ANNENBERG
BROWN UNIVERSITY

EdWorkingPaper No. 20-177

The Longitudinal Effects of School Improvement Grants

Min Sun
University of Washington

Alec Kennedy
San Francisco Unified Public Schools

Susanna Loeb
Brown University

School Improvement Grants (SIG) represent one type of governments' capacity-building investment to spur sustainable changes in America's persistently under-performing public schools. This study examines both short- and long-run effects of the first two cohorts of SIG schools from two states and two urban districts across the country. Using dynamic event analyses, we observe that SIG showed larger effects in the second and third years of the intervention than the first year on 3-8th grade student test scores—a pattern of gradually increase over the course the intervention. These positive effects are largely sustained three or four years after the funding ended. In high schools, the SIG effects on 4-year graduation rates were steadily increasing throughout the period of six or seven years after the initial start of the intervention. These patterns of SIG effects mostly apply to each of the four locations, but the magnitude of effects varies across locations, suggesting differential implementations. Moreover, SIG effects on students of color or low-socioeconomic students are similar to, and sometimes a bit larger than, the overall SIG effects. We also conduct a variety of sensitivity and robustness checks. Lastly, we discuss the policy implications of our findings on states' continuing efforts of transforming public organizations and building their long-term capacity for better performance.

VERSION: January 2020

The Longitudinal Effects of School Improvement Grants

Min Sun

University of Washington

Alec Kennedy

San Francisco Unified Public Schools

Susanna Loeb

Brown University

Acknowledgement

Min Sun's contribution to this study is supported by grants from the National Science Foundation under Grant No. DRL-1506494 and from the Spencer Foundation under Grant No. 201900028. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funders. We also thank our partners at each of the four locations (WA OSPI, NCERDC, SFUSD, and BCPS) for sharing their data.

Abstract

School Improvement Grants (SIG) represent one type of governments' capacity-building investment to spur sustainable changes in America's persistently under-performing public schools. This study examines both short- and long-run effects of the first two cohorts of SIG schools from two states and two urban districts across the country. Using dynamic event analyses, we observe that SIG showed larger effects in the second and third years of the intervention than the first year on 3-8th grade student test scores—a pattern of gradually increase over the course the intervention. These positive effects are largely sustained three or four years after the funding ended. In high schools, the SIG effects on 4-year graduation rates were steadily increasing throughout the period of six or seven years after the initial start of the intervention. These patterns of SIG effects mostly apply to each of the four locations, but the magnitude of effects varies across locations, suggesting differential implementations. Moreover, SIG effects on students of color or low-socioeconomic students are similar to, and sometimes a bit larger than, the overall SIG effects. We also conduct a variety of sensitivity and robustness checks. Lastly, we discuss the policy implications of our findings on states' continuing efforts of transforming public organizations and building their long-term capacity for better performance.

Introduction

Substantial investments by government entities in building organizations' capacity are frequently used as policy instruments to spur sustainable, positive changes in public sectors, for example, public health (e.g., Decorby-Watson, Mensah, Bergeron et al., 2018) and education (e.g., Sun, Penner, Loeb, 2017; Strunk, McEachin, & Westover, 2014). The evaluation on these strategies are often lack of rigor and lack of longer-term data that captures the sustainability at the organizational level (Decorby-Watson et al., 2018). Information on both short-run and the long-run effects can help policy makers assess the returns on these investments. Information on short run impacts can provide more immediate insights into the likely effects of the program, but these impacts may build over time with short term results underestimating long-run returns, or their effects may fade, even negating initially positive indications.

School Improvement Grants (SIG) exemplify such capacity-building investment. SIGs were a signature policy of the Obama's administration. In an effort to incentivize the transformation of the nation's persistently lowest performing public schools, Congress appropriated $3.5 billion for the first cohort of SIGs through the American Recovery and Reinvestment Act (ARRA) of 2009. Congress continued the investment raising funds to a total of approximately $7 billion for subsequent five cohorts of SIGs (U.S. Chamber of Commerce Foundation, 2010; U.S. Department of Education, 2010; U.S. Department of Education, 2010). These funds typically doubled the grantee schools' regular budget and were available to schools for at least 3 years. All identified schools needed to undertake dramatic transformations, such as replacing the principal, implementing curricular reform, and tying teacher evaluation results to personnel decisions. States also provided technical assistance and coaching, aiming to build these schools' capacity to remedy underperformance.

While SIGs were a substantial investment, the research estimating their effects in the first few years have been mixed, with some finding quite positive results and others no evident change (Sun et al., 2017; Dee, 2012; Dragoset et. al., 2017). For example, Dee (2012) used a "fuzzy" regression discontinuity design and found significant improvements in post-treatment performance in California schools whose baseline proficiency rate just met the lowest achieving threshold. In contrast, Dickey-Griffith (2013) used a difference-in-differences approach to assess one-year impacts in Texas and found mixed results, including negative impacts on student achievement in elementary and middle school, but positive effects on high school graduation rates. No study that we know of has looked at the effects of SIG interventions beyond the three intervention years.

In this paper, we examine the effects of the first two SIG cohorts of 99 schools from two states and two urban districts: North Carolina (NC), Washington state (WA), San Francisco Unified School District (SFUSD), and Beachfront County Public Schools (BCPS, Pseudonym[1]). Our study has several strengths. First, the current paper constitutes the first study that comprehensively documents the longitudinal effects of SIGs on school performance. We compiled extended data that include three years before the SIG award and six or seven years after the award. This longitudinal analysis is more aligned with the policy intent and design—building schools' capacity for long-run success. Second, our data from these four locations across the nation represent a geographically diverse group of states and local districts. Although SIG programs are highly prescriptive compared to other federal capacity-building initiatives, local contexts and capacities play a role in the implementation of SIG programs (Carlson & Lavertu, 2018; Ginsburg & Smith, 2018). This study includes both a pooled analysis across these

[1] We are waiting on the approval of using this district's real name. We expect to hear back from them soon.

four locations so that we can see the overall effects and separate analysis for each location so that we can observe the heterogenous effects across locations. Third, the study assesses effects for subgroups of students as well as for the full sample. We analyze SIG effects for historically underserved students of color and for students from low socio-economic status (SES) families. Lastly, we conduct a battery of sensitivity analyses and robustness check to rule out other possible explanations for identified SIG effects (Wong, Valentine, & Miller-Baine, 2017).

We use dynamic event analyses based on a difference-in-differences (DiD) framework to unravel gradually increasing positive effects of SIG in math and reading/English Language Arts (ELA) test scores over the course of the intervention years in elementary and middle school grades (grades 3-8). SIG showed larger effects in the second and third years of the intervention than the first year. After SIG funding ends, we find that these positive effects on math and reading/ELA test scores started to slightly decrease; however, the positive policy effects for math sustained by the third or fourth year post policy (i.e., six or seven years after the school initially received the grant). We also observe that turnaround schools that adopted more dramatic reform strategies saw larger increases in test scores than transformation schools during treatment years and were more able to sustain these positive effects after the funding ended. High schools had steady increases in their 4-year graduation rates throughout the six or seven years, with a five percentage point increase in year one of the reform and a 14 percentage point increase in year six and beyond. Moreover, the patterns of SIG effects in separate locations are largely consistent in pattern, but show different magnitudes, suggesting differential implementations across locations. In addition, SIG effects on historically underserved student groups are similar to, and sometimes a bit larger than, the overall SIG effects.

The findings are relevant for states and localities beginning to implement the Every

Student Succeeds Act (ESSA) of 2015, as the new federal law continues to require that states spent seven percent of state Title I funds, over $1 billion a year, on turning around states' lowest performing schools. While states now have wider flexibilities to design, implement, and monitor this new generation of school improvement work, an understanding of the nation's return on its investment in SIGs can help state and school leaders to select programs to promote student learning in struggling schools. The next sections briefly summarize the SIG program design and implementation and review the literature that has assessed the short-term impacts of SIG programs across the country. The following sections outline the data and methods, and provide results. The final section discusses the policy implications of our findings, contribution to literature, and opportunities for future research.

Policy Background—the SIG Program: SIGs, authorized under section 1003(g) of Title I of Elementary and Secondary Education Act of 1965 (ESEA), were grants to state educational agencies (SEAs) for SEAs to use to make competitive subgrants to local educational agencies (LEAs). The awarded LEAs were supposed to demonstrate great need for the funds and strong commitment to use the funds to provide adequate resources to substantially raise the achievement of students in their lowest-performing schools.

Prior to ARRA, the federal funding levels for remedying underperformance in the nation's public schools had been substantially lower. For instance, the US Department of Education appropriated $491,265 in fiscal year 2008, whereas in fiscal year 2009 under ARRA, they appropriated more than $3.5 billion to states to be used over a three-year implementation period by Cohort 1 SIG schools (2010–11 to 2012–13 school years). In fiscal years 2010, 2011, and 2012, they appropriated a total of $1.6 billion to fund a second round of SIG schools (Cohort 2). In Cohort 1 and 2, each grantee school received between $50,000 and $2 million per year—an award that was double the size of their regular school budget—to implement reforms for three years. Most states made fewer SIG awards in Cohort 2 than in Cohort 1, but on average, annual per-pupil award amounts were larger in Cohort 2.

The SIG program targeted the persistently lowest-achieving schools in each state. In Cohorts 1 and 2, the persistently lowest-achieving schools were typically defined as schools among the lowest five percent in terms of academic achievement level or growth. Among Cohort 1 and 2 schools, the majority of schools were selected based on placing in the bottom five percent of schools in the 3-year average proficiency rate for all students on state assessments in reading/ELA and math (combined). High schools were also eligible if their adjusted five-year cohort graduation rate for all students was less than 60 percent. LEAs applied to the SEAs on behalf of some or all eligible schools. The states then competitively awarded to schools that met the eligibility criteria and other determination criteria, such as district capacity and commitment to support school turnaround or geographical locations of the schools in the states.

SIG schools were required to adopt one of four school reform models that aimed to overhaul existing practices. The *transformation model* required replacing the principal, implementing significant instructional reform, increasing learning time, and developing teacher- and leader-evaluation systems that took student progress into account and were tied to personnel decisions (e.g., rewards, promotions, retentions, and firing). The *turnaround model* included all of the transformation model requirements, along with replacing at least 50 percent of the staff. The *restart model* required the school to close and reopen under the leadership of a charter or education management organization. Finally, the *closure model* simply closed the school. Over 75 percent of SIG schools choose the transformation model, while 20 percent chose the turnaround, and five percent selected the "restart" model. The "closure" model was rarely chosen (Ginsburg & Smith, 2018). As a result, over 99 percent of the SIG funds went to turnaround and transformation schools (Hurlburt, Therriault, Le Floch, & Wei, 2012). In this study, we assess effects overall and compare the effects of turnaround with those of transformation.

Schools used the funds for a variety of improvement purposes, such as funding additional professional development for school leaders and teachers, hiring additional staff members, providing co-planning time for teachers, extending school days, providing differential rewards to teachers based on their performance, implementing curriculum reforms, using student data to adjust instruction, and engaging parents and communities (Sun et al., 2017; Sun et al., 2019). States typically supplemented local investments by providing designated support staff (e.g., WA provided school improvement coaches to each SIG school), professional development for SIG-awarded districts or schools (e.g., NC offered a principal leadership institute), and improvement tools (such as Indistar—an online planning and implementation tools adopted by both NC and WA). States in all our locations monitored schools' progress annually to determine whether the SIG funds should continue. Monitoring strategies include a combination of in-person site visits (CA, BCPS, NC, WA), designation of staff assigned to specific districts or schools (BCPS), check-in meetings (e.g., in person, or telephone, BCPS, WA), and online tools (e.g., Indistar used by WA). States used a variety of measures to monitor school progress, including student academic progress (e.g., proficiency levels, graduation and dropout rate, academic growth patterns, and percentage of students completing advanced coursework), student connection and school climate (e.g., student attendance rate, discipline incidents, and truants), and staff talent management (e.g., teacher or principal performance distributions and teacher attendance rate).

California where SFUSD is located did not apply for NCLB waivers, therefore, all of its SIG schools discontinued at the end of the three years. NC, WA, and BCPS received NCLB waivers in 2012-13. SIG schools in these three locations could exit the designation of underperformance if they showed substantial progress. The federal funds to those schools were discontinued as a result. SIG schools could also be identified as priority or focus schools after the three years of SIG programs, if they continued to be the lowest five percent based on student performance (Priority schools) or if they were within the lowest ten percent in subgroup performance (Focus schools). Of the 99 SIG schools in our sample, 84 were designated as Priority after SIG funds were discontinued. Of those 84 schools, 14 were moved to a Focus designation after the initial 3 years following the end of SIG. The 15 other SIG schools exited Priority/Focus status after SIG ended. Although less intense interventions than the SIG program, priority and focus schools continued to receive monitoring, technical assistance, and financial supports from the states. These financial supports are substantially lower than those from the SIGs.

Literature Review: Extant research evidence shows a mixed record of SIG impacts. Early research on SIG examined the first-year impact of these often multi-year reforms, finding mixed results (Dee, 2012; Dickey-Griffith, 2013). More recent work examined the impacts beyond the first year of the school improvement efforts, again showing mixed evidence of their effectiveness. The author's study in SFUSD showed a pronounced, positive impact of Cohort I SIG interventions on student achievement in year three (Sun et al., 2017). Similar positive effects emerged in the first year of reform and grew through the fourth year in Massachusetts (Papay & Hannon, 2018). Positive effects in math and ELA were found in the 19 SIG award schools (including both Cohort 1 and Cohort 2) in Colorado (Colorado Department of Education, 2015). Substantial positive effects were also identified in Ohio Cohort 1 and 2 SIG schools: the effect of 0.24 standard deviation in math and reading, and 7-9 percentage points increase in high school graduation rate (Carlson & Lavertu, 2018). In contrast, a study commissioned by the U.S.

Department of Education using data from 22 states did not find positive effects on test scores, high school graduation, or college enrollment for the Cohort 1 SIG schools, though the estimates were not precise enough to rule out positive effects in line with the other studies (Dragoset et al., 2017). More in contrast, studies in North Carolina showed null to significantly negative effects of SIG efforts on state test scores and graduation rates (Heissel, & Ladd, 2016; Henry, Guthrie, & Townsend, 2015). Some of the disparities in these results may be explained by sample selection and estimation strategies, as Guthrie and Henry's (2016) work in North Carolina illustrates, though differences in findings across studies likely also stem from variation in the design and implementation of school reform interventions across schools, districts, and states.

Some prior studies have compared the differences in effects between schools that chose the turnaround model and those that chose the transformation model. Most of these studies to date show that the turnaround model is more effective than the transformation model (Sun et al., 2017; Carlson & Lavertu, 2018; Dee, 2012; Dragoset et al., 2017), though a New Jersey report found that the transformation model was more effective than turnaround (Kyse, Swann-Jackson, Marini et al., 2014).

Several studies reveal trends in improvement over time rather. These studies tend to find a pattern of gradual, rather than sudden, gains. For studies that measure interim gains over three years of the grant period, improvement was greater in year two than in year one (such as in Colorado, Massachusetts, Ohio, San Francisco, and Tennessee). While San Francisco and Massachusetts show better outcomes in year three than year two, the achievement levels are not substantially better in year three than year two in Colorado, Ohio, and Tennessee. These gradually emerging program effects suggest the importance of understanding whether the effects sustained beyond the third year of implementation, particularly after the withdrawal of the substantial financial support and intense public accountability. This study aims to shed light on this question.

Data and Sample

Our analysis focuses on estimating the SIG program effects on student achievement and graduation rates in the first two SIG cohorts across four locations: NC, WA, BCPS, and SFUSD. We remove schools that were selected as part of later SIG cohorts (Cohort 3 and 4 were offered in WA and NC, respectively) which include two Cohort 1 schools in Washington. We also only include schools that adopted either transformation or turnaround models in our analysis. This sample includes 66 Cohort 1 schools that were awarded funding starting in the 2010-11 school year: 23 schools in NC, 15 schools in WA, 19 schools in BCPS, and 9 schools in SFUSD. The second cohort of 33 SIG schools were awarded their funding in the next year (2011-12): 17 schools in NC, 10 schools in WA, 6 schools in BCPS, and no schools in SFUSD.

Our analysis uses state and district administrative datasets on schools' student characteristics, state spring standardized tests in math and reading/ELA, graduation rate, and school contexts. The longitudinal nature of the data spans nearly a decade, from the 2007-08 to 2016-17 school years, covering three years prior to the start of SIG in the first cohort of schools and three years after the withdrawal of funding.[2] We collected and matched graduation rate data from publicly available sources (state or district websites). Further, by linking students' geo-addresses with the U.S. Census's American Community Survey (ACS) data, we obtained the five-year characteristics of neighborhoods where the students lived, including the log of median household income, percent with a bachelor's degree or higher among residents who are 25 and

[2] At the time of analysis, the data we have access to for each location spanned over different time periods. NC had data starting in 2007-08 and ending in 2016-17. WA had test score data starting in 2009-10 (only one pre-reform year) through 2016-17. BCPS had data from 2007-08 through 2016-17. Our data for SFUSD began in 2007-08 and ended in 2016-17 (in 2014-15, no students were tested on a state standardized test).

older, percent of residents 18 or under living below the poverty threshold, and the log of median housing value (owner occupied). Student address information was available for all locations *except* BCPS. To define students' SES, we use students' geocoded ACS data to derive a composite factor score across the three neighborhood dimensions. We then define the bottom 20% of the composite score as low-SES students.

Table 1 provides descriptive statistics of baseline student and school attributes during the three-year pre-reform period comparing SIG and non-SIG schools. These two types of schools differ significantly on almost all of the observed pre-intervention school characteristics for both the pooled sample and within each location. For example, SIG schools served students who were lower-achieving, had lower graduation rates, and were more likely to be people of color, English learners (EL), and in special education programs.

[Insert Table 1 here]

Analytic Strategies

We use an event study model, also known as a Granger-style difference-in-differences model, in order to examine the dynamic nature of the SIG treatment effects, both during and after SIG years (Angrist & Pischke, 2008; Autor, 2003; Sun et al., 2017; Taylor & Tyler, 2012). Our approach essentially tests whether treatment schools experienced higher performance (e.g., higher test score or graduation rate) during or after the intervention when compared to non-treatment schools, relative to pre-reform differences between treatment and non-treatment schools, controlling for changes in their students' demographic characteristics.

Our analysis is conducted at school-by-year level. The main benefit of conducting a school-by-year level analysis is that we can follow the school even as cohorts of students move through it. This approach of using school as the unit of analysis is consistent with the program design of building organizational capacity. The approach is also necessary for estimating the sustainable effects of SIGs, as the average elementary student has at most 3 years of test scores. Equation (1) describes our basic model specification.

$$S_{jt} = \alpha_0 + \sum_{i=1}^{3} \beta_i (SIG_j * Year_i) + \sum_{i=4}^{6+} \beta_i (EverSIG_j * Year_i) + \tau_t + \gamma_j + \tau_t * Location_j + X_{jt} + \varepsilon_{jt} \quad (1)$$

where S_{jt} is the math, reading/ELA standardized test score (grade 3-8), or graduation rate for school j in year t. We standardize the scores for a given test, grade, location, and year to account for differences in tests across locales. Although the subscript for subjects is omitted, we conduct the estimation separately for math and reading/ELA. We use the 4-year graduation rate as an outcome for high schools. $Year_i$ is an indicator for the i^{th} year since SIG started (e.g., $i = 1$, indicates 2010-11 for Cohort 1 SIG schools and 2011-12 for Cohort 2 SIG schools). $(SIG)_j$ is a time-invariant school-level indicator for the schools who were selected to receive SIGs. β_1, β_2, or β_3 indicates the treatment effect estimate during each of the treatment years. β_4, β_5, and β_{6+} indicates treatment effect estimates in post-treatment year 1, year 2, and year 3+. Equation (1) provides a flexible model specification to examine non-linear school reform effects.

Our goal is to estimate the SIG effect on average achievement and graduation, net of location-wide and school-specific factors that may also influence the change in student outcomes. We include school fixed effects, γ_j, to control for time-invariant heterogeneity across schools. We also include year fixed effects (τ_t) to control for yearly shocks and general trends affecting student outcomes across all schools. Additionally, other factors might influence student outcomes within the school's geographic region over time, which could bias the estimates if such factors do not change at a national level or get picked up by year fixed effects. In the pooled analysis, we include region-by-year fixed effects, as indicated by $\tau_t * Location_j$, to account for region-specific changes in policy, economy, demographic, or social aspects that might influence the extent to which schools can improve their performance.

Even after we control for school-specific, region-year specific factors, as well as yearly shocks, time-varying school factors may influence the changes in student outcomes. For example, students were not randomly assigned to schools and the student populations might have changed during the course of SIG interventions. To address student selection bias, time-varying controls, $\mathbf{X}_{jt}$, include school averages of students' race and ethnicity, gender, and EL status, as well as logged school enrollment and school level (primary, middle, high, or other). This set of factors often correlate with other unobserved school changes that may influence school average performance. To illustrate, the change in students' demographics may signal the amount of community resources or parental supports to the school that may, in turn, affect student outcomes. Controlling for these time-varying school characteristics, as well as the full set of fixed effects, allows us to estimate SIG effects net of both observed and unobserved factors. ε_{st} is the error term. We estimate clustered robust standard errors at the school level to adjust for correlations within schools and the influence of a small number of treatment clusters on standard error estimates (Cameron & Miller, 2015). Lastly, we estimate potential differential effects of transformation and turnaround models[3], and for each cohort respectively by interacting the SIG model type or SIG cohort with each of the $Year_i$dummy variables.

We conduct several robustness and sensitivity checks. First, DiD designs assume that trends in treatment schools would have been the same as in non-treatment schools if they had not received the reforms. We examine pre-reform trends to assess the validity of this assumption. We

[3]Turnaround schools, on average, were lower achieving during pre-SIG years than transformation, thus, we did not pursue the heterogeneity analysis by pre-reforming performance level.

also control for pre-reform trends to examine whether non-parallel pre-trends would change the inferences of estimated treatment effects (e.g., Sun et al., 2017; Strunk, McEachin, & Westover, 2014).

Second, our estimation of SIG effects could be biased if the pre-reform performance levels differed between SIG and non-SIG schools. For example, SIG schools with low graduation rates would have more room to grow to reach the maximum of 100% than non-SIG schools with already high graduation rates. For achievement and graduation rate separately, we thus generate a set of comparison schools for each of SIG school in terms of both pre-reform performance level and trend from 2008 to 2010, student body served, grade span, and region fixed effects. We used nearest neighbor one-to-one propensity score matching without replacement.[4] Combining the DiD framework with propensity score matching gains internal validity, while it reduces the precision of the estimation due to reduced sample size and makes the heterogeneity effect estimations implausible for certain location and certain variable (e.g., high school graduation rate for WA). We use the matched estimates to triangulate the inferences from our benchmark models.

Third, other factors might have changed at the same time as the treatment reforms and these changes might affect student outcomes and bias our estimates of SIG effects. To address this potential bias, we first assessed synchronous education policy changes. To our knowledge, none of the four locations undertook other significant reforms targeting the treatment schools concurrently with their designation as SIG. However, we do know that many SIG schools were designated as either Priority or Focus schools after they had exited the SIG designation at the end

[4] We implemented alternative matching strategies, such as one-to-one matching with different caliper widths and one to multiple matching with replacement. The nearest neighbor one-to-one matching without replacement provides the most robust results. Alternative matching strategies yielded similar findings.

of their third year in reform. We thus examine the degree to which the post-reform effects were driven by the Priority or Focus designation for these schools.

Fourth, changes in student selection into SIG schools may mask the true effects of the SIG reforms on school quality. Entering cohorts of students during SIG years, for example, may have been higher or lower performing than in pre-reform years. We address this concern in part with the controls for student demographics but these controls may not be sufficient. To understand whether the student body did change in SIG schools, we use a model similar to Equation (1) but with the dependent variables as test scores of newly entering students prior to when they entered the SIG school. If we see that the prior performance of entering students changed during the SIG years, then we would have reason to be concerned about the validity of the estimated effects, and we would be able to predict whether our estimates likely underestimate or overestimate the true effect of program on the schools.

Finally, to assess the heterogeneity of SIG effects across locations and student subgroups, we run Equation (1) in each location separately, for historically underserved students of color (defined as non-White and non-Asian), and for economically disadvantaged students (defined using student neighborhood SES factors).

Results

For each outcome measure, we present the results for the pooled sample along with differential SIG effects for transformation and turnaround schools, and for each of the first two cohorts. We then discuss the robustness and sensitivity of these main estimates. We lastly include results of the consistency and variation of SIG effects across locations and for student subgroups.

SIG Effects from the Pooled Analysis on 3-8 Grade Student Achievement: As shown in Table 2 and Figure 1, SIG interventions significantly increased the average student achievement in math and reading/ELA for grades three through eight during the treatment years. The treatment effects are more pronounced in the second and third year of the intervention than the first year in the pooled sample. Figure 1 compares the trends in average student achievement in math and reading/ELA for Cohort 1/Cohort 2 SIG and non-SIG schools. Figure 1a and Figure 1b shows that prior to reform the average math scores of the SIG schools were considerably lower than the average of the non-SIG schools: Cohort 1 schools were 0.5 standard deviations (SD) lower and Cohort 2 schools were 0.6 SD lower. Notably, the pre-trends are nearly parallel between the SIG schools and non-SIG schools. After Fall 2010, in obvious contrast to the pre-trend, the mean math achievement raised much more quickly in SIG schools than in non-SIG schools. By the third year of intervention, the gaps in average math achievement were the smallest between SIG and non-SIG schools.

After the SIG award ended, the effects mostly sustain after one year, and show some positive results even three years post intervention, especially for math. Table 2 show that three years after the end of the program the math effects are still positive and statistically significant (0.154 SD), while the estimates for ELA are approximately half as big as they were at the end of the program (0.109) and not statistically distinguishable from zero.

[Insert Table 2]

[Insert Figure 1]

Although transformation and turnaround schools adopted many similar interventions, turnaround schools also replaced leaders and staff, potentially resulting in different treatment effects. As shown in the "Trans" and "Turn" columns in Table 2 (abbreviations for

Transformation and Turnaround models, respectively), turnaround schools show somewhat larger effects than transformation schools across all years in math, though the differences are often not statistically different from zero.[5] For example, in year 1 of the treatment, the estimated effect on average improvement in math is 0.10 SD in transformation schools, which is smaller than the estimated 0.14 SD change in turnaround schools. Similarly, in reform year three, the estimated average effect is 0.19 SD in transformation schools, compared to a much larger estimate of 0.30 SD in turnaround schools. One-year post treatment, transformation schools had an estimated average effect of 0.12 (a decrease from the third year of treatment) and turnaround schools had an estimated effect of 0.31 (approximately the same level of gain as the third year of the treatment). The patterns for reading/ELA are not as consistent, though in the majority of cases the point estimates for turnarounds are greater than that for transformation schools, but the differences are not statistically significant. The finding that slightly more pronounced effects are observed in turnaround schools is consistent with previous evidence from California and Ohio (Sun et al., 2017; Carlson & Lavertu, 2018; Dragoset et al., 2017; Dee, 2012).

In terms of the differential effects between Cohort 1 and Cohort 2 schools, we see that Cohort 2 schools have somewhat larger effects in reading/ELA than Cohort 1 across all years (again, most of those differences are not statistically significant). This pattern may suggest organizational learning in that useful lessons from cohort 1 schools may pass along to cohort 2, and state policymakers may learn better ways to support schools. Similar patterns that favor Cohort 2 are observed for math in the short run (the year 1 and year 2 of the reform), but the differences are not evident in the long run.

SIG Effects from the Pooled Analysis on High School Graduation Rates: Figure 2

[5] Only the first post-SIG year estimates in math are significantly different between transformation and turnaround schools. None of estimates for reading/ELA are significantly different.

and Table 3 provide the results for high school graduation rates. Again, SIG schools show an improvement relative to non-SIG schools, and the graduation effects persist post reform.[6] The SIG effects start as 5.96 percentage points in year 1, increase to 10.58 percentage points in year 3, and continuously raise to 15.59 percentage points in the third year of post reform. A possible explanation for the sustained effects is that graduation rates capture the delayed program effects when treated students moved through the school system. Another explanation might that students in later years experienced more years of treatment than the initial years of intervention.

The estimated effects on graduation rates are largely similar for transformation and turnaround schools. Although turnaround schools have lower graduation rates to start, they grow at the same pace as transformation schools. In contrast to the achievement results, we find consistently larger effects in Cohort 1 than Cohort 2 schools for the graduation effects. Although some Cohort 1 effects look twice as big as those in Cohort 2 schools, the differences are mostly not distinguishable given the precision of the estimates. The differential cohort effects may be explained by the fact that Cohort 2 schools had a higher graduation rate than Cohort 1 schools prior to the SIG intervention, as shown in Figure 2. Graduation rates have a ceiling (100%), which may restrict growth in higher graduation rate schools. However, while smaller, the effects in Cohort 2 schools are meaningful. As shown in Figure 2, the graduation rate of Cohort 2 high schools caught up with that of non-SIG schools both during and after the reform.

[Insert Table 3]

[Insert Figure 2]

[6] When conducting analysis of graduation rates for high schools, we excluded schools whose graduation rates were calculated based on fewer than 20 students, which results in 34 SIG high schools included in the analysis, including 25 Cohort 1 and 9 Cohort 2 high schools. The excluded high schools were often not conventional high schools or secondary schools. These schools also often had graduation rates of zero percentage points, which are either very skewed or inappropriately calculated. This criterion excluded 60 school-by-year observations across four locations.

Robustness of the Estimated SIG Effects from the Pooled Analyses: One key assumption of the difference-in-differences approach is that the changes from pre- to post-intervention in non-SIG schools provide a valid counterfactual for what would have happened in SIG schools if the interventions had not been implemented. Although we cannot prove this assumption, we closely examine the pre-trends to assess a possible violation. As shown in Figure 1, the pre-trends in achievement measures were almost parallel between SIG and non-SIG schools in the overall samples, suggesting common pre-trends. However, Figure 2 of graduation rates show a greater increase in SIG schools than non-SIG schools prior to the intervention. We test this threat to the common trends assumption for both achievement and graduation rate by adding pre-treatment yearly differential changes between SIG and non-SIG schools to Equation (1). As shown in Appendix Table A1 for achievement, the mostly null effects of pre-SIG differences between SIG and non-SIG schools support the causal interpretation of the positive SIG effects on student achievement. After accounting for pre-SIG indicators, the estimated SIG effects during and after the reform are very similar to those in our main models in Table 2. We, however, see some evidence that the parallel pre-trend assumption is violated when estimating the effects of SIG on graduation rates (see Appendix Table A2). The year prior to SIG funds showed a significant increase in graduation rates. However, even after controlling for pre-trends, the results show positive trends in graduation rates both during and after the reform period.

We also match on both the pre-reform average level of performance and yearly trends and re-estimated SIG effects with the matched sample. As shown in Appendix Table A3, the estimates of SIG effects on grade three through eight achievement continue to be positive in both math and reading/ELA over time. The estimates are somewhat smaller than corresponding estimates from our benchmark models, but the differences are neither substantial in magnitude nor statistically distinguishable from zero. The estimates for graduation rates are included Appendix Table A4. In comparison to the results in Table 3, the estimated effects from the matched sample analysis are about 3 or 4 percentage points lower than the whole sample estimates in each of the years, although the patterns are the same in that the effects on graduate rates gradually increased over time and raised up to 11 percentage points three years after SIG interventions. Many estimates are statistically nonsignificant, at least in part because the matched sample has a much smaller sample size, which reduces the precision of estimates. Overall, these results concur with overall positive effects of SIG programs on school performance.

Among the 90 SIG schools in NC, WA, and BCPS, only 15 of them had completely exited under-performing school designations after their three-year SIG interventions, while the majority of these SIG schools were continuously identified as either Priority or Focus schools. To examine the degree to which the sustained post-reform SIG effects can be driven by post-SIG designations as Priority or Focus schools, we apply Equation (1) to a sample of schools including only the ones (both SIG and non-SIG) that had either Priority or Focus designations in post-reform years. If the post-reform effects were primarily driven by continuous supports through Priority and Focus designations, then we would have observed close-to-zero post-reform effects in this estimation—namely, the SIG and non-SIG schools should have received the same treatment and thus should have improved similarly. As shown in Appendix Table A5 and A6, the estimated post-SIG effects are very consistent with those in Table 2 and 3, pointing to the effects being driven by the SIG intervention and not the Priority or Focus designation.

Lastly, we examined whether the identified effects could be driven by the changes in entering cohorts of students. We use an approach similar to Equation (1) with entering cohorts' achievement prior to joining the school as the left-hand side variables. This approach estimates the extent to which SIG schools became more or less likely to attract academically prepared students. As Appendix Table A7 shows, although some coefficients are significantly either positive or negative, most of the effects are indistinguishable from zero, and no common pattern emerges. These results lend support to the hypothesis that the identified SIG effects are not driven by changes in the characteristics of entering cohorts.

Heterogeneity in SIG Effects: For 3-8 grade achievement, the patterns we described from the pooled analysis are generally consistent across locations. As shown in Table 4, in each location and for both subjects, we observe gradual improvements during the SIG reform years. All the effects are statistically significant by the third year of the reform except for the SFUSD and BCPS samples in reading/ELA. The most salient across-location difference is that SIG schools in WA show the most sustainable effects in math after the SIG grants ended.

[Table 4 Here]

Table 5 shows SIG's longitudinal effects on graduation rate across locations. WA had the largest gain in graduation rates attributable to SIG programs, ranging from 16.6 percentage points in the first year of the reform to 29.9 percentage points in the third-year post-reform. One caution of interpreting WA and SFUSD results is that WA only had five SIG high schools and SFUSD only had two SIG high schools, so the estimates are imprecise.

[Table 5 Here]

We further examine SIG effects specifically for historically underserved students of color (non-White and non-Asian students) and for low-SES students. We analyze outcomes for these subsets of students only in schools who have at least 20 of these students. As shown in Table 6, for math and reading/ELA test scores in grades 3-8, the SIG effects for historically unserved students of color are on par with the overall SIG effects in Table 2. We again see a gradually increasing trend during the reform years with the largest effect in the third year of reform year three, and most of the effects sustained in post SIG reform years. The contrasts between turnaround and transformation and between the two cohorts again are similar to the SIG effects for all students. In the analysis of the graduation rate in high schools, the SIG effects, although positive and significant, are slightly smaller than for the overall sample. The effects range from about 5.2 percentage points in reform year two to about 9.5 percentage points in post-reform year three. Our sample size does not support the estimation of SIG effects by turnaround and transformation models separately.

[Table 6 Here]

We then analyze the effects separately for students from low-SES families. Given the neighborhood linkage of our SES measure, the majority of SES-disadvantaged students are present in a smaller selection of schools. As a result, we see a drop in the sample size used for estimating SIG effects on this subpopulation of students. As shown in the column of “All SIG” in Table 7, SIG effects for low-SES students are largely similar to the SIG effects for all students in the column of “All SIG” in Table 2. When contrasting the effects across reform types, we see that low-SES students in turnaround schools improved more than peers in transformation schools in math particularly. Moreover, within turnaround schools, SIG effects appear larger for low-SES students in math in grades 3-8 than the estimates for all students in Table 2 column “Math, Turn.” Then within Cohort 2 schools, the effects for low-SES students are substantially larger than the effects for all students, particularly during the reform years and the early years after the funds ended. Our data do not allow us to estimate separate SIG effects for SES disadvantaged students in graduation rates because the state public available data do not separately report graduation rates for SES disadvantaged students.

[Table 7 Here]

Discussion

The School Improvement Grant program provides an unprecedent opportunity for studying whether and how governments' substantial investment in building capacity of public organizations can pay off in both the short- and long-run. This study is the first, to our knowledge, to estimate the longer-term effects of SIG on student achievement and graduation rates across multiple locations.

The findings for the SIG program are relevant as states and districts aim to improve their lowest performing schools. Although the SIG program ended after the reauthorization of Elementary and Secondary Education Act (ESEA) in 2015, known as the Every Student Succeeds Act (ESSA), states under ESSA are required to use evidence-based practices to transform underperforming schools. Many states plan to continue similar capacity-building approaches with some variations in program designs (Sun et al., 2019). Given that the aim of SIG programs is to spur dramatic change and build organizational capacity, understanding whether SIG effects increase, sustain or decline over the long run, particularly after the SIG funds were removed, is central to understanding the potential advantages and disadvantages of this type of capacity-building policy. Moreover, given that prior studies show variations of program effects across locations, by examining multiple locations - two urban districts and two states across regions of the country – this study can assess both average effects and variation in effects.

The results provide some reasons for optimism regarding the efficacy of the SIG program. We find positive short-term effects of SIG on test score and graduation rates, that increase with more years of implementing the interventions. This gradual emergence of SIG effects during reform years echo earlier findings that capacity-building programs take times to yield impacts (Sun et al., 2017; Borman et al., 2003; Bryk et al., 2010). Schools need time to adopt new curricula, hire and train staff, and develop new organizational climate and culture. In addition, we find that while SIG effects diminish over time after the funding of the program ends, the positive effects, particularly in math and particularly in turnaround schools, are sustained at least three years after the funding ends. Moreover, the continuous, substantial improvement in high school graduation rates provide the first evidence that some SIG schools were able to generate lasting improvements in the performance of awarded high schools.

Our estimates suggest that SIG programs may be more successful than many other government-driven programs that aim to build organizational capacity to remedy underperformance in public schools. For example, another significant government initiative is the Comprehensive School Reform (CSR) originally funded in 1998 with $145 million. Later, this program became part of No Child Left Behind Act in 2001, and congress budgeted an annual funding between $200-$310 million per year up until 2015. Nearly 7,000 schools nationwide received three-year awards to implement CSR models between 1998 and 2006 (U.S. Department of Education, 2010). Evidence indicates that five years after initially receiving their CSR awards, schools receiving awards did not demonstrate larger achievement growth in either math or reading than matched comparison schools not receiving CSR grants (U.S. Department of Education, 2010). These CSR models include several features similar to SIG programs, such as evidence-based reforms, comprehensive design, professional development for school staff, measurable goals, parent and community involvement, and external assistance. The null effects of CSR are largely attributable to the low-level implementation of designed program components. Moreover, compared to SIG, CSR schools received lower funds per year, and were not required to undertake staff changes (e.g., replacing principals and 50% of staff members) or tie personnel decisions with student performance growth.

Similarly, under the No Child Left Behind Act of 2001, states were required to improve under-performing schools with both sanctions and capacity building approaches through additional resources and technical assistance (Dee & Jacob, 2011; Hanushek & Raymond, 2005; Strunk, McEachin, & Westover, 2014). Strunk, McEachin, & Westover (2014) assessed the effects of District Assistance and Intervention Teams who were state-approved external experts who provided technical assistance to some under-performing school districts in California. They found that students in districts with this support performed significantly better on state standardized tests in math, but not in ELA. The effect size of 0.005 to 0.045 SDs on math in the first two years of implementation are considerably smaller than the estimates in this study (0.1-0.16 SD). Although these external experts provided an array of supports to districts and schools, these treatment schools and districts did not receive as substantial an influx of resources as those SIG schools and did not have prescribed school turnaround models.

The identified SIG effects on test scores in this study are similar to the effects on student test scores estimated for the market-based reforms in New Orleans after the Hurricane Katrina in 2005 (Harris & Larsen, 2018). In the New Orleans reforms, the state took over almost all public schools, which, in turn, turned over management to autonomous non-profit charter management organizations working under performance contracts. Harris and Larsen’s study showed average effects on test scores after 4.5 years of market-based reform ranging 0.10 to 0.40 SD, which are similar to the estimated effects of 0.14 to 0.3 SD that we estimate for SIG turnaround schools. Our identified SIG effects on high school graduation rates, particularly from the matched sample, are also similar to estimates for the market-based reforms in New Orleans.

Our study also corroborates prior research showing variation SIG effects, perhaps as a result of variation in design and implementation. For example, we estimate larger effects for Washington state, suggesting the benefits of further investigation of reform strategies employed by this state and its SIG schools. Sun et al. (2019) applied text analysis techniques to WA schools' improvement planning and implementation reports and found several promising reform strategies that were associated with either the reduction in student absences or gains in state standardized test scores during reform years. Those reform strategies include teachers' use of data to inform instruction and develop targeted interventions for at-risk students, and setting improvement targets for both students and teachers and providing incentives for meeting these targets. Other prior studies also indicate that hiring highly effective teachers and school leaders partially explained positive effects in these schools, while several factors can suppress positive effects, including but not limited to hiring more novice educators (Henry et al., 2019). Knowledge on when school turnaround programs such as SIG were more successful can provide guidance for states and districts as they develop their own evidence-based school turnaround strategies under ESSA.

While this study provides initial evidence of the longer-run effects of SIG it has shortcomings. First, we are able to examine only a few years post SIG implementation. More years of data would eventually allow researchers to better understand the longer-term impacts of SIG on school effectiveness. Second, this study examines SIG program effects on schools. It does not look at the effects of attending a SIG school on the longer-run effects for students. Following cohorts of students who attended SIG schools to examine SIG programs' long-term effects on outcomes such as future educational attainments and employment outcomes – such as

Sass, Zimmer, Gill, and Booker’s (2016) examination of charter high schools’ effects and Gormley Jr., Phillips, & Anderson’s (2018) study Tulsa’s Pre-K program’s effects on middle school student performance – would provide further evidence on the benefits of the SIG approach.

In sum, states around the country are searching for ways to transform under-performing schools to build their long-term capacity for better instruction. Because these schools often educate large proportions of students from traditionally underserved groups, improving chronically underperforming schools serves as a critical lever for reducing educational inequality. Yet, these types of programs require substantial investment and, often, substantial upheaval in schools. Such dramatic transformation makes it even more critical to learn what reform practices work and how they work. The findings can inform school turnarounds but may also provide lessons for organizational capacity building in other public sectors.

References

Angrist, J. D., & Pischke, J. S. (2008). Mostly harmless econometrics: An empiricist's companion. Princeton university press.

Borman, G. D., Hewes, G. M., Overman, L. T., & Brown, S. (2003). Comprehensive school reform and achievement: A meta-analysis. Review of educational research, 73(2), 125-230.

Bryk, A. S., Sebring, P. B., Allensworth, E., Easton, J. Q., & Luppescu, S. (2010). Organizing schools for improvement: Lessons from Chicago. University of Chicago Press.

Cameron, A. C., & Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. Journal of Human Resources. 50(2). 317-372.

Carlson, D., & Lavertu S. (2018). School Improvement Grants in Ohio: Effects on student achievement and school administration. Educational Evaluation and Policy Analysis, 40(3), 287-315.

Colorado Department of Education (2015). Evaluation of SIG Schools (Cohorts 1 and 2). Unit of Federal Programs. Available May 2017. Online:http://www.cde.state.co.us/fedprograms/briefreport-for-cohort-1-and-2-sig-schools

DeCorby-Watson, K., Mensah, G., Bergeron, K., Abdi, S., Rempel, B., & Manson, H. (2018). Effectiveness of capacity building interventions relevant to public health practice: a systematic review. BMC public health, 18(1), 684.

Dee, T. (2012) School Turnarounds: Evidence from the 2009 Stimulus. NBER Working Paper No. 17990, April 2012.

Dee, T. S., & Jacob, B. (2011). The impact of No Child Left Behind on student achievement. Journal of Policy Analysis and management, 30(3), 418-446.

Dickey-Griffith, D. (2013). Preliminary effects of the school improvement grant program on student achievement in Texas. The Georgetown Public Policy Review, 21-39.

Dragoset, L., Thomas, J., Herrmann, M., Deke, J., James-Burdumy, S., Graczewski, C., Boyle, A., Upton, R., Tanenbaum, C., & Giffin, J. (2017). School Improvement Grants: implementation and effectiveness (NCEE 2017-4013). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Ginsburg, A., & Smith, M. S. (2018). Revisiting SIG: Why critics were wrong to write off the federal School Improvement Grant program. FutureEd. Georgetown University.

Gormley Jr, W. T., Phillips, D., & Anderson, S. (2018). The effects of Tulsa's Pre-K Program on middle school student performance. Journal of Policy Analysis and Management, 37(1), 63-87.

Guthrie, J. E., & Henry, G. T. (2016). When the LATE ain't ATE: Comparing alternative methods for evaluating reform impacts in low-achieving schools (Working Paper). Annual meeting of the Association for Public Policy Analysis and Management, Washington DC.

Hanushek, E. A., & Raymond, M. E. (2005). Does school accountability lead to improved student performance?. Journal of Policy Analysis and Management: The Journal of the Association for Public Policy Analysis and Management, 24(2), 297-327.

Harris, Douglas N. & Larsen, Matthew F. (2018). The effects of the New Orleans Post-Katrina market-based school reforms on student achievement, high school graduation, and college outcomes. Education Research Alliance for New Orleans. Retrieved from https://educationresearchalliancenola.org/files/publications/Harris-Larsen-Effects-of-New-Orleans-Post-Katrina-Market-Based-School-Reforms.pdf

Heissel, Jennifer A. and H. F. Ladd (2016). School turnaround in North Carolina: A regression discontinuity analysis. National Center for the Analysis of Longitudinal Data in Education Research (CALDER). Working Paper 156. Online: http://www.caldercenter.org/sites/default/files/WP%20156.pdf.

Henry, G. T., Pham, L., Kho, A., & Zimmer, R. (2019). Peeking into the black box of school turnaround: A formal test of mediators and suppressors.

Henry, Gary T., J.Guthrie & L. W. Townsend, (2015). Outcomes and impacts of North Carolina's Initiative to turn around the lowest-achieving schools. Consortium for Educational Research and Evaluation–North Carolina. Online: http://cerenc.org/rtttevaluation/district-and-school-transformation/.

Hurlburt, S., Therriault, S. B., & Le Floch, K. C. (2012). School Improvement Grants: Analyses of atate applications and eligible and awarded schools. NCEE 2012-4060. National Center for Education Evaluation and Regional Assistance.

Kyse, Eden N.,R. Swann-Jackson, J. Marini, J. Benton, A. Byrne, A. Sceppaguercio,& K. Wilson.(2014). Final Evaluation Report for the School Improvement Grant (SIG) Evaluation: Summary and Recommendations (2014). Montclair State University. Online: http://www.nj.gov/education/title1/sig/EvaluationReportC1and2.pdf.

Papay, J., & Hannon, M. (2018). The effects of school turnaround strategies in Massachusetts. A paper presented at the annual meeting of the Association for Public Policy Analysis & Management: Washington D.C.

Sass, T. R., Zimmer, R. W., Gill, B. P., & Booker, T. K. (2016). Charter high schools' effects on long-term attainment and earnings. Journal of Policy Analysis and Management, 35(3), 683-706.

Strunk, K. O., McEachin, A., & Westover, T. N. (2014). The use and efficacy of capacity-building assistance for low-performing districts: The case of California's district assistance and intervention teams. Journal of Policy Analysis and Management, 33(3), 719-751.

Sun, M., Liu, J., LeClair, Z., & Zhu, J. (2019). Using a text-as-data approach to understand reform processes: A deep exploration of school improvement strategies. *Educational Evaluation and Policy Analysis,* 41(4), 510-536.

Sun, M., Penner, E., & Loeb, S. (2017). Resource- and approach-driven multi-dimensional change: Three-year effects of School Improvement Grants. *American Educational Research Journal, 54*(4), 607–643.

Taylor, E. S., & Tyler, J. H. (2012). The effect of evaluation on teacher performance. American Economic Review, 102(7), 3628-51.

U.S. Chamber of Commerce Foundation. (2010, March). ARRA School Improvement Grants (SIG): Testing school district commitment. Retrieved September 11, 2015, from http://www.uschamberfoundation.org/content/arra-school-improvement-grants-sig-testing-school-district-commitment

U.S. Department of Education, Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service, Evaluation of the Comprehensive School Reform Program implementation and outcomes: Fifth-Year Report, Washington, D.C., 2010.

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, DC: Office of Elementary and Secondary Education. Retrieved from http://www2.ed.gov/programs/sif/sigguidance11012010.pdf

U.S. Department of Education. (2010). Guidance on school improvement grants under section 1003(g) of the Elementary and Secondary Education Act of 1965. Washington, DC: Office of Elementary and Secondary Education. Retrieved from http://www2.ed.gov/policy/elsec/leg/esea02/pg1.html#sec1003

Wong, V.C., Valentine, J. C., & Miller-Bains, K. (2017). Empirical performance of covariates in education observational studies. Journal of Research on Educational Effectiveness, 10(1). 207-236.

Table 1: Pre-reform School Characteristics for Analytical Sample by Location

	Pooled		SFUSD		BCPS		NC		WA	
	SIG	Non SIG	SIG	Non SIG	SIG	Non SIG	SIG	Non SIG	SIG	Non SIG
% Black	48.3%	18.2%	21.8%	14.8%	73.5%	28.6%	60.9%	28.3%	13.4%	5.1%
% Hispanic	24.1%	16.4%	61.1%	25.3%	24.4%	60.5%	10.2%	10.3%	32.2%	14.7%
% Asian	3.0%	4.6%	8.3%	40.5%	0.2%	1.0%	1.6%	2.2%	5.7%	6.3%
% Other	6.5%	6.0%	6.2%	7.8%	0.7%	1.1%	4.2%	5.5%	15.7%	7.4%
% White	18.2%	54.8%	2.5%	11.6%	1.2%	8.7%	23.1%	53.7%	32.9%	66.4%
% Female	47.1%	48.1%	48.6%	48.5%	48.1%	48.7%	44.9%	48.5%	49.0%	47.7%
% Special Education	7.5%	5.3%	12.8%	11.4%	14.7%	12.5%	0.6%	0.3%	13.5%	17.0%
% English Learner	13.6%	7.9%	44.9%	34.4%	10.0%	16.1%	7.5%	6.2%	15.8%	7.1%
% Gifted	3.4%	7.6%	7.3%	14.3%	3.7%	9.0%	2.6%	9.3%	1.3%	2.1%
% Low SES	59.4%	33.8%	32.9%	28.9%	NA	NA	49.6%	29.9%	64.6%	29.4%
Average Math Score	-0.66	-0.07	-0.73	-0.11	-0.53	-0.05	-0.71	-0.07	-0.68	-0.09
Average Reading Score	-0.69	-0.06	-0.64	-0.13	-0.58	-0.04	-0.80	-0.06	-0.60	-0.06
Graduation Rate (%)	53.2%	72.6%	76.0%	82.3%	52.0%	67.2%	51.1%	76.8%	55.9%	68.6%
Avg. Enrollment	707	570	428	508	1105	811	609	598	568	496
Elementary School	30	2848	5	67	6	196	10	1368	9	1217
Middle School	18	922	1	11	7	73	1	482	9	356
High School	35	958	2	20	10	66	18	385	5	487
Other	16	566	1	0	2	86	11	262	2	218
Reform Model: Transformation	70		5		11		33		21	
Reform Model: Turnaround	29		4		14		7		4	
Cohort 1	66		9		19		23		15	
Cohort 2	33		0		6		17		10	
Post-reform: Priority	70		NA		13		37		20	
Post-reform: Priority, then Focus	14		NA		9		0		5	
N	99	5,294	9	98	25	421	40	2,497	25	2,278

Note: The mean statistics presented above are unweighted averages of school-level characteristics for the three year pre-reform period (2007-08 to 2009-10). WA only had one year of pre-reform test score data. Graduation rate is the four-year graduation rate. School level (ES, MS, HS, Other) is defined using the NCES definition. SIG schools that were part of later cohorts of SIG were removed from the sample and only schools present during the first year of SIG (2010-11) were included. SFUSD = San Francisco Unified School

District. BCPS = Beachfront County Public Schools. NC = North Carolina state public schools. WA = Washington state public schools.

Table 2: Estimated Longitudinal Effects of SIG on Student Achievement 3-8 Grades (Elementary and Middle Schools)

		Math					Reading/ELA				
		All SIG	Trans	Turn	C1	C2	All SIG	Trans	Turn	C1	C2
Reform Years	1st Year	0.115***	0.108*	0.138**	0.091*	0.158***	0.044	0.050	0.041	0.018	0.087*
		(0.033)	(0.045)	(0.044)	(0.046)	(0.042)	(0.030)	(0.041)	(0.042)	(0.042)	(0.038)
	2nd Year	0.171***	0.161***	0.192**	0.134**	0.242***	0.073*	0.065	0.086	0.036	0.137***
		(0.035)	(0.042)	(0.059)	(0.046)	(0.048)	(0.032)	(0.040)	(0.051)	(0.043)	(0.040)
	3rd Year	0.228***	0.190***	0.300***	0.224***	0.231***	0.122***	0.117**	0.131**	0.086*	0.185***
		(0.037)	(0.044)	(0.062)	(0.051)	(0.050)	(0.031)	(0.039)	(0.049)	(0.038)	(0.049)
Post-Reform Years	1st Post Year	0.180***	0.117	0.313***	0.171*	0.193***	0.114***	0.117**	0.105	0.118*	0.101*
		(0.048)	(0.060)	(0.066)	(0.069)	(0.052)	(0.034)	(0.039)	(0.064)	(0.046)	(0.049)
	2nd Post Year	0.178***	0.153*	0.227***	0.177**	0.173*	0.108**	0.090*	0.145*	0.094*	0.126*
		(0.047)	(0.065)	(0.053)	(0.060)	(0.075)	(0.035)	(0.041)	(0.062)	(0.045)	(0.056)
	3+ Post Year	0.116**	0.088	0.175**	0.116*	0.091	0.106*	0.086	0.140	0.088	0.140**
		(0.045)	(0.059)	(0.062)	(0.055)	(0.080)	(0.043)	(0.048)	(0.083)	(0.055)	(0.050)
N		35,200	35,200	35,200	35,200	35,200	35,048	35,048	35,048	35,048	35,048

Notes: The table shows the estimated longitudinal SIG effects on Grade 3-8 (HS excluded) Math and Reading/ELA performance (SD) in the pooled sample. Scores have been standardized by location, grade, year, and test. 3+ Post Year is the estimated effect of SIG in the third year and beyond (when data are available). Trans = Transformation model adopted, Turn = Turnaround model adopted, C1 = SIG Cohort 1, C2 = SIG Cohort 2. Robust standard errors clustered at the school level are presented in parentheses below estimates. * p<0.05, ** p<0.01, * p<0.001

Table 3: Estimated Longitudinal Effects of SIG on 4-year High School Graduation Rate

		All SIG	Trans	Turn	C1	C2
Reform Years	1st Year	6.232***	6.087***	6.894**	6.162***	6.785*
		(1.535)	(1.760)	(2.525)	(1.732)	(3.213)
	2nd Year	8.789***	8.886***	8.410*	10.605***	4.961
		(1.831)	(2.099)	(3.292)	(1.989)	(3.551)
	3rd Year	10.759***	10.826***	10.506***	13.082***	5.063
		(1.979)	(2.319)	(2.559)	(1.685)	(4.905)
Post-Reform Years	1st Post Year	12.397***	11.972***	13.979***	14.911***	6.279
		(1.937)	(2.302)	(2.243)	(1.786)	(4.131)
	2nd Post Year	13.157***	13.224***	12.911***	15.296***	8.028
		(2.052)	(2.466)	(1.845)	(1.881)	(4.839)
	3+ Post Year	14.171***	14.614***	12.755***	15.450***	11.002**
		(1.838)	(2.184)	(1.897)	(1.900)	(4.165)
N		7,984	7,984	7,984	7,984	7,984

Notes: The table shows the estimated longitudinal SIG effects on high school 4-year graduation rate in the pooled sample. 3+ Post Year is the estimated effect of SIG in the third year and beyond (when data are available). Trans = Transformation model adopted, Turn = Turnaround model adopted, C1 = SIG Cohort 1, C2 = SIG Cohort 2. Robust standard errors clustered at the school level are presented in parentheses below estimates.
* $p<0.05$, ** $p<0.01$, *** $p<0.001$

Table 4: Estimated Longitudinal Effects of SIG on Student Achievement by Location

		Math					Reading/ELA				
		Pooled	SFUSD	BCPS	NC	WA	Pooled	SFUSD	BCPS	NC	WA
Reform Years	1st Year	0.115***	0.140*	0.091*	0.136	0.171***	0.044	0.110*	-0.015	0.081	0.071*
		(0.033)	(0.055)	(0.044)	(0.076)	(0.049)	(0.030)	(0.055)	(0.035)	(0.069)	(0.034)
	2nd Year	0.171***	0.279***	0.152**	0.130*	0.248**	0.073*	0.129	-0.007	0.083	0.140**
		(0.035)	(0.080)	(0.047)	(0.062)	(0.077)	(0.032)	(0.089)	(0.030)	(0.067)	(0.045)
	3rd Year	0.228***	0.360**	0.162***	0.219**	0.307***	0.122***	0.157	0.050	0.135*	0.199***
		(0.037)	(0.108)	(0.048)	(0.072)	(0.073)	(0.031)	(0.086)	(0.032)	(0.057)	(0.058)
Post-Reform Years	1st Post	0.180***	n/a	0.235**	0.077	0.304***	0.114***	n/a	0.023	0.140*	0.190***
	Year	(0.048)	n/a	(0.079)	(0.088)	(0.080)	(0.034)	n/a	(0.051)	(0.065)	(0.047)
	2nd Post	0.178***	0.158	0.163**	0.122	0.318***	0.108**	0.172*	0.056	0.109	0.157*
	Year	(0.047)	(0.081)	(0.061)	(0.096)	(0.090)	(0.035)	(0.079)	(0.059)	(0.063)	(0.064)
	3+ Post	0.116**	0.137	0.060	0.066	0.278**	0.106*	0.130	0.118*	0.094	0.144
	Year	(0.045)	(0.079)	(0.068)	(0.085)	(0.088)	(0.043)	(0.071)	(0.052)	(0.091)	(0.078)
N		35,200	759	3,400	18,319	12,722	35,048	759	3,404	18,296	12,589

Notes: The table shows the estimated longitudinal SIG effects on Grade 3-8 (HS excluded) Math and Reading/ELA performance (SD). Scores have been standardized by location, grade, year, and test. 3+ Post Year is the estimated effect of SIG in the third year and beyond (when data are available). SFUSD = San Francisco Unified School District. BCPS = Beachfront County Public Schools. NC = North Carolina public schools. WA = Washington state public schools. Robust standard errors clustered at the school level are presented in parentheses below estimates. n/a indicates that CA did not have state standardized tests in these years. * p<0.05, ** p<0.01, * p<0.001

Table 5: Estimated Longitudinal Effects of SIG on 4-year High School Graduation Rate by Location

		The Whole Sample				
		Pooled	SFUSD	BCPS	NC	WA
Reform Years	1st Year	6.232***	-2.712	4.020	5.937**	16.570***
		(1.535)	(2.506)	(2.479)	(2.006)	(1.525)
	2nd Year	8.789***	-3.527	12.232***	6.471**	17.770***
		(1.831)	(3.585)	(3.255)	(2.169)	(4.533)
	3rd Year	10.759***	3.283	13.276***	8.825**	17.644**
		(1.979)	(4.512)	(2.303)	(2.783)	(5.660)
Post-Reform Years	1st Post Year	12.397***	1.687	13.390***	10.442***	21.983***
		(1.937)	(2.441)	(2.650)	(2.950)	(3.655)
	2nd Post Year	13.157***	1.317	12.614***	11.519***	25.780***
		(2.052)	(2.641)	(2.286)	(2.994)	(4.730)
	3+ Post Year	14.171***	10.321	11.780***	12.241***	29.900***
		(1.838)	(6.303)	(2.316)	(2.539)	(4.247)
N		7,984	177	600	3,539	3,668

Notes: The table shows the estimated longitudinal SIG effects on high school 4-year graduation rate. 3+ Post Year is the estimated effect of SIG in the third year and beyond (when data are available). SFUSD = San Francisco Unified School District. BCPS = Beachfront County Public Schools. NC = North Carolina public schools. WA = Washington state public schools. Robust standard errors clustered at the school level are presented in parentheses below estimates.
* p<0.05, ** p<0.01, *** p<0.001

Table 6: Estimated Longitudinal Effects of SIG for Historically Underserved Students of Color

		Math					Reading/ELA					4-year Graduation Rate		
		All SIG	Trans	Turn	C1	C2	All SIG	Trans	Turn	C1	C2	All SIG	C1	C2
Reform Years	1st Year	0.066	0.052	0.109*	0.042	0.148***	-0.001	-0.000	0.015	-0.023	0.063	3.522	5.351*	3.618
		(0.035)	(0.045)	(0.046)	(0.058)	(0.043)	(0.031)	(0.037)	(0.046)	(0.049)	(0.037)	(1.893)	(2.673)	(2.774)
	2nd Year	0.124***	0.112**	0.154**	0.089	0.226***	0.042	0.033	0.061	0.017	0.111*	5.161*	8.066***	3.819
		(0.035)	(0.040)	(0.059)	(0.050)	(0.057)	(0.034)	(0.040)	(0.057)	(0.052)	(0.044)	(2.306)	(2.368)	(4.019)
	3rd Year	0.183***	0.154***	0.246***	0.192***	0.212***	0.085*	0.071	0.117*	0.053	0.165**	6.610*	11.457***	2.220
		(0.036)	(0.039)	(0.065)	(0.058)	(0.050)	(0.035)	(0.042)	(0.057)	(0.049)	(0.052)	(2.740)	(1.803)	(5.250)
Post-Reform Years	1st Post	0.173***	0.118*	0.293***	0.184**	0.199**	0.048	0.048	0.049	0.047	0.078	7.245*	12.650***	1.900
	Year	(0.043)	(0.052)	(0.064)	(0.068)	(0.063)	(0.035)	(0.039)	(0.065)	(0.053)	(0.046)	(2.966)	(2.627)	(4.895)
	2nd Post	0.116**	0.088	0.185***	0.129*	0.141	0.026	0.013	0.058	0.014	0.077	8.559**	12.903***	4.671
	Year	(0.040)	(0.049)	(0.049)	(0.061)	(0.076)	(0.033)	(0.039)	(0.053)	(0.052)	(0.054)	(2.877)	(2.074)	(5.408)
	3+ Post	0.055	0.021	0.139*	0.092	0.055	0.042	0.022	0.079	0.046	0.107*	9.542**	12.990***	4.347
	Year	(0.045)	(0.053)	(0.065)	(0.065)	(0.079)	(0.049)	(0.044)	(0.099)	(0.065)	(0.050)	(2.920)	(2.877)	(4.737)
N		32,424	32,424	32,424	32,424	32,424	32,281	32,281	32,281	32,281	32,281	4,645	4,645	4,645

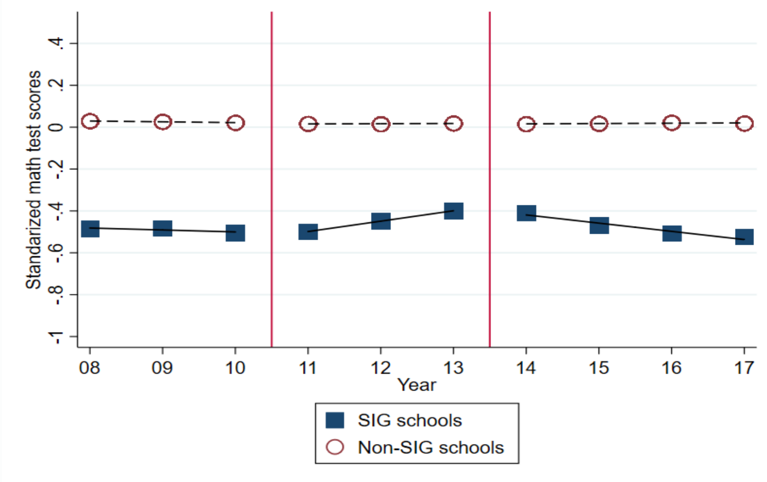
Notes: The table shows the estimated longitudinal SIG effects on Grade 3-8 (HS excluded) Math and Reading/ELA performance (SD) and high school 4-year graduation rate in the pooled sample. SFUSD and BCPS have been omitted from the analysis on graduation rates since consistent data was unavailable for this sample of students. Scores have been standardized by location, grade, year, and test (using the full sample of students). 3+ Post Year is the estimated effect of SIG in the third year and beyond (when data are available). Historically Underserved Students of Color = non-White, non-Asian. If a school had fewer than 20 Historically Underserved Students of Color, they are omitted from the analysis. No high schools who chose the turnaround model in this sample had a 4-year graduation rate making it impossible to estimate the effects of both reform models on graduation rates. Trans = Transformation model adopted, Turn = Turnaround model adopted, C1 = SIG Cohort 1, C2 = SIG Cohort 2. Robust standard errors clustered at the school level are presented in parentheses below estimates. * $p<0.05$, ** $p<0.01$, *** $p<0.001$

Table 7: Estimated Longitudinal Effects of SIG for Low-SES Students

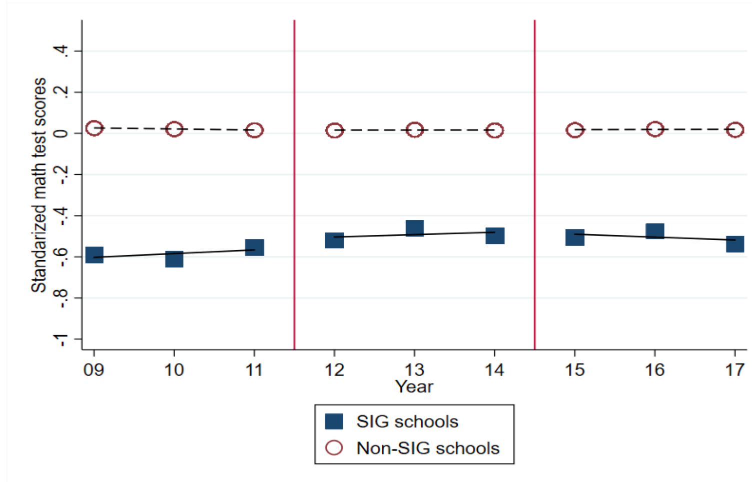
		Math					Reading/ELA				
		All SIG	Trans	Turn	C1	C2	All SIG	Trans	Turn	C1	C2
Reform Years	1st Year	0.196**	0.156*	0.333***	0.179	0.233***	0.109**	0.099*	0.164**	0.076	0.156**
		(0.063)	(0.066)	(0.083)	(0.101)	(0.049)	(0.040)	(0.044)	(0.054)	(0.057)	(0.056)
	2nd Year	0.203**	0.157**	0.348**	0.189	0.234**	0.112**	0.083	0.201*	0.055	0.207***
		(0.065)	(0.057)	(0.133)	(0.099)	(0.073)	(0.043)	(0.043)	(0.086)	(0.058)	(0.051)
	3rd Year	0.248***	0.180**	0.467***	0.275**	0.205**	0.222**	0.219**	0.227*	0.199	0.248***
		(0.068)	(0.058)	(0.131)	(0.106)	(0.067)	(0.069)	(0.079)	(0.092)	(0.105)	(0.070)
Post-Reform Years	1st Post Year	0.182*	0.100	0.444***	0.168	0.209***	0.178***	0.142**	0.291***	0.162*	0.194***
		(0.075)	(0.073)	(0.118)	(0.119)	(0.062)	(0.049)	(0.054)	(0.070)	(0.076)	(0.046)
	2nd Post Year	0.255**	0.210*	0.395***	0.305*	0.162	0.165***	0.137**	0.250***	0.138*	0.200***
		(0.091)	(0.103)	(0.092)	(0.134)	(0.095)	(0.041)	(0.043)	(0.064)	(0.058)	(0.059)
	3rd Post Year	0.154*	0.056	0.372***	0.159	n/a	0.109	0.075	0.148	0.086	n/a
		(0.078)	(0.075)	(0.100)	(0.104)	n/a	(0.075)	(0.064)	(0.150)	(0.084)	n/a
N		13,194	13,194	13,194	13,194	13,194	13,125	13,125	13,125	13,125	13,125

Notes: The table shows the estimated longitudinal SIG effects on Grade 3-8 (HS excluded) Math and Reading/ELA performance (SD) in the pooled sample. We only have geocoded address information available through 2015-16. Scores have been standardized by location, grade, year, and test (using the full sample of students). We use student neighborhood characteristics (median household income, median house value, % of 25+ population with a BA) to identify SES-disadvantaged students. If a school had fewer than 20 SES-disadvantaged students, they are omitted from the analysis. Given that we constructed the measure of SES-disadvantage ourselves, we were not able to match state-reported graduation rates to this population of students.
Trans = Transformation model adopted, Turn = Turnaround model adopted, C1 = SIG Cohort 1, C2 = SIG Cohort 2. Robust standard errors clustered at the school level are presented in parentheses below estimates. * $p<0.05$, ** $p<0.01$, *** $p<0.001$

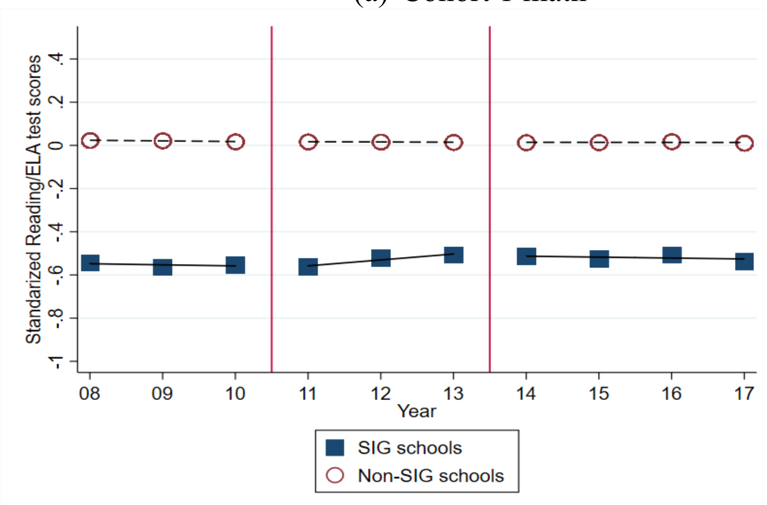
Figure 1: Trends in 3-8 Grade Student Achievement for both SIG and non-SIG Schools



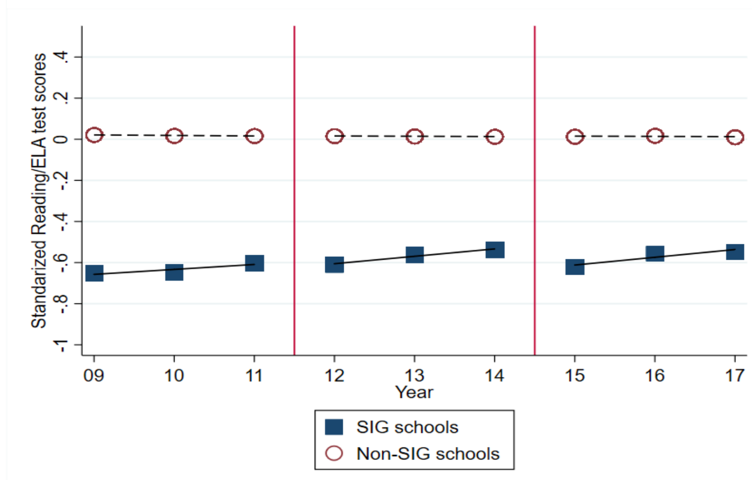
(a) Cohort 1 math



(b) Cohort 2



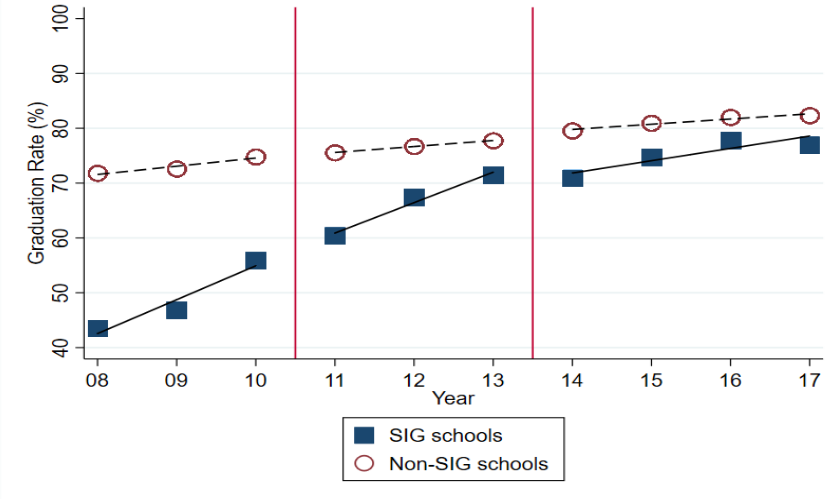
(a) Cohort 1 Reading/ELA



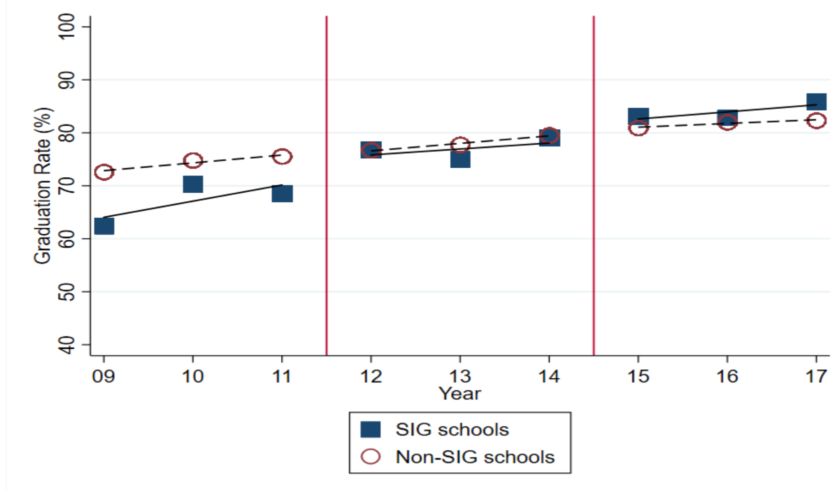
(b) Cohort 2 Reading/ELA

Notes: Plots are of weighted average reading/ELA achievement across our four locations for the 11 year period that our sample covers (weighted by student enrollment size). Reading/ELA scores are standardized by location, grade, and year. Not all locations have data for each year presented in the plots. SFUSD has data for 2008 through 2017 (but is missing data for 2014 since no standardized test was given that year). BCPS has data from 2008 through 2016. NC has data from 2008 through 2017. WA has data from 2010 through 2018.

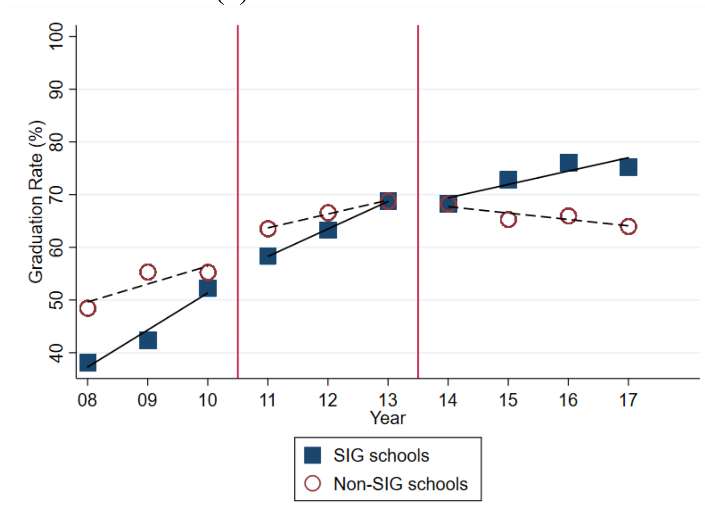
Figure 2: Trends in 4-year Graduation Rates for both SIG and non-SIG Schools



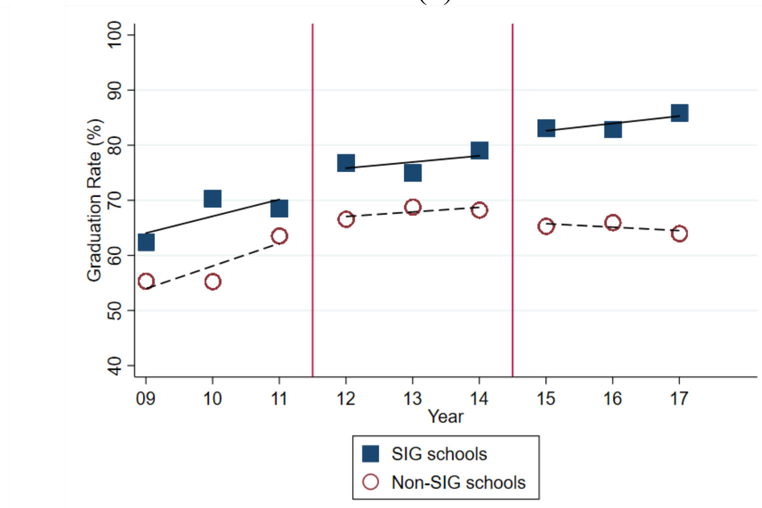
(a) Cohort 1



(b) Cohort 2



(c) Matched Sample Cohort 1



(d) Matched Sample Cohort 2

Notes: Plots are of average 4-year graduation rate across our four locations for the 10 year period that our sample covers. All locations had data across the 10 year period except for BCPS which went only from 2008 to 2016.

Appendix Tables

Table A1: Estimated Longitudinal Effects of SIG on Student Achievement (Controlling for Differential Pre-trends)

		Math					Reading/ELA				
		All SIG	Trans	Turn	C1	C2	All SIG	Trans	Turn	C1	C2
Pre-Reform Years	2nd Pre Year	-0.058*	-0.060	-0.061	-0.038	-0.086	-0.052	-0.038	-0.071	-0.038	-0.076*
		(0.028)	(0.043)	(0.031)	(0.031)	(0.051)	(0.028)	(0.040)	(0.037)	(0.039)	(0.037)
	3rd Pre Year	0.005	0.010	-0.012	0.012	-0.010	-0.005	0.019	-0.041	-0.032	0.025
		(0.034)	(0.048)	(0.044)	(0.042)	(0.059)	(0.030)	(0.044)	(0.032)	(0.039)	(0.044)
Reform Years	1st Year	0.100*	0.095	0.116*	0.086	0.128*	0.027	0.047	0.006	-0.006	0.071
		(0.040)	(0.056)	(0.056)	(0.056)	(0.060)	(0.037)	(0.054)	(0.038)	(0.051)	(0.051)
	2nd Year	0.156***	0.148**	0.170**	0.128*	0.212***	0.056	0.061	0.051	0.012	0.121**
		(0.037)	(0.045)	(0.064)	(0.055)	(0.041)	(0.036)	(0.049)	(0.050)	(0.051)	(0.042)
	3rd Year	0.213***	0.176***	0.277***	0.218***	0.201***	0.105**	0.113**	0.096*	0.061	0.168***
		(0.038)	(0.042)	(0.069)	(0.054)	(0.049)	(0.032)	(0.044)	(0.047)	(0.043)	(0.049)
Post-Reform Years	1st Post Year	0.165***	0.103*	0.291***	0.166*	0.163***	0.097**	0.113**	0.069	0.094	0.085
		(0.045)	(0.052)	(0.074)	(0.069)	(0.044)	(0.036)	(0.044)	(0.061)	(0.051)	(0.046)
	2nd Post Year	0.163***	0.140*	0.205***	0.171**	0.143*	0.091*	0.086	0.109*	0.069	0.110*
		(0.044)	(0.058)	(0.060)	(0.062)	(0.057)	(0.036)	(0.046)	(0.056)	(0.050)	(0.051)
	3+ Post Year	0.100*	0.074	0.153*	0.110	0.060	0.088*	0.083	0.104	0.063	0.124**
		(0.044)	(0.055)	(0.070)	(0.058)	(0.072)	(0.042)	(0.053)	(0.072)	(0.057)	(0.045)
N		35,200	35,200	35,200	35,200	35,200	35,048	35,048	35,048	35,048	35,048

Notes: The table shows the estimated longitudinal SIG effects on Grade 3-8 (HS excluded) Math and Reading/ELA performance (SD) in the pooled sample. Test scores have been standardized by location, grade, year, and test. 2nd Pre Year and 3rd Pre Year are represent the second and third year of the three year pre-reform period, respectively. 3+ Post Year is the estimated effect of SIG in the third year and beyond (when data are available). Trans = Transformation model adopted, Turn = Turnaround model adopted, C1 = SIG Cohort 1, C2 = SIG Cohort 2. Robust standard errors clustered at the school level are presented in parentheses below estimates. * p<0.05, ** p<0.01, *** p<0.001

Table A2: Estimated Longitudinal Effects of SIG on 4-year Graduation Rate (Controlling for Differential Pre-trends)

		Graduation Rate				
		SIG	Trans	Turn	C1	C2
Pre-Reform Years	2nd Pre Year	-0.265	-0.001	-1.177	0.653	-0.010
		(1.433)	(1.309)	(3.960)	(1.825)	(2.008)
	3rd Pre Year	5.633**	4.990*	8.315	6.907**	4.465
		(2.032)	(2.013)	(4.948)	(2.403)	(3.376)
Reform Years	1st Year	8.002***	7.712**	9.336	8.746**	8.066
		(2.337)	(2.536)	(4.766)	(2.705)	(4.396)
	2nd Year	10.510***	10.458***	10.843**	13.166***	6.179
		(2.402)	(2.718)	(3.830)	(2.652)	(4.238)
	3rd Year	12.509***	12.428***	12.946***	15.646***	6.349
		(2.508)	(2.878)	(3.185)	(2.287)	(5.450)
Post-Reform Years	1st Post Year	14.146***	13.573***	16.423***	17.477***	7.562
		(2.557)	(2.917)	(3.653)	(2.504)	(4.861)
	2nd Post Year	14.880***	14.797***	15.353***	17.863***	9.223
		(2.613)	(2.998)	(3.617)	(2.599)	(5.242)
	3+ Post Year	15.915***	16.209***	15.200***	18.018***	12.196**
		(2.369)	(2.671)	(3.630)	(2.582)	(4.443)
N		7,984	7,984	7,984	7,984	7,984

Notes: The table shows the estimated longitudinal SIG effects on 4-year graduation rate in the pooled sample. 2nd Pre Year and 3rd Pre Year are represent the second and third year of the three year pre-reform period, respectively. 3+ Post Year is the estimated effect of SIG in the third year and beyond (when data are available). Trans = Transformation model adopted, Turn = Turnaround model adopted, C1 = SIG Cohort 1, C2 = SIG Cohort 2. Robust standard errors clustered at the school level are presented in parentheses below estimates. * p<0.05, ** p<0.01, *** p<0.001

Table A3. Estimated Longitudinal Effects of SIG on 3-8 Grade Student Achievement on Matched Samples

		Math					Reading/ELA				
		SIG	Trans	Turn	C1	C2	SIG	Trans	Turn	C1	C2
Reform Years	1st Year	0.094*	0.089	0.118*	0.070	0.130*	0.047	0.051	0.050	-0.000	0.109*
		(0.039)	(0.049)	(0.052)	(0.053)	(0.058)	(0.033)	(0.044)	(0.045)	(0.051)	(0.043)
	2nd Year	0.118*	0.113*	0.135	0.095	0.167**	0.076*	0.065	0.096	0.040	0.143**
		(0.046)	(0.052)	(0.071)	(0.060)	(0.059)	(0.038)	(0.046)	(0.053)	(0.050)	(0.043)
	3rd Year	0.171***	0.144**	0.229**	0.168**	0.195**	0.115**	0.102*	0.137*	0.095*	0.157**
		(0.047)	(0.052)	(0.079)	(0.061)	(0.060)	(0.038)	(0.045)	(0.058)	(0.047)	(0.055)
Post-Reform Years	1st Post	0.154**	0.096	0.281***	0.162*	0.157*	0.120**	0.118*	0.115	0.101	0.156**
	Year	(0.057)	(0.068)	(0.079)	(0.080)	(0.062)	(0.044)	(0.049)	(0.072)	(0.059)	(0.056)
	2nd Post	0.172**	0.141*	0.239***	0.192**	0.138	0.161***	0.141**	0.195**	0.151**	0.174**
	Year	(0.055)	(0.070)	(0.064)	(0.069)	(0.077)	(0.043)	(0.049)	(0.068)	(0.055)	(0.056)
	3+ Post	0.117*	0.085	0.193*	0.121	0.073	0.157**	0.130*	0.192*	0.141*	0.165**
	Year	(0.054)	(0.063)	(0.076)	(0.065)	(0.081)	(0.051)	(0.053)	(0.094)	(0.064)	(0.052)
N		1,301	1,301	1,301	1,301	1,301	1,296	1,296	1,296	1,296	1,296

Notes: The table shows the estimated longitudinal SIG effects on Grade 3-8 (HS excluded) Math and Reading/ELA performance (SD) in the pooled sample. Scores have been standardized by location, grade, year, and test. 3+ Post Year is the estimated effect of SIG in the third year and beyond (when data are available). Trans = Transformation model adopted, Turn = Turnaround model adopted, C1 = SIG Cohort 1, C2 = SIG Cohort 2. Robust standard errors clustered at the school level are presented in parentheses below estimates.

We used 1:1 nearest neighbor propensity score matching, and matched schools on pre-reform characteristics (2007-08 to 2009-10) in each of the 3 years prior to SIG: % race/ethnicity, % female, % ELL, average math and ELA/Reading score, natural log of total enrollment, grade span, and region.

* p<0.05, ** p<0.01, * p<0.001

Table A4. Estimated Longitudinal Effects of SIG on 4-year High School Graduation Rate on Matched Samples

		All SIG	Trans	Turn	C1	C2
Reform Years	1st Year	2.495	2.424	3.227	2.466	3.797
		(2.137)	(2.105)	(4.655)	(3.264)	(2.558)
	2nd Year	2.493	3.192	-3.027	3.812	3.648
		(2.237)	(2.297)	(4.722)	(2.101)	(3.162)
	3rd Year	5.204	5.534	3.739	9.677**	1.053
		(2.873)	(2.878)	(4.533)	(2.965)	(4.365)
Post-Reform Years	1st Post Year	6.456*	6.086*	9.249*	10.879***	-0.436
		(2.553)	(2.689)	(3.761)	(2.433)	(3.805)
	2nd Post Year	7.212**	7.127*	7.192*	9.249***	3.091
		(2.536)	(2.725)	(2.808)	(2.157)	(4.630)
	3+ Post Year	9.558***	9.548***	9.652**	11.175***	6.030
		(2.413)	(2.488)	(3.206)	(2.524)	(4.162)
N		528	528	528	528	528

Notes: The table shows the estimated longitudinal SIG effects on high school 4-year graduation rate in the pooled, matched sample. 3+ Post Year is the estimated effect of SIG in the third year and beyond (when data are available). Trans = Transformation model adopted, Turn = Turnaround model adopted, C1 = SIG Cohort 1, C2 = SIG Cohort 2. Robust standard errors clustered at the school level are presented in parentheses below estimates.
We used 1:1 nearest neighbor propensity score matching, and matched schools on pre-reform characteristics (2007-08 to 2009-10) in each of the 3 years prior to SIG: % race/ethnicity, % female, % ELL, average math and ELA/Reading score, graduation rate, natural log of total enrollment, grade span, and region.
* p<0.05, ** p<0.01, *** p<0.001

Table A5. Examining the Influence of Priority or Focus Schools Designations on Student Achievement

		Math					Reading/ELA				
		SIG	Trans	Turn	C1	C2	SIG	Trans	Turn	C1	C2
Reform Years	1st Year	0.132***	0.109*	0.185**	0.093	0.185***	0.076*	0.065	0.100	0.029	0.133***
		(0.038)	(0.046)	(0.062)	(0.059)	(0.043)	(0.035)	(0.044)	(0.051)	(0.054)	(0.035)
	2nd Year	0.167***	0.155***	0.195*	0.118*	0.241***	0.108**	0.089*	0.154*	0.051	0.189***
		(0.040)	(0.043)	(0.081)	(0.055)	(0.052)	(0.037)	(0.044)	(0.061)	(0.053)	(0.038)
	3rd Year	0.224***	0.182***	0.325***	0.203**	0.253***	0.161***	0.145***	0.199***	0.113*	0.228***
		(0.042)	(0.045)	(0.084)	(0.062)	(0.050)	(0.035)	(0.041)	(0.059)	(0.047)	(0.048)
Post-Reform Years	1st Post Year	0.190***	0.137*	0.318***	0.166*	0.223***	0.168***	0.162***	0.181*	0.157**	0.175***
		(0.052)	(0.063)	(0.078)	(0.078)	(0.054)	(0.038)	(0.042)	(0.077)	(0.054)	(0.048)
	2nd Post Year	0.215***	0.178**	0.304***	0.219**	0.199**	0.174***	0.150***	0.231**	0.153**	0.192***
		(0.053)	(0.066)	(0.066)	(0.073)	(0.072)	(0.039)	(0.044)	(0.074)	(0.055)	(0.053)
	3+ Post Year	0.163**	0.130*	0.239**	0.159*	0.135	0.162**	0.144**	0.203	0.137*	0.185***
		(0.050)	(0.060)	(0.077)	(0.067)	(0.073)	(0.050)	(0.053)	(0.108)	(0.069)	(0.049)
N		5,038	5,038	5,038	5,038	5,038	5,000	5,000	5,000	5,000	5,000

Notes: The table shows the estimated longitudinal SIG effects on Grade 3-8 (HS excluded) Math and Reading/ELA test scores in the pooled sample that only includes schools that had priority and focus designations in post-reform years. Scores have been standardized by location, grade, year, and test. 3+ Post Year is the estimated effect of SIG in the third year and beyond (when data are available).
* $p<0.05$, ** $p<0.01$, * $p<0.001$

Table A6. Examining the Influence of Priority or Focus Schools Designations on Graduation Rates

		All SIG	Trans	Turn	C1	C2
Reform Years	1st Year	7.233**	6.910**	10.835*	9.164**	7.245
		(2.307)	(2.272)	(4.486)	(2.735)	(3.913)
	2nd Year	6.671*	7.009*	0.999	10.946**	4.798
		(3.214)	(3.288)	(5.589)	(3.397)	(3.928)
	3rd Year	8.822*	9.009*	7.505	14.926***	2.919
		(3.856)	(3.813)	(6.775)	(3.858)	(5.159)
Post-Reform Years	1st Post Year	11.087**	10.854**	13.985**	16.433***	3.893
		(3.261)	(3.259)	(4.973)	(3.048)	(4.237)
	2nd Post Year	13.729***	13.664***	13.642**	16.686***	9.237
		(3.229)	(3.298)	(4.123)	(2.884)	(5.031)
	3+ Post Year	16.349***	16.265***	16.974***	19.263***	12.021**
		(3.023)	(3.099)	(3.462)	(3.017)	(4.482)
N		629	629	629	629	629

Notes: The table shows the estimated longitudinal SIG effects on high school 4-year graduation rate. The sample includes only schools that had priority and focus designations in post-reform years. 3+ Post Year is the estimated effect of SIG in the third year and beyond (when data are available). Robust standard errors clustered at the school level are presented in parentheses below estimates.
* p<0.05, ** p<0.01, *** p<0.001

Table A7. Examining the Influence of Entering Cohorts

		Prior Math			Prior Reading		
		NC	SFUSD	BCPS	NC	SFUSD	BCPS
Reform Years	1st Year Effect	0.061	-0.104	0.007	0.144**	0.193	-0.076*
		(0.053)	(0.119)	(0.037)	(0.054)	(0.129)	(0.034)
	2nd Year Effect	0.040	0.041	0.037	0.085	0.004	-0.090**
		(0.055)	(0.100)	(0.037)	(0.056)	(0.109)	(0.034)
	3rd Year Effect	0.110*	-0.118	0.056	0.070	-0.246	-0.060
		(0.053)	(0.154)	(0.039)	(0.054)	(0.172)	(0.036)
Post-Reform Years	1st Post Effect	0.083	0.045	0.070	-0.004	0.019	-0.034
		(0.056)	(0.207)	(0.041)	(0.058)	(0.225)	(0.038)
	2nd Post Effect	-0.034	n/a	0.093*	-0.009	n/a	0.026
		(0.058)	n/a	(0.044)	(0.059)	n/a	(0.041)
	3 Post Effect	0.089	-0.052	0.023	0.027	-0.059	-0.070
		(0.062)	(0.133)	(0.043)	(0.063)	(0.151)	(0.040)
N		429,480	6,221	111,701	427,639	6,079	112,184

Notes: This table show if SIG schools became more able to recruit higher performing students during and after SIG reform than the achievement of entering cohorts prior to SIG, compared to the trends in non-SIG schools. We used DiD regressions with incoming students' prior test scores in reading and math before they entered SIG their current schools as outcomes, but anything else similar to those illustrated in Equation (1). The regressions are run at student-level just for students who are new to the school (removing transition grades). The analyses do not include WA because WA do not have prior test scores for the entering cohorts of 2009-10 to support the DiD regressions. n/a indicates that California did not offer statewide standardized tests in two years.

* p<0.05, ** p<0.01, *** p<0.001