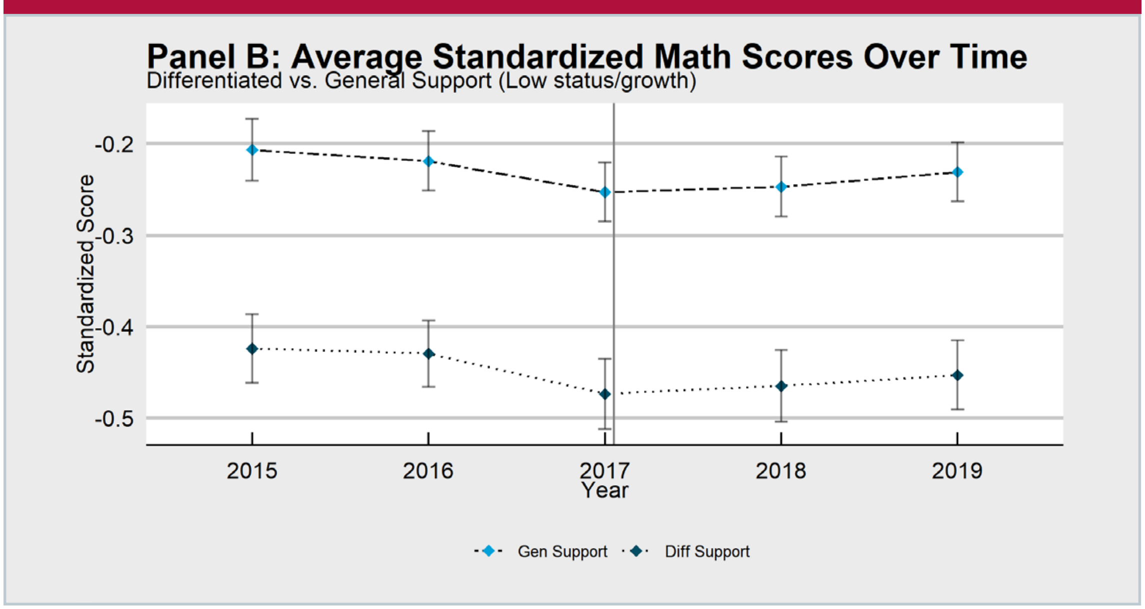


FIGURE 8. Average Standardized Math Scores Over Time.



Panel A depicts the differences in trends in standardized ELA achievement between the assistance and our subset of non-assistance districts; Panel B illustrates the difference for mathematics. The overall trends across all years prior to the policy change appear parallel. Because selection into treatment is a factor of declining achievement, recovering a causal effect of assistance is possible by comparing assistance districts to districts experiencing a similar decline. Both groups appear as likely to receive assistance prior to the policy activating in 2016-17.

The dip in achievement prior to eligibility determination is an interesting phenomenon first identified by Ashenfelter (1978). The 'Ashenfelter dip' occurs when outcomes of potential program participants decline prior to program entry. This shock or decline in achievement actually drives eligibility for DA. A temporary dip in performance would bias our causal estimates if we did not observe a similar shock to achievement in the control group, which is the case according to Figure 9.

FIGURE 9. Average Standardized ELA Achievement of Low Status and Growth General Assistance, Differentiated Assistance, and High Status General Assistance.

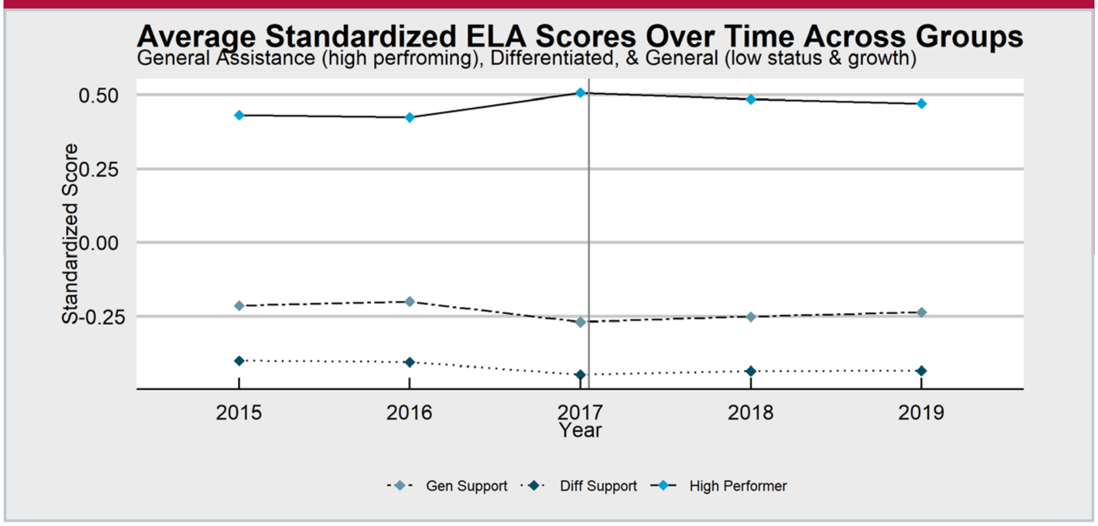
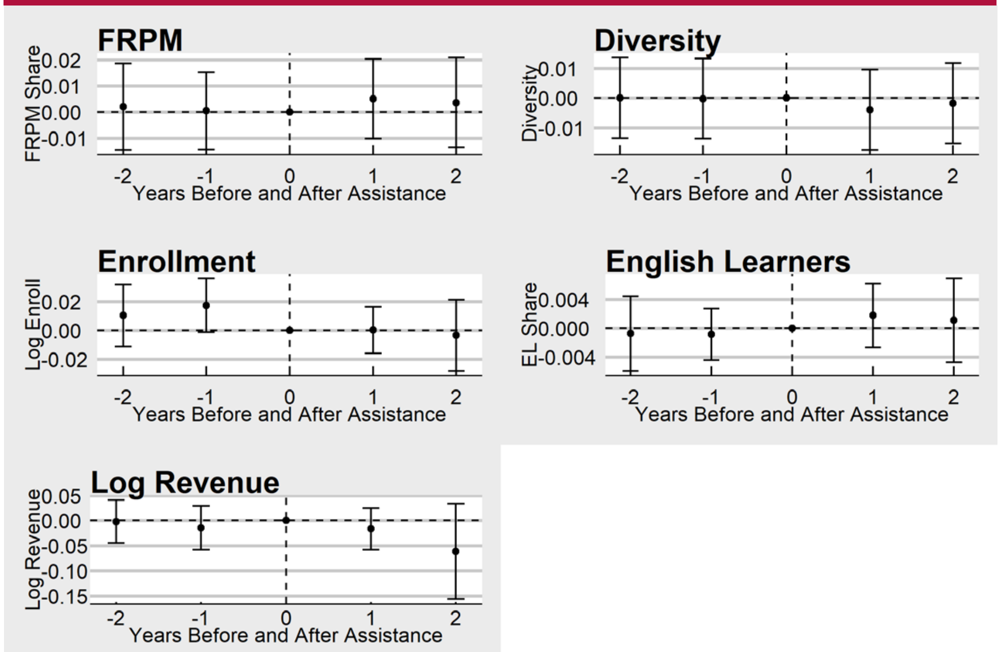


Figure 9 depicts differences between three groups. The dotted line represents the performance of differentiated assistance districts across our analytic window, while the dashed line represents the performance of the subset of 270 low status and growth general assistance districts. Their performance illustrates the same 'Ashenfelter dip' during the eligibility year. Comparatively, the solid line represents districts with high status, those districts in the green and blue on the California Schools Dashboard. It is highly improbable that any of these districts would become eligible for DA. As evidenced in the graph, their performance actually improved during the eligibility year. This finding further substantiates Ashenfelter's claim regarding the existential threat of accountability on performance. Additionally, another disturbing attribute of this visualization is that most low-performing districts have yet to return to their initial performance levels prior to the accountability policy shift.

COMPOSITIONAL CHANGES AND ALTERNATIVE OUTCOMES

Our second robustness check examined the difference between assistance eligibility and district characteristics over time. The inclusion of district-fixed effects controls for unobservable heterogeneity between districts averaged over the analytic period of this study. However, bias could exist if assistance districts experienced changes within our analytic window different from the changes experienced by non-assistance districts, which may have led to their eligibility for assistance (e.g., a loss of higher-performing students and a subsequent compositional increase in the percentage of students receiving free and reduced-price lunch). Districts might have also experienced compositional changes during treatment that affected outcomes of interest. For example, one can imagine a district where students exited based on changes implemented because of DA assignment. Other students may exit when the state identified their district as low-performing. To account for these potential confounding situations, we examined changes in time-invariant characteristics including (a) proportion of students receiving free and reduced-price meals, (b) proportion of English learners, (c) proportion of student racial/ethnic composition, (d) the log of district enrollment, and (e) the log of pupil revenue. We expected to observe no effect of assistance on these compositional outcomes.

FIGURE 10. Effects of Differentiated Assistance on District Characteristics.



Most point estimates are small and less than zero in Figure 10 suggesting that district composition is stable with respect to these characteristics during our analytic window. Assistance had no effect on district composition. Furthermore, these estimates also serve as another robustness check. Because these district characteristics represent alternative outcomes to testing, the share of these characteristics for our assistance districts was no different in the first or second year after the policy change compared to the share in the non-assistance group. Had compositional changes been significant, we could not rule out the extent to which these changes produced differences in achievement versus assistance.

SECTION 8

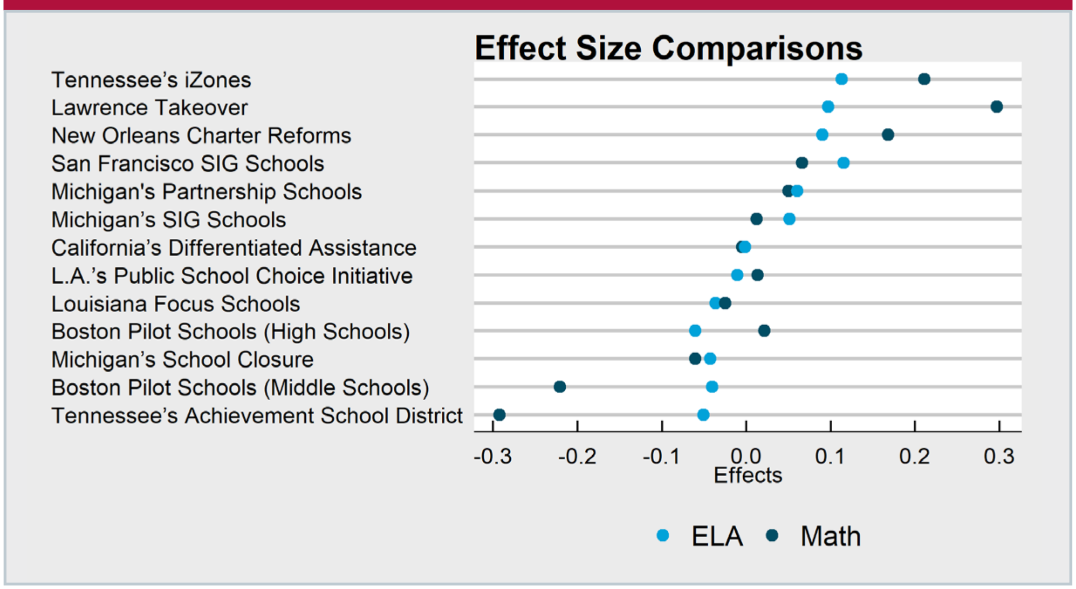
TAKEAWAYS AND POLICY IMPLICATIONS

SECTION 8: KEY TAKEAWAYS AND POLICY IMPLICATIONS

Accountability provides an important set of policies that states deploy to hold schools responsible for raising student achievement. The present report used panel data from California public schools along with a difference-in-differences analysis to simulate an experimental research design to determine the effects of differentiated assistance on student achievement. Results suggested no benefit to district achievement when receiving additional assistance. Districts receiving assistance, on average, had achievement score declines of .002 in ELA and .006 in mathematics during the post-policy period compared to districts in general assistance. To put the declines in context, a .002 standard deviation decline on a 4th-grade ELA assessment (Mean = 2460, SD 96) translates into a drop of less than one point in a school district's 4th-grade average ELA performance. The magnitude of the declines are small and not noticeably different from zero. The DDD analysis suggested a slight narrowing of the achievement gap in ELA and math after the first year of implementation and a return to eligibility year levels in year two. None of the changes in the achievement gap was significantly different from zero. Based on the findings, little evidence exists to suggest differentiated assistance improves achievement and reduces achievement gaps in a meaningful way. Given these results, what are some key takeaways and implications moving forward?

Improvement is relative and it may take longer than two years for assistance to have a meaningful effect on student achievement. Figure 11 depicts the effects of accountability reform efforts over the past 10 years.

FIGURE 11. Effects of Differentiated Assistance Compared to Other District-Level Interventions.



One observation from Figure 11 is that an accountability policy with the effects of differentiated assistance (essentially no effect) places it near the center of the distribution in terms of accountability policy effectiveness. Policymakers making decisions about the efficacy of differentiated assistance must recognize that most accountability initiatives fail to show effects on student achievement within the first two years. Research on the effects of accountability interventions suggest they vary by years of treatment, intervention type, and key features of the intervention, but a limited number of interventions make a detectable difference on student achievement immediately (Schueler, B., Armstrong-Asher, C., Larned, K.E., Mehrotra, S., & Pollard, C. 2020).

Additionally, the findings in this report align with similar results in Boston, Michigan, and Tennessee; however, they differ considerably in the context of the reform effort. California's System of Support most closely resembles both the reform and effects found in Michigan's Partner School Model and Boston's Pilot Schools. In other reform efforts, states closed schools or reconstituted schools into new districts that operated independently from the schools' home districts. Tennessee's ASD schools, for example, mandated that districts adopt significant management and personnel changes and make substantial operational changes, like reforms prescribed by turnaround models. Michigan's school closure policies displaced students into neighboring schools. More aggressive structural changes assume that districts lack the capacity or resolve to change the status quo in these schools internally (Chubb & Moe, 1990). California's DA model performs better than other reform models while maintaining local decision-making about student needs. Furthermore, California's technical support provides additional financial resources to build capacity of districts to improve their systems without forcing districts to make impulsive structural changes to personnel and resources. Unlike other reform models, the DA process focuses on the system's capacity to drive change locally, on understanding schools' needs, knowing what actions will address the needs, while embedding the work in a data culture that values improvement over judgement. If DA has the right theory of action, research suggests it may take three to five years to see improvement (Berends, Bodilly, & Kirby, 2002). With only two years of post-policy data, it may take more time to determine whether DA improves achievement or not.

Avoiding the unintended consequences of accountability efforts may be as important as accomplishing the intended goals of the reform. California's System of Support may avoid some of the inauspicious issues associated with consequential accountability. California has not experienced widespread unintended consequences from its accountability model. One only needs to recall "erasure parties" in Atlanta (Bowers, Wilson, & Hyde, 2011), where teachers took extraordinary steps to improve achievement, to understand the unintended consequences of accountability. One reason for fewer unintended consequences with California's support system may be two-fold. First, the California Schools Dashboard drives performance in the accountability system. The Dashboard emphasizes numerous measures of school quality along with change over time in these measures, while reducing the emphasis on any single test score. Stressing the importance of multiple measures of school quality may lead to less corruption in achievement data (Nichols & Berliner, 2007). The second reason may be due to the emphasis of building system capacity at the district level. Former iterations of the NCLB Act focused on schools as the level of change. Schools educate students, but schools rarely command their own budgets, have complete autonomy for program selection, or set their own policies. These decisions occur at the district level often with input from community stakeholders. Therefore, building capacity at the district level and holding a district accountable for improvement might mitigate many of the unintended consequences of accountability that result when schools feel punished for actions beyond their control. Unlike other reform models that focus energy on replacing the structural components in a low performing system such as school principals and teachers, DA builds the internal capacity of leaders to improve the systems responsible for the structural components. While we report limited empirical evidence of the effects of DA on achievement, the state's focus on using multiple measures of quality in its Dashboard along with using DA as the primary consequence for low performance may avoid many of the unintended consequences that plague accountability efforts. In fact, policymakers might consider expanding the criteria for who receives assistance. If DA is the right theory of action, then many districts in our control group might potentially benefit from this assistance.

Unfortunately, not all the consequences of accountability are avoidable. Policymakers should initiate accountability reform policies with the knowledge that shifts in accountability may contribute to initial declines in achievement that affect low-performing schools disproportionally. The mere label of increased accountability leads to improvements in some performances and contractions in others (Merton, 1968). The achievement gap between high-performing districts in this report compared to DA districts increased significantly during the eligibility year. High performers perform better under pressure while low performers perform worse. Policymakers must consider how accountability labels create inequities in student outcomes and figure out ways to address these consequences.

Consistency in DA implementation may contribute heavily to variability in outcomes. Another important reason for a lack of evidence linking DA to student achievement may involve the implementation of DA across the state. Prior research on the implementation of DA and the California System of Support suggested (a) increasing COEs' capacity and expertise in order to improve consistency in improvement work across the state, (b) "tightening up" the relatively loose practice of differentiated assistance and extending the window of support from six months to a year, and (c) speeding up the release of accountability data so the work commences at the beginning of a school year versus the middle of the school year (Humphrey & O'Day, 2019). While state law mandates many of the ways COEs support districts, COEs vary considerably in the Ad Hoc services provided to school districts (such as migrant education programs and preschool initiatives) and in the number and size of school districts they serve (some serve 90 districts with millions of students while others serve seven districts with several thousand students). Humphrey and O'Day's work suggest these structural and geographic elements create various levels of expertise in COEs and varying levels of coherence in COE services. Future work should focus on answering these questions:

PROCESS

1. **How do COEs measure and monitor the DA process to ensure districts are on track to achieve longer-term outcomes? How are COEs consolidating and using what it learns from implementing DA?**
2. **How does the DA process address equity? How do COEs weave equity through the root cause analysis, action planning, and other aspects of the intervention?**
3. **How does the DA process use the California Schools Dashboard to make decisions and how do COEs and districts navigate the messiness of multiple layers of data?**

CHANGES

4. **How does the DA process ensure that districts are making adaptive changes to their systems versus purely technical changes? What challenges do DA districts face implementing adaptive changes?**

DA ELIGIBILITY/EXIT

5. **How are DA eligibility and exit criteria validated? Are districts exiting DA after demonstrating sustained improvements across the eight state priorities? Are all districts that truly need support eligible?**

DA THEORY OF ACTION

6. **How is DA intended to work, and how is the theory of action being refined over time? What are the actions and expectations for DA in year 1? Year 2? Year 3?**

OTHER

7. **How have DA districts' acceptance, awareness, and attitudes about DA changed?**
8. **How do COEs address capacity issues to provide the right intensity of support needed to make change happen in the district system?**

A better understanding of how to address implementation challenges may consolidate DA practices across COEs and provide policymakers with important connections between practice and outcomes.

A final caveat about the limited evidence of DA on student achievement involves making causal claims about district-level interventions. First, we cannot entirely rule out the possibility that assistance and non-assistance districts differed on unobserved dimensions and biased our estimates. We cannot randomly assign districts to assistance; rather the state assigns districts to assistance because districts fail to meet criteria for improvement. Non-random assignment to DA may introduce selection bias into our estimates as well. While we have taken certain steps and present evidence for the complementarity of counterfactuals, the possibility remains that assistance and non-assistance districts differ in ways that may influence our causal estimates. Additionally, it is also possible that aggregate, district-level data might mask certain school influences and specific school district characteristics fundamental to the relationship between student achievement and DA. Even with certain limitations, this study provides valuable insight into California's efforts to hold schools accountable for improving student achievement. Future work will undoubtedly lead to refinements to the state's accountability model that support greater opportunities for all of California's children.

REFERENCES

REFERENCES

Abadie, A., & Cattaneo, M. D. (2018). Econometric methods for program evaluation. *Annual Review of Economics*, 465-503.

Abdulkadiroglu, A., Angrist, J., Dynarski, S., Kane, T., & Pathak, P. (2011). Accountability and flexibility in public schools: Evidence from Boston's charters and pilots. *The Quarterly Journal of Economics*, 699-748.

Angrist, J., & Pischke, J. (2009). *Mostly harmless econometrics*. Princeton: Princeton University Press.

Ashenfelter, O. (1978). Estimating the effect of training programs on earnings. *Review of Economics and Statistics* (60), 47-57.

Berends, M., Bodilly, S., & Kirby, S. (2002). Facing the challenges of whole-school reform: New American schools after a decade. Santa Monica: RAND Corporation.

Bertrand, M., Duflo, E., & Mullainathan, S. (2002). How much should we trust difference-in-differences estimates? Cambridge: National Bureau of Econcomic Research.

Bonilla, S., & Dee, T. (2020). The effects of school reform under NCLB waivers: Evidence from focus schools in Kentucky. *Education Finance and Policy*, 15(1), 75-103.

Booher-Jennings, J. (2005). Below the bubble: Educational triage and the Texas accountability system. *American Educational Research Journal*, 42(2), 231-268.

Bowers, M., Wilson, R. E., & Hyde, R. L. (2011). Special investigation into test tampering in Atlanta's school system. Atlanta: Office of the Governor: Special Investigators.

Bryk, A., Gomez, L., Grunow, A., & LeMahieu, P. (2015). *Learning to improve: How America's schools can get better at getting better*. Cambridge: Harvard Education Press.

California Department of Education. (2020, September 9). California School Dashboard and System of Support. Retrieved from California School Dashboard and System of Support - Accountability (CA Dept. of Education): https://www.cde.ca.gov/ta/aC/cm/

Chingos, M., Whitehurst, G., & Gallaher, M. (2013). School districts and student achievement. Providence: The Brown Center on Education Policy.

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

Crocco, M., & Costigan, A. (2007). The narrowing of the curriculum and pedagogy in the age of accountability: Urban educators speak out. *Urban Education* (42), 512-535.

Cszjka, J. C., Hirabayashi, S., Little, R., & Rubins, D. B. (1992). Projecting from advance data using propensity modeling. *Journal of Business and Economic Statistics*, 117-131.

Dee, T. (2012). School turnarounds: Evidence from the 2009 stimulus. Washington. D.C.: NBER Working Paper 17990. Retrieved from http://www.nber.org/papers/w17990.pdf

Dee, T., & Dizon-Ross, E. (2019). School performance, accountability, and waiver reforms: Evidence from Louisiana. *Educational Evaluation and Policy Analysis*, 41(3), 316-349.

Dragoset, L., Thomas, J., Hermann, M., Deke, J., James-Burdumy, S., Graczewsld, C., & Giffin, J. (2017). School improvement grants: Implementation and effectiveness. U.S. Department of Education, Institute of Education Sciences. Washington, D.C.: National Center for Education Evaluation and Regional Assistance. Retrieved from http://ies.ed.gov/ncee

Hanushek, E., & Raymond, M. (2005). Does school accountability lead to improved student performance? *Journal of Policy Analysis and Management* (24), 297-327.

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

Humphrey, D. C., & O'Day, J. (2019). The early implementation of California's system of support: Counties, differentiated assistance, and the new school dashboard. Stanford: Policy Analysis for California Education.

Johnson, S., Marietta, G., Higgins, M., Mapp, K., & Grossman, A. (2015). *Achieving coherence in district improvement*. Cambridge: Harvard Education Press.

Kane, T., & Staiger, D. (2002). The promise and pitfalls of using imprecise school accountability measures. *Journal of Economic Perspectives* (16), 91-114.

Knudson, J., Shamburgh, L., & O'Day, J. (2011). Beyond the school: Exploring a systematic approach to school turnaround. San Mateo: California Collaborative on District Reform.

Merton, R. K. (1968). The Matthew effect in science. *Science* (5), 56-63.

Neal, D., & Schanzenbach, D. (2007). Left behind by design: Proficiency counts and test-based accountability. NBER Working Paper No. 13293, JEL No. D78, I2, J41.

Nichols, S. L., & Berliner, D. C. (2007). *Collateral damage: How high-stakes testing corrupts America's schools.* Cambridge: Harvard Education Press.

Ouchi, W. (2006). Power to the Principals: Decentralization in three large school districts. Organizational Science, 17(2), 298-307.

Robins, J. M., Hernan, M. A., & Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology*, 550-560.

Rubin, D. B. (1973). The use of matched sampling and regression adjustment to remove bias in observational studies. *Biometrics*, 29(1), 185-203.

Schueler, B. E., Goodman, J. S., & Deming, D. J. (2017). Can states take over and turn around school districts? Evidence from Lawrence, Massachusetts. *Educational Evaluation and Policy Analysis*, 1-24.

Schueler, B. E., Armstrong-Asher, C., Larned, K.E., Mehrotra, S., Pollard, C. (2020). Improving low-performing schools: A metanalysis of impact evaluation studies. (EdWorkingPaper: 20-274). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/qxjk-yq91

Strunk, K. O., Marsh, J. A., Hashim, A., Bush, S., & Weinstein, T. (2016). The impact of turnaround reform on student outcomes: Evidence and insights from the Los Angeles Unified School District. *Education Finance and Policy* (11), 251-282.

U.S. Department of Education. (2010). Guidance on fiscal year 2010 school improvement grants under section 1003(g) of the Elementary and Secondary Education Act of 1965. Washington, D.C.: Office of Elementary and Secondary Education.

White, K.W. & Rosenbaum, J.E. (2013). Inside the black box of accountability: How high-stakes accountability alters school culture and the classification and treatment of students and teachers. In *No Child Left Behind and the Reduction of the Achievement Gap: Sociological Perspectives on Federal Educational Policy*. Taylor and Francis, pp. 97-116.

Wong, K., & Meyer, S. (1998). Title I school wide programs: A synthesis of findings from recent evaluations. Educational *Evaluation and Policy Analysis* (20), 115-136.

Wong, K., & Shen, F. (2002). Politics of state-led reform in education: Market competition and electoral dynamics. *Educational Policy*, 161-192.

Wong, K., & Shen, F. (2003). Measuring the effectiveness of city and state takeover as a school reform strategy. *Peabody Journal of Education*, 88-119.

Zavadsky, H. (2012). *School turnarounds: The essential role of districts*. Cambridge: Harvard Education Press.

Zimmer, R., Kho, A., Henry, G., & Viano, S. (2015). Evaluation of the effects of Tennessee's Achievement School District on student test scores. Nashville: Tennessee Consortium on Research, Evaluation and Development.

FOR MORE INFORMATION CONTACT:
Shannon E. Coulter, Ph.D.
Director, Research And Evaluation

San Diego Office of Education
6401 Linda Vista Drive
San Diego, Ca 92111
scoulter@sdcoe.net
858-295-8825

san diego county office of
EDUCATION
FUTURE WITHOUT BOUNDARIES℠