American Educational Research Journal
August 2017, Vol. 54, No. 4, pp. 607–643
DOI: 10.3102/0002831217695790

Resource- and Approach-Driven Multidimensional Change: Three-Year Effects of School Improvement Grants

Min Sun
University of Washington
Emily K. Penner
University of California, Irvine
Susanna Loeb
Stanford University

Hoping to spur dramatic school turnaround, the federal government channeled resources to the country's lowest-performing schools through School Improvement Grants (SIG). However, prior research on SIG effectiveness is limited and focuses primarily on student achievement. This study uses a difference-in-differences strategy to estimate program impacts on multiple dimensions across the 3-year duration of the SIG award in one urban school district. Following 2 years of modest improvement, we find pronounced, positive effects of SIG interventions on student achievement in Year 3, consistent with prior literature indicating that improvements from comprehensive school turnarounds emerge gradually. We also identify improvements indicating the process through which change occurred, including reduced unexcused absences, increased family preference for SIG schools, improved retention of effective teachers, and greater development of teacher professional capacity.

MIN SUN is an assistant professor in education policy at the College of Education at the University of Washington, 2012 Skagit Lane, M205 Miller Hall (Box 353600), Seattle, WA 98195; *misun@uw.edu*. Her research focuses on educator quality, school accountability, and school improvement.

EMILY K. PENNER is an assistant professor in the School of Education at the University of California, Irvine. Her research focuses on educational inequality and policy and considers the ways that districts, schools, teachers, peers, and parents can contribute to or ameliorate educational inequality.

SUSANNA LOEB is the Barnett Family Professor of Education at Stanford University. She specializes in education policy with a focus on school governance and finance and educator labor markets.

KEYWORDS: School Improvement Grants, school turnaround, school performance, school capacity

> Teachers in The Zone continue to work together to develop effective and engaging instructional practices. Principals, instructional coaches and school support teams provide strong leadership focused on continuous improvement. In addition, the adoption of a community-schools approach provides for enhanced student supports and aligned community partnerships. This combination of essential school supports is resulting in significantly improved outcomes for students.
>
> —Guadalupe Guerrero
> Deputy Superintendent for Instruction,
> Innovation, and Social Justice (SFUSD, 2012)

School Improvement Grants (SIGs) were part of a broader package of targeted federal initiatives intended to spur state and local school improvement introduced by the former Secretary of Education, Arne Duncan (e.g., Race to the Top, No Child Left Behind [NCLB] priority schools). In an effort to incentivize dramatic school transformations, Congress appropriated $3.5 billion for the first wave of SIGs through the American Recovery and Reinvestment Act (ARRA) to support states' "persistently lowest achieving" (PLA) schools (U.S. Chamber of Commerce Foundation, 2010; U.S. Department of Education, 2010a; U.S. Department of Education, 2010b). The U.S. Department of Education awarded California, which had the largest number of PLA schools in the country, nearly $416 million in SIG funds. The San Francisco Unified School District (SFUSD) received $45 million of this SIG funding to transform its 10 PLA schools between the academic years 2011 and 2013 (hereafter, we use the spring to refer to the academic year; e.g., 2010–2011 as 2011). SIG funding doubled these schools' budgets during the grant period (Wentworth, Khanna, & Piper, 2016).

SFUSD's SIG schools serve as examples of the potential effectiveness of the SIG program because of SFUSD's concerted efforts to implement the reforms using evidence-based guidelines. SFUSD designed its SIG reform plans using the five "essential supports" from the comprehensive school reform guidelines drawn from improvements in student-learning outcomes in Chicago Pubic Schools (Bryk, Sebring, Allensworth, Easton, & Luppescu, 2010). Moreover, SFUSD created "the Superintendent's Zone," an administrative structure aimed at providing administrative and curricular support to SIG schools to promote the successful implementation of SIG reforms (Wentworth et al., 2016). Its evidenced-based, comprehensive school improvement framework and focus on quality implementation make SFUSD a useful site for assessing the effects of SIG reforms in an urban district that attempts to use best practices to reform its most struggling schools.

In spite of the considerable resources marshaled for SIGs and the high expectations that SIG awards would produce substantial improvements in chronically underperforming schools, research demonstrating the effectiveness of SIGs and other whole-school reforms is inconsistent. A comprehensive report summarizing the research on whole-school reform efforts finds limited high-quality evidence assessing their effectiveness (Herman et al., 2008). To date, the majority of the existing work on SIGs is descriptive in nature and focuses on implementation (Council of the Great City Schools, 2015; Lachlan-Haché, Naik, & Casserly, 2012; Scott & McMurrer, 2015). A small number of recent studies that have estimated the causal impacts of SIG reforms on student outcomes either only gauge the effects after just the first year of the SIG award (Dee, 2012; Dickey-Griffith, 2013) or only focus on a limited set of academic outcome measures—mainly student test scores (e.g., de la Torre et al., 2013; Papay, 2015; Player & Katz, 2013).

In this study, we use nearly a decade of longitudinal data to examine SIG-program impacts across the full 3-year grant duration in SFUSD. Following gradual improvements in the first 2 years of reform, we find pronounced, positive effects of SIG interventions on student achievement in the third year. This pattern is consistent with the hypothesis that comprehensive school turnarounds need time for positive changes to occur in schools. We find evidence of the process of these changes, including the development of "essential supports" for organizing for school improvement identified by Bryk and colleagues (2010). Our analyses show a reduction in unexcused student absences in SIG schools. Families, particularly those with high-achieving students and from higher socioeconomic backgrounds, demonstrated increased preferences for SIG schools. SIG schools became better able to retain effective teachers and provide them with professional supports. These additional outcomes allow us to not only estimate temporal changes but also to examine the longer run effects of the SIG supports on these historically low-performing schools. To our knowledge, this is the first study that estimates SIG program impacts on student achievement across the 3-year period of the grant, incorporates multiple measures of SIG impacts on lowest performing schools, and attempts to uncover the mechanisms of change via staff capacity building.

Background

Research on Whole-School Reform Efforts

School reformers have promoted a variety of strategies to remedy underperformance in American schools (Coleman et al., 1966; Kantor & Lowe, 1995; National Commission on Excellence in Education, 1983). Several decades of whole-school reform efforts sought to spark improvements by modifying or restructuring struggling schools. In the 1980s and 1990s, Schoolwide Programs (SWPs) gave schools flexibility to reduce class size, hire staff, expand professional development offerings, increase teacher and parent involvement in decision making, and change classroom instruction (Wong & Meyer, 1998). Evaluations of the effects of these programs were minimal, limited by the use of small, nonrandom samples of participating schools, and often lacked causal rigor (Sunderman, 2001; Wang, Wong, & Kim, 1999; Wong & Meyer, 1998). Pushing for more dramatic improvement, Congress enacted the Comprehensive School Reform (CSR) Demonstration program in 1997, intending to improve curriculum, instruction, organization, professional development, and parental involvement (Desimone, 2002). Evaluations of CSR programs found mixed impacts (Bifulco, Duncombe, & Yinger, 2005; Bloom, Ham, Melton, & O'Brien, 2001; Cook et al., 1999; Gross, Booker, & Goldhaber, 2009) and wide variation in implementation and comprehensiveness of implementation (Aladjem et al., 2006; Berends, 2000; Berends, Bodilly, & Kirby, 2002; Rowan & Miller, 2007). G. D. Borman et al.'s (2003) meta-analysis of studies of 29 CSR programs showed positive impacts of several CSR reforms, with the largest effects resulting from reforms that were implemented for the longest amount of time (i.e., 5 years or more). Title I of NCLB also funded turnaround efforts, prescribing dramatic restructuring in hopes of improving student achievement and attainment. Under NCLB, schools that failed to meet annual yearly progress goals for multiple years in a row were closed and restructured. Ahn and Vigdor (2014) found that the threat of closure and leadership change improved student test score performance for schools first entering the NCLB sanction regime, but schools under threat of weaker consequences showed no evidence of improvement.

Seeking to draw lessons from the mixed track record across multiple waves of whole-school reforms, researchers worked to identify best practices and develop a theory of action to guide restructuring schools. These guidelines highlighted the importance of capacity building among school and district leaders and teachers, garnering faculty and parent support and commitment through relationship-building, implementing strategies from research-based plans, giving greater flexibility to adapt reform and financial resources to specific contexts, and making visible improvements early on in the turnaround process (K. M. Borman, Carter, Aladjem, & LeFloch, 2004; Herman et al., 2008; Hess, 1999; Malen & Rice, 2004; Mintrop & Trujillo, 2005; Spillane & Thompson, 1997). In addition, Bryk et al. (2010) generated conclusions from fieldwork in Chicago to create a "theory of practice" around school transformations that echoed many of the conclusions drawn from other whole-school reform efforts.

Research on School Improvement Grants

In spite of multiple policy efforts to spur changes in low-performing schools, many schools continued to struggle. The Obama Administration drew national focus to chronically underperforming schools by targeting SIG funds to PLA schools, making this policy a marquee component of the American Recovery and Reinvestment Act in 2009. PLA schools were defined as schools that were eligible for Title I assistance with baseline achievement in the lowest 5% (based on 3-year average proficiency rates) and that had made the least progress in raising student achievement over the previous 5 years. To receive funding, SIG schools were required to adopt one of four intervention models beginning in 2011 (U.S. Department of Education, 2010a). The *transformation model* required replacing the principal, implementing curricular reform, introducing teacher evaluations based in part on student performance, and incorporating evaluation results into personnel decisions (e.g., rewards, promotions, retentions, and firing). The *turnaround model* included all of the requirements of the transformation model, as well as replacing at least 50% 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.

Early research examining SIGs is primarily descriptive, providing progress reports focused on implementation (Council of the Great City Schools, 2015; Lachlan-Haché et al., 2012; Scott, Krasnoff, Davis, & Northwest, 2014; Scott & McMurrer, 2015). Evidence of SIG impacts on student outcomes is emerging. Two studies—Dee (2012) and Dickey-Griffith (2013)—examine first-year impacts. Dee (2012) uses a "fuzzy" regression discontinuity design based on two school-level eligibility thresholds—"lowest achieving" and "lack of progress"—and finds significant improvement in posttreatment performance in schools whose baseline proficiency rate just met the lowest achieving threshold but not among schools on the "lack of progress" margin. Dee also finds some evidence that SIG awards contribute to reductions in suspensions and truancy rates, but primarily among "turnaround" schools, which undergo more dramatic staff and principal replacement than other models. In contrast, Dickey-Griffith (2013) uses a difference-in-differences approach to assess 1-year impacts in Texas and finds mixed results, including negative impacts on student achievement in elementary and middle school and positive effects on high school graduation rates.

Recent work also examines SIG impacts beyond the first year, again providing mixed evidence of SIG effectiveness. Papay (2015) finds large, positive effects on math and English language arts (ELA) scores of being identified as SIG-eligible, which grow from the first to the third year of implementation in Massachusetts. A new report from the U.S. Department of Education uses data from 22 states and finds a null impact on test scores, high school graduation, and college enrollment for the cohort of schools funded in 2010 (Dragoset et al., 2017). A possible explanation for the difference in findings across studies is the variation in the design and implementation of SIG interventions across districts and states. Another possible explanation is that the heterogeneous results may result from sample selection and estimation strategies, as illustrated in Henry and Guthrie (2016)'s work in North Carolina.

Several large urban districts embedded SIG schools within other reform approaches, particularly "portfolio models" (Hill, 2006). For example, Los Angeles Unified School District (LAUSD) implemented a Public School Choice Initiative (PSCI), in which stakeholders compete to turn around the district's lowest performing "focus" schools. Initial research finds that reform plans were only sometimes associated with reported implementation and had inconsistent effects on student achievement across three rounds of PSCI-driven turnarounds (Strunk, Marsh, Bush-Mecenas, & Duque, 2015; Strunk, Marsh, Hashim, & Bush-Mecenas, 2016; Strunk, Marsh, Hashim, Bush-Mecenas, & Weinstein, 2016). In contrast, research examining New Orleans portfolio district reforms indicates positive effects on both student achievement and behavior (Barrett & Harris, 2015; Harris & Larsen, 2016; McEachin, Welsh, & Brewer, 2016; Welsh, Duque, & McEachin, 2016).

A companion initiative under the Obama administration, Race to the Top (RttT), funded similar, highly prescribed, school turnaround strategies. Evidence from several states that won RttT funding provides mixed evidence of effectiveness. Heissel and Ladd (2016) find negative effects of the program in North Carolina, and Zimmer, Henry, and Kho (2015) find some positive effects in Tennessee, particularly among Innovation Zone schools that were managed by school districts.

SIG and School Turnaround Models in SFUSD

As districts across the country drafted SIG proposals, SFUSD's central office prepared an application and conducted a needs assessment, examining the challenges and priority needs of each of the 10 SIG-eligible schools. The needs assessment indicated the 10 schools had incoherent curricula, assessments, and instructional guidance; insufficient resources and classroom materials; a lack of comprehensive interventions and monitoring of student progress; and haphazard implementation of improvement strategies that rarely lasted beyond a few years. SIG-eligible schools lacked resources to engage with families and did not comprehensively meet the needs of the community. Principals in some of the schools lacked the instructional leadership needed to dramatically improve student performance. In addition, the secondary schools experienced low engagement and high truancy (SFUSD, 2010).

In response, a district committee created a joint application for the 10 schools, first identifying the reform model each school would adopt. Because SFUSD had more than nine SIG-eligible schools, it could only use the transformation model in up to half of the eligible schools (Norton, 2010). During this process, district leaders sought strategic input from stakeholders and held school-site discussions and community meetings at each school; five schools chose the transformation model, four schools chose turnaround, and the lowest performing school chose closure (California Department of Education, 2010).[2]

Based on the needs assessment, the district adopted the five "essential supports" from Bryk et al.'s (2010) "Organizing Schools for Improvement: Lessons from Chicago" to develop a coordinated effort for school improvement in these nine schools. This plan highlights the ways in which each support fits within the required components of the SIG application, which is summarized below.

- *Activating school leadership as the driver for change:* In addition to removing principals who had been at a SIG school for more than 2 years and providing new principals with more flexibility over hiring, SFUSD redesigned the ways in which the district central office provided support to schools. The SIG schools were organized into two zones—Bayview and Mission—with corresponding district resources to strengthen management and provide continuous support and mentoring for school personnel (SFUSD, 2010; Wentworth et al., 2016).
- *Developing professional capacity among teachers:* SFUSD provided job-embedded teacher professional development featuring one-on-one coaching. Moreover, SFUSD instituted a performance management system using common interim assessments and other evidence of student learning to improve teaching practice (SFUSD, 2012).
- *Cultivating cohesive instructional guidance that promotes ambitious academic achievement for every child:* SIG schools were required to implement a Common Core curriculum that clearly specified what students should know and be able to do and set high standards for rigor and instructional quality. The schools also administered common interim assessments that tracked students' progress in meeting the standards. The schools partnered with third parties (e.g., Teacher's College, WRITE Institute, Algebraic Thinking & The Algebra Project, Project SEED, Tools for Schools, etc.) to focus on improving math and literacy instruction (SFUSD, 2010).
- *Nurturing a student-centered learning climate:* SIG schools extended learning time for students both after school and during the summer and implemented an early-warning monitoring system of student progress. In addition, secondary SIG schools promoted a college-going culture (SFUSD, 2010).
- *Fostering parent-community ties:* All SIG schools implemented a community-school approach beyond parent workshops that built family and community involvement and outreach (SFUSD, 2010).

SFUSD's application was successful. It received nearly all of the $45 million it requested. The district used the SIG money to implement the reforms outlined in the proposal, continuing to adopt the language of the Bryk at al. (2010) comprehensive reforms. Although the staffing transitions were more comprehensive at the turnaround schools than the transformation schools, all nine used the guidelines outlined in the application to structure their reforms.

The theory of change behind SFUSD's reforms is based on the belief that an incremental change is not sufficient to reform the "dysfunctional organizations" in low-performing schools and that dramatic restructuring of staff, curriculum, and environment is necessary (Malen & Rice, in press). This theory is similar to that articulated by Strunk, Marsh, Hashim, and Bush-Macenas (2016) suggesting that districts rely both on *incentives* to improve the productivity of staff in reconstituted schools, specifically the threat of additional reconstitution, and on *school capacity reinforcements*, including new staff, additional training, and funding, to support reforms. Research on multiple waves of whole-school reforms also demonstrates the need to build school capacity in order to yield systematic and sustained positive changes (K. M. Borman et al., 2004; Herman et al., 2008; Hess, 1999; Malen & Rice, 2004; Mintrop & Trujillo, 2005; Spillane & Thompson, 1997). This capacity-building strategy often requires time for schools to implement reconfiguration and install supports for teaching and learning.

Because SFUSD chose to model its reforms on Bryk et al.'s (2010) five-essential supports model, we assess impacts in several areas that might be indicative of such targeted reform efforts. First, we evaluate whether SIG schools progressed toward nurturing student-centered learning climates, coherent instructional guidance, and improved parent-community ties by examining changes in student achievement, attendance, and parent preferences for school placement. Developing a student-centered learning climate should result in improved student performance and increased attendance (Harris & Larsen, 2016; Jackson, 2012; McEachin et al., 2016). Improved parent-community ties and increased curricular rigor should make SIG schools more popular among families rather than stigmatize these schools as undesirable (Heissel & Ladd, 2016; Welsh et al., 2016), as revealed by preferences in school placement processes (Hastings, Kane, & Staiger, 2005; Hastings & Weinstein, 2008). In addition, changes in leadership, instructional guidance, and teacher professional capacity should be evident in improved retention of skilled teachers and in teacher-reported working environment, collaboration, administrative support, and mentorship (Barrett & Harris, 2015; Strunk, Marsh, Hashim, & Bush-Mecenas, 2016). We examine changes to the teacher workforce on measures of effectiveness and use survey results that include self-reports about teacher professional support, collaboration, and mentoring from school leadership.

The approaches embedded in SFUSD's SIG application and subsequent reforms feature prominently in current policy prescriptions for improving struggling schools under ESSA. Yet evidence supporting their effectiveness is limited to a sparse set of outcome measures that do little to illuminate the mechanisms driving change. This study seeks to address this limitation by providing a multiyear, in-depth evaluation of SIG reforms. It examines not only student outcome measures but also several indicators of organizational change that speak to the process through which SIG schools conducted their turnaround. This study provides the most thorough evaluation of SIG reforms to date.

Data and Methods

The data used in this study come from SFUSD. In the 2014–2015 school year, SFUSD was California's sixth largest district, serving approximately 58,000 students (California Department of Education, 2015). SFUSD's student body is both racially and socioeconomically diverse: 26% of its students identify as Latino, 41% as Asian, 11% as White, 10% as African American, 1% as Native American, and 10% as other. Twenty-seven percent speak English as a second language, and 61% are eligible for free or reduced-price lunch. SFUSD employs over 3,500 teachers to serve this student body. SFUSD's teaching force is also more diverse than the national average: 53% of teachers identify as educators of color compared with 18% of public school teachers nationwide (National Center for Education Statistics, 2015). SFUSD teachers have 11 years of teaching experience on average, which is slightly less than the national average of 14 years (National Center for Education Statistics, 2015; SFUSD, 2015, 2016). SFUSD schools demonstrate a substantial amount of performance heterogeneity, including both gold ribbon schools, acknowledged for outstanding and innovative performance, and SIG schools, identified as within the bottom 5% of the persistently lowest performing across the state in the same year.

Our analyses use SFUSD administrative data on students, teachers, and their schools from 2005 to 2013. We supplement the administrative data with 4 years of personnel survey data from 2010 to 2013. We exclude the one closure school from the analysis because the SIG award to this school was mainly used to facilitate students' transitions to new schools at the end of spring 2011 rather than invested in improving school capacity to raise students' learning outcomes. The SIG schools in our analysis sample, thus, include the five transformation and four turnaround schools.

Analytic Samples

Because of concerns regarding parent responses to SIG reforms that might motivate them to transfer their student in or out of SIG schools in response to the reform efforts, we estimate SIG effects using two approaches. The first approach compares student outcomes during the grant period (e.g., from 2011 through 2013) between those who were in SIG schools in fall 2010 (i.e., right at the beginning of the reform) and those who were in non-SIG schools at the same time, regardless of whether they transferred out of these schools in subsequent years. Hereafter, we call this group the "all starters." The estimate of SIG effects on this sample is analogous to what is called an intent-to-treat (ITT) effect in the experimental research literature, in that it represents the average effects for students who were in their assigned "treatment" conditions—either SIG or non-SIG—prior to the implementation of the intervention. The analysis sample does not include cohorts of students who newly enrolled in the schools in 2011 and 2012 after the intervention started, because parents might select or avoid SIG schools due to the SIG awards. Inclusion of these new cohorts may introduce bias in the estimation of SIG treatment effects.

While the "all starters" sample most cleanly removes issues of selection based on SIG assignment from the estimation, it may not accurately estimate the SIG effect because many of the "all starters" do move from the SIG schools and thus are not subject to the advantages or disadvantages of the SIG intervention. Our second approach further limits the sample to include only students who were in the same schools for at least 1 year prior to and 1 year after fall 2010 and did not transfer between schools during the intervention period. We call this group "stayers," which is somewhat analogous to the traditional "treatment-on-treated" sample in the experimental research literature. The estimate of SIG effects from this sample covers only those students who actually received the treatment of attending a SIG school for at least 1 year. Results are largely consistent between these two samples. We focus on the "all starters" sample in the main text, which provides the more conservative estimates, and include all corresponding results for the "stayers" sample in the online appendix.

Table 1 summarizes the descriptive comparisons of baseline student attributes in 2010 between SIG and non-SIG schools and between turnaround and transformation SIG schools for the two samples described above. Almost all of the observed preintervention student characteristics differ significantly between SIG and non-SIG schools. For example, SIG schools served students who were lower performing, had more disciplinary issues as indicated by the days of unexcused absences and suspensions, were more likely to be minorities, to be English language learners (ELLs), and to come from socioeconomically disadvantaged families. Similarly, among SIG schools, turnaround schools served lower performing, higher minority, and more socioeconomically disadvantaged students in 2010 than did transformation schools.

We describe changes in student composition prior to and during each year of the reform in online appendix Table B1. SIG schools kept students with higher prereform average math and ELA scores in postreform years in the "stayers" sample, underscoring the importance of controlling for prereform differences in student characteristics and achievement in our estimation of SIG impacts. These controls account for peer changes and sample selection.[3]

Analytical Approaches by Types of Outcome Measures

To examine a broad spectrum of SIG impacts in response to reforms grounded in the "essential supports" outlined by Bryk and colleagues

Table 1

Comparisons of Student Characteristics for "All Starters"

	SIG	Non-SIG	Transformation	Turnaround
Math standardized test scores[a]	–0.69 (0.70)	0.11 (1.00)	–0.65 (0.70)	–0.80 (0.71)
ELA standardized test scores[a]	–0.62 (0.82)	0.10 (0.99)	–0.56 (0.81)	–0.75 (0.83)
Days of excused absences	5.81 (8.33)	4.31 (5.96)	5.60 (8.13)	6.24 (8.72)
Days of unexcused absences[a]	11.93 (15.56)	6.74 (13.37)	12.43 (15.95)	10.89 (14.68)
Days suspended[a]	0.19 (1.04)	0.08 (0.68)	0.16 (0.91)	0.27 (1.25)
Race: White	0.02 (0.15)	0.11 (0.31)	0.03 (0.16)	0.02 (0.14)
African American[a]	0.18 (0.38)	0.09 (0.29)	0.12 (0.33)	0.30 (0.46)
Hispanic[a]	0.61 (0.49)	0.21 (0.41)	0.65 (0.48)	0.54 (0.50)
Asian[a]	0.12 (0.33)	0.51 (0.50)	0.16 (0.36)	0.06 (0.23)
Other[a]	0.06 (0.24)	0.07 (0.26)	0.05 (0.22)	0.08 (0.28)
Students in special education programs	0.14 (0.34)	0.11 (0.31)	0.14 (0.35)	0.13 (0.34)
English language learners[a]	0.46 (0.50)	0.28 (0.45)	0.48 (0.50)	0.42 (0.49)
Log of neighborhood median household income[a]	10.95 (0.48)	11.10 (0.45)	11.02 (0.43)	10.81 (0.53)
N (students)	2,644	37,094	1,782	862

Note. Means are reported with standard deviations in parentheses.
[a]Significant differences in means between turnaround and transformation schools.

(2010), we examine student attendance, student achievement, family preferences for SIG schools, teacher retention based on effectiveness and seniority, and teacher support. While these are not by any means an exhaustive set of indicators of these changes, our outcome measures do provide several pieces of evidence indicating the extent to which changes aligned with the reform's theory of action. Due to the varied nature of the outcome measures, we employ several analytic strategies and functional forms, which we describe in detail below.

Student Achievement

Because restricting analysis to either the "all starters" or "stayers" sample leads to a small number of students with test scores prior to 2008, the main analysis on student achievement uses 6 years of data from 2008 through 2013. We define whether a student was in a SIG school or a non-SIG school by his or her school attendance in the year before implementation (i.e., fall 2010). We then test whether students who initially attended SIG schools showed higher achievement over the subsequent 3 years than students who initially attended non-SIG schools[4] relative to prereform differences between SIG and non-SIG schools, controlling for their preintervention characteristics. The logic is similar to a traditional difference-in-differences (DD) approach.[5] Equation 1 describes the model.[6]

$$A_{igst} = \alpha_0 + \beta_1 Year_{2011} + \beta_2 Year_{2012} + \beta_3 Year_{2013} + \beta_4 (Year_{2011})(SIG_s) + \beta_5 (Year_{2012})(SIG_s) + \beta_6 (Year_{2013})(SIG_s) + X_{igst}\gamma_1 + \boldsymbol{\omega}_g + \boldsymbol{\nu}_s + \varepsilon_{igst}, \quad (1)$$

where A_{igst} is the math or ELA standardized test score of student *i* in grade *g*, school *s*, and year *t* on the California Standards Tests (CST). Although the subscript for subjects is omitted, we conduct the estimation separately for math and ELA. $Year_{2011}$ is the dummy indicator for observations in 2011—the first year of SIG interventions; $Year_{2012}$ indicates the second year; and $Year_{2013}$ indicates the third year. SIG_s is a time-invariant school-level indicator for the nine SIG schools. β_4, β_5, and β_6 indicate the treatment effect estimate in each of the treatment years by contrasting the difference in the average student achievement between the pre- and post-2010 school years in SIG schools with the difference in the average achievement between pre- and post-2010 in non-SIG schools.

Students were not randomly assigned to schools before the intervention. To account for student selection bias, student controls, X_{igst}, are added to the model, including students' race and ethnicity (Black, Hispanic, Asian, others), gender, ELL and disability designations, and whether either parent has a BA degree or higher. Instead of using lunch subsidies as a proxy for student socioeconomic background, we use several measures of student neighborhood socioeconomic status via their geocoded home addresses. By linking students' geo-coded addresses with the U.S. Census Bureau American Community Survey (ACS) data, we obtained the 5-year characteristics of neighborhoods (2007–2012) where the students lived, including the log of median household income, percentage with a bachelor's degree or higher among residents who are 25 and older, percentage of residents 18 or under living below the poverty threshold, and the log of median housing value (owner occupied). We include students' average achievement in the subject area prior to 2010 in order to further account for their preintervention differences. A second specification controls for students' prior-year test scores to capture SIG effects on the year-to-year student improvement, rather than controlling for students' average achievement prior to 2010. Although this second model may underestimate the treatment effects since it adjusts for a score that is likely a function of the treatment, it has the potential advantage of absorbing more of the differences between students in the SIG and non-SIG schools. Both models also include grade fixed effects, $\boldsymbol{\omega}_g$, to account for differences in academic tests across grades, and school fixed effects, $\boldsymbol{\nu}_s$, to control for time-invariant heterogeneity across schools. ε_{igst} is the error term. Because SIG strategies are whole-school reform efforts, we estimate cluster robust standard errors at the school level to adjust for correlations within schools and the influence of small number of treatment clusters on standard error estimates.[7]

We then use a similar strategy to estimate potential differential effects of transformation and turnaround models:

$$\begin{aligned}A_{igst} = {} & \alpha_0 + \beta_1 Year_{2011} + \beta_2 Year_{2012} + \beta_3 Year_{2013} + \beta_4 (Year_{2011})(Transformation_s) \\ & + \beta_5 (Year_{2012})(Transformation_s) + \beta_6 (Year_{2013})(Transformation_s) \\ & + \beta_7 (Year_{2011})(Turnaround_s) + \beta_8 (Year_{2012})(Turnaround_s) \\ & + \beta_9 (Year_{2013})(Turnaround_s) + X_{igst}\gamma_1 + \boldsymbol{\omega}_g + \boldsymbol{\nu}_s + \varepsilon_{igst},\end{aligned} \tag{2}$$

where $Transformation_s$ is a time-invariant school-level indicator for schools that implemented the transformation model, and $Turnaround_s$ indicates schools that chose the turnaround model. β_4, β_5, and β_6 estimate SIG effects in transformation schools (relative to non-SIG schools) separately by each intervention year, while β_7, β_8, and β_9 capture SIG effects in turnaround schools (relative to non-SIG schools). The remaining are consistent with those in Equation 1. The numbers of Transformation and Turnaround schools are small and, as a result, we treat these estimates with caution.

Student Absences

We use data from 2008 through 2013 to estimate the effects of SIG reforms on student absences. Because absences are a relatively rare occurrence for most students, we use negative binomial models to estimate the count of students' full-day absences as a function of SIG policy treatment and student and family characteristics.[8] The identification strategy is the same as illustrated in Equation 1, estimating SIG effects using the comparison of the change in the average probability of student absences before and after the reform, between SIG and non-SIG schools. Beyond including student controls that are used to model student achievement, when modeling absences, we also control for the distance from a student's home to his or her school to account for absences due to transportation difficulties.

We estimate effects separately for excused and unexcused absences. A full-day absence was recorded for pupils who were absent for more than 84% of the regularly scheduled school day. The State of California Education Code 48205 states that a legitimate excused absence has to be initiated by parents or legal guardians. Excused absences can be due to student illness, medical appointment, or justifiable personal reasons, including an appearance in court, attendance at a funeral service, religious holiday or ceremony, or a visit to a college or university.[9] Because of the relatively restrictive rules for legitimate excused absences, we anticipate less variation in excused absences across schools and over time, as well as less change within schools following the implementation of SIG reforms. In contrast, we expect unexcused absences to align more closely to student and parent school engagement, as well as monitoring systems of student progress, all of which are the targets of SIG reforms.

Family Preferences

As an indication of community and parent responses to SIG improvement efforts, we use student-family school choice on enrollment preference forms. SFUSD uses a Student Assignment System, which has been in place since 2003, to assign all students to all of its schools through a choice process designed to provide equitable access to the range of opportunities available in San Francisco's public schools. This process is described in greater detail in online appendix C. The large majority of students submit choice forms when they enter kindergarten, sixth, and ninth grades, when they initially enter the district, or if they want to transfer schools (ranging from 60% to 70% in the early 2000s to about 90% in more recent years).

Our choice analysis restricts the sample to all students who could have chosen the SIG schools over 9 years from 2005 through 2013. This includes all students who applied for the grade level in which a school receives a new cohort of students. For example, all kindergarten applicants are the potential choosers of an elementary school; similarly, all sixth-grade applicants, of a middle school; and all ninth-grade applicants, of a high school. Although many students listed more than one choice, we model students' first choices, because these schools are families' most desired choice.[10]

The identification strategy of comparing top preferences for SIG schools and non-SIG schools would not be appropriate for analyzing choices, because the increase in desirability of SIG schools would, by default, result in a decrease in non-SIG schools' desirability. In other words, the change in non-SIG schools' trends is dependent on the change in SIG schools' trends, and vice versa. This interdependency would violate the common-trends assumption of a difference-in-differences approach. Instead, we use an interrupted time series (ITS) approach to identify the postintervention deviations from the preintervention trend in student choice of the SIG schools. We model the likelihood that a student chooses a SIG school as his or her first choice ($y_{it} = 1$) as a function of postintervention duration ($Year_{2011}$, $Year_{2012}$, and $Year_{2013}$), controlling for the year trend (*Year*), the same student characteristics as in Equation 1 and the proximity from his or her home to the school in the vector of $\mathbf{X}_{it}$, and school fixed effects ($\boldsymbol{\nu}_s$) (Burgess, Greaves, Vignoles, & Wilson, 2014; Hastings & Weinstein, 2008). A logit regression model is summarized in Equation 3, where *Year* is a year linear term, centered on the reform year of 2010, and the standard errors are school-level cluster robust standard errors.

$$\log\left(\frac{p(y_{it}=1)}{1-p(y_{it}=1)}\right) = \alpha_0+\alpha_1 Year+\beta_1 Year_{2011}+\beta_2 Year_{2012}+\beta_3 Year_{2013} + \mathbf{X}_{it}\gamma_1+\nu_s+\varepsilon_{it}. \tag{3}$$

Effective Teacher Retention

SIG schools aim to disrupt retention policies based solely on seniority and implement a system that prioritizes the hiring and retention of *effective teachers*. Without knowing all dimensions that principals use to select teachers, we measure teacher effectiveness using an annual value-added measure, which gauges teachers' contribution to raising student achievement (see online appendix A for details on the estimation of value-added). We average 3 years of value-added measures in their respective subjects—the current year and two prior years—to create our teacher effectiveness measure. This measure accounts for concerns about year-to-year fluctuation of value-added measures due to the variation in true teacher performance over time and measurement error (Loeb & Candelaria, 2012).[11] In each year, we have between 66 and 88 teachers in the nine SIG schools with value-added, which represents approximately 22% of teachers in these schools.

As shown descriptively in online appendix Table B2, SIG schools kept teachers with higher value-added scores during the reform period than the prereform year. To formalize this observation, we use a strategy, similar to the difference-in-difference-in-differences (DDD) framework, with a conditional logit function to examine whether the relationship between teacher effectiveness and retention became stronger in SIG schools relative to non-SIG schools in post-SIG years, compared to the pre-SIG years.

$$\begin{aligned}\log\left(\frac{p(y_{jst}=1)}{1-p(y_{jst}=1)}\right) &= \alpha_0+\beta_1 Year_{2011}+\beta_2 Year_{2012}+\beta_3 Year_{2013}+\beta_4(Effectiveness)_{jst}\\ &+\beta_5(Year_{2011})(SIG_s)+\beta_6(Year_{2012})(SIG_s)+\beta_7(Year_{2013})(SIG_s)\\ &+\beta_8(Year_{2011})(Effectiveness)_{jst}+\beta_9(Year_{2012})(Effectiveness)_{jst}\\ &+\beta_{10}(Year_{2013})(Effectiveness)_{jst}+\beta_{11}(SIG_s)(Effectiveness)_{jst}\\ &+\beta_{12}(Year_{2011})(SIG_s)(Effectiveness)_{jst}+\beta_{13}(Year_{2012})(SIG_s)(Effectiveness)_{jst}\\ &+\beta_{14}(Year_{2013})(SIG_s)(Effectiveness)_{jst}+\mathbf{X}_{jst}\gamma_1+\boldsymbol{\nu}_s+\varepsilon_{jst},\end{aligned} \tag{4}$$

where y_{jst} is the retention status of teacher j in school s and year t ("1" = stays in current school in the following year, excluding retirement; "0" =

otherwise). Although the subscript for subjects is omitted, we conduct the estimation separately for math and ELA teachers. $(Effectiveness)_{jst}$ indicates the 3-year average teacher value-added estimates. The coefficients of the three-way interactions, β_{12}, β_{13}, and β_{14}, indicate the SIG effects on retaining effective teachers in Year 1, Year 2, and Year 3, respectively. $\mathbf{X}_{jst}$ includes teacher demographics and professional background (e.g., having a master's degree, majored in education in the highest degree, and in first 3 years of teaching), as well as school characteristics (percentage of White, Black, Hispanic, or Asian students; average school-level days of suspension, percentage novice teachers, average of student socioeconomic characteristics). $\boldsymbol{\nu}_s$ indicates school fixed effects. Again, we used cluster robust standard errors at the school level. In a second model, we replace effectiveness with being an experienced teacher (e.g., >3 years of teaching experience) to assess whether experienced teachers became more or less likely to stay in SIG schools during reform period.

Teacher Supports

We also investigate how well SIG schools succeeded in developing the professional capacity of their teachers, creating cohesive instructional guidance, and using leadership as a driver for change using annual teacher survey data between 2010 and 2013.[12] We use a set of the questions from the surveys that focused specifically on teachers' reports of the supportiveness of their school environments, their mentoring from school leaders, and their collaboration and mutual support as a teaching team. Teachers were asked, on a 7-point scale including *never* (0), *once* (1), *twice* (2), *3 or 4 times* (3), *5–9 times* (7), and *10 or more times* (10) within each year, about the frequency of (a) visiting another teacher's classroom to watch him or her teach; (b) having a colleague observe your classroom; (c) inviting someone in to help your class; (d) going to a colleague to get advice about an instructional challenge you faced; (e) receiving useful suggestions for curriculum material from colleagues; (f) receiving meaningful feedback on your teaching practice from colleagues; (g) receiving meaningful feedback on your teaching practice from your principal; and (h) receiving meaningful feedback on your teaching practice from another school leader (e.g., AP, instructional coach). We derive a composite measure of *teacher supports* by taking the mean across these items.[13]

We examine changes in teacher supports from 2010 to 2013 in SIG schools relative to non-SIG schools using an approach similar to Equation 1. The analysis includes responses from all teachers present in each year, because these survey results are intended to take the pulse of the current teaching climate in SIG versus non-SIG schools in both the pre-SIG and during-SIG periods. These models include the composite measure of *teacher supports* as the dependent variable, post-SIG year indicators, the school and teacher controls described above, and school fixed effects, and use school-level robust standard errors.

Robustness and Falsification Tests

There are three potential threats to the causal inference of the DD design. First, DD designs assume that trends in SIG schools would have been the same as those in non-SIG schools without the reforms. We examine prereform trends to assess the validity of this assumption. A second concern is that other factors produced or contributed to any changes in SIG schools at the same time as the SIG reforms, which is difficult to assess. However, we provide some evidence of the prominence of the SIG reforms relative to any other concurrent factors. A third potential concern is mean-reversion, in which the lowest achieving schools experience larger than average gains in years following the SIG intervention (e.g., Ahn & Vigdor, 2014; Figlio & Rouse, 2006). In other words, the increase in student achievement in years following SIG implementation is not due to a SIG treatment effect but rather due to the unusually low scores prior to the intervention. Using techniques similar to those used by Figlio and Rouse (2006), we assess the threat of mean reversion in our data.

Results

For three of our outcome measures (achievement, attendance, and family preferences), we present a series of figures that graphically illustrate our analytic approach, followed by regression estimates in the "all starters" samples. We then present differential SIG effects for transformation and turnaround schools. For our remaining outcomes (teacher turnover and teacher supports), we present regression estimates of SIG effects across the 3 years of SIG reform.

Student Achievement

Our analysis provides evidence that SIG interventions significantly increased average student achievement in math and ELA and that the treatment effect is most pronounced in the third year of the intervention. Figure 1 compares the trends in average student achievement between SIG and non-SIG schools. Figure 1a shows that prior to reform, the average math score of the SIG "all starters" sample was –0.68 standard deviations (*SD*) in spring 2008 and –0.69 *SD* in spring 2010. The average math score of non-SIG schools was considerably higher, 0.17 *SD* in spring 2008 and 0.11 *SD* in spring 2010, resulting in a significant 0.80 *SD* gap in average math achievement right before the SIG intervention started. Notably, the pre-SIG trends are almost parallel in these two types of schools. After fall 2010, in obvious contrast to the pre-SIG trend, the mean math achievement raised much more

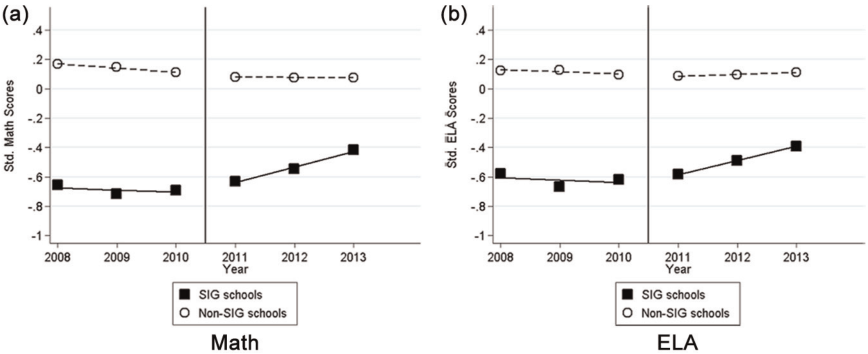

Figure 1. **Comparison of trends in student achievement between SIG and non-SIG schools for "all starters."**

Note. The "all starters" sample includes students those who were the district in fall 2010, regardless of whether they transferred between schools in subsequent years.

quickly in SIG schools than in non-SIG schools. By spring 2013, the third intervention year, the gap in average math achievement declined to 0.50 *SD* (i.e., 0.08 – [–0.42]). Figure 1b shows analogous results for ELA.

Table 2 presents regression estimates that formalize the patterns we observe in Figure 1. For each subject area, we have two model specifications: one which includes a control for average student achievement observed prior to the SIG reforms (Model 1), and a second which includes lagged student achievement from the prior year (Model 2). As indicated in the columns of "all SIG schools" in Model 1, the estimated SIG effect in math is 0.12 *SD* in 2011 and 0.07 *SD* in ELA. The Year 2 point estimates are higher in both math and ELA. In Year 3, the estimates are positive and significant: relative to the change in non-SIG schools, we estimate that SIG interventions improved student achievement by 0.24 *SD* in math with controls for average achievement prior to SIG and generated an average year-to-year improvement of 0.15 *SD*. We estimate that the SIG interventions significantly increased average ELA achievement by 0.12 *SD* in Year 3 and generated year-to-year improvements of 0.02 *SD*. Results for the "stayers" sample in online appendix Table B3 show consistent, somewhat larger positive effects of SIG interventions in both math and ELA in Year 3.

Although transformation and turnaround schools adopted many similar interventions, turnaround schools also replaced leaders and staff, potentially resulting in different treatment effects. We use Equation 2 to estimate SIG effects on student achievement in these two types of schools. Our analyses by reform type may be more exploratory than causal, because as illustrated in Figure B2 in the online appendix, the common trends assumption may not

Table 2

Estimated SIG Effects on Student Achievement for ‘‘All Starters’’

	Math					ELA				
	All SIG Schools		Transformation	Turnaround		All SIG Schools		Transformation	Turnaround	
	Model 1	Model 2	Model 1	Model 1	*F*	Model 1	Model 2	Model 1	Model 1	*F*
SIG effect in 2011	0.115***	0.107**	0.088*	0.170***	3.67†	0.066**	0.065*	0.065*	0.069	0.01
	(0.034)	(0.038)	(0.043)	(0.014)		(0.023)	(0.025)	(0.027)	(0.043)	
SIG effect in 2012	0.134	0.078	0.065	0.324***	3.45†	0.090*	0.025	0.076†	0.128	0.22
	(0.102)	(0.092)	(0.114)	(0.084)		(0.043)	(0.050)	(0.041)	(0.104)	
SIG effect in 2013	0.237*	0.149*	0.155	0.470***	3.34†	0.116*	0.021	0.106†	0.145†	0.17
	(0.119)	(0.070)	(0.116)	(0.129)		(0.051)	(0.029)	(0.058)	(0.077)	
Student achievement prior to SIG	X		X	X		X		X	X	
Student prior-year achievement		X					X			
Student covariates	X	X	X	X		X	X	X	X	
Grade fixed effects	X	X	X	X		X	X	X	X	
School fixed effects	X	X	X	X		X	X	X	X	
Adjusted R^2	.645	.662	.646			.706	.729	.706		
N (student-years)	112,856	121,995	112,856			112,680	122,026	112,680		

Note. Since school fixed effects are included in all model specifications, the main effect of the SIG treatment is then omitted. Model 1 includes student covariates and school fixed effects. Student covariates include students’ race-ethnicity, gender, disability, ELL, and whether either parent has a BA, as well as several measures of students’ neighborhood averaged across 5 years (from 2007 through 2012), such as the log of median household income, percentage of residents 25 or older who have a bachelor’s degree, percentage of residents 18 or under living in poverty, and log of median housing value (owner occupied). The models also include student average achievement prior to 2010–2011. Model 2 replaces the pre-reform student average in Model 1 with 1-year lagged achievement. School-level cluster robust standard errors are reported in parentheses. *F* statistics test the null hypothesis that the SIG effects between transformation and turnaround schools are the same.

$^{\dagger}p \leq .1$. $^{*}p \leq .05$. $^{**}p \leq .01$. $^{***}p \leq .001$.

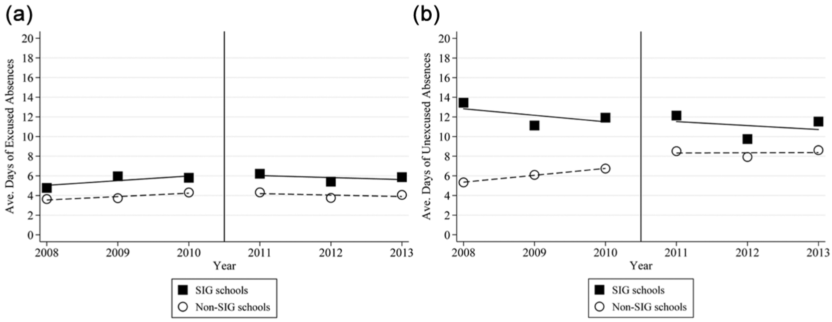

Figure 2. **Trends in full-day student absences in both SIG and non-SIG schools for "all starters"**

Note. The "all starters" sample includes students who were the district in fall 2010, regardless of whether they transferred between schools in subsequent years.

hold for some model specifications. However, it is still worthwhile to present the analyses, because prior literature suggested differential effects between these two types of reform models (Dee, 2012). As shown in the "Transformation" and "Turnaround" columns in Table 2, consistent across both reform models and in both subjects, SIG effects in the third year are generally larger than in the first 2 years of the intervention. Additionally, between-group comparisons suggest larger increases in mean math achievement in turnaround than transformation schools across all 3 years. For example, for the "all starters" sample in 2011, the estimated effect on average improvement controlling for average scores prior to the reform is 0.09 *SD* in transformation schools, which is smaller than the estimated 0.17 *SD* change in turnaround schools ($F = 3.67$, $p \leq 0.1$). Similarly, in 2012, the estimated average effect is 0.07 *SD* in transformation schools, compared with a much larger estimate of 0.32 *SD* in turnaround schools ($F = 3.45$, $p \leq 0.1$). In 2013, transformation schools had an estimated average effect of 0.16 and turnaround schools had an estimated effect of 0.47 ($F = 3.34$). Although none of the differences in the estimated effects on ELA between these two types of schools are statistically significant, the estimated effects in turnaround schools are still slightly larger than those in transformation schools.

Absences

Our analyses provide some evidence of changes in student attendance in response to increased monitoring of student progress under SIGs, but not strong evidence of effects. Figure 2 illustrates the changes in average full-day excused and unexcused absences from 2008 to 2013, separately for SIG and non-SIG schools. Figure 2a shows the average days of excused absences in the "all starters" sample and shows that SIG reforms did not lead to meaningful decreases in excused absences, because the gap in average days of excused absences between SIG and non-SIG schools did not change meaningfully from the pre- to postintervention periods. Figure 2b shows the average days of unexcused absences in the "all starters" sample and indicates decreases in both SIG and non-SIG schools after program implementation. Figure 2 suggests that SIG reforms may have reduced unexcused absences, as the gap between SIG and non-SIG schools shrank a little in the postintervention period, but the causal effects are not as clear as for achievement because the differences between SIG and non-SIG schools were closing in the years before reform as well.

The regression estimates in Table 3 confirm that SIG reforms had close-to-zero influence on students' excused absences but did reduce the likelihood that students had unexcused absences. For example, the incidence rate for full-day unexcused absences decreased by 18% in Year 1 in the "all starters" sample, by 24% in Year 2, and by 12% in Year 3. Although only Year 2 SIG effect estimates are consistently significant across model specifications, all estimates for unexcused absences are negative. These discrepant findings across the two types of absences are understandable given that unexcused absences are more likely to be malleable and a function of factors such as parental engagement and a student attendance monitoring system.

When we compare transformation with turnaround schools, the results in Table 3 do not show systematic differences between these two types of schools in estimated intervention effects on either excused or unexcused absences.

Family Preferences

Among all students who submitted school preferences, roughly 35% applied for kindergarten, 18% applied for sixth grade, 31% applied for ninth grade, and 1%–2% applied for each of the remaining grade levels. Figure 3a plots the percentage of all students who listed a SIG school as their first choice among those submitting choice preferences. The trend for SIG schools' popularity among families declined from 2005 to 2010, while the negative trend reversed after 2011. This pattern also emerges in logit regression results in the first column of Table 4. Among all students submitting choices in Year 1, the odds that students selected a SIG school as their first choice significantly increased by 31% relative to the odds of making the same choice before the intervention, after accounting for student characteristics, distance from their home to the school, and school fixed effects. The odds that students listed a SIG school as their first choice increased by 65% in Year 2 and 117% in Year 3.

Table 3

Estimated SIG Effects on Student Absences for "All Starters"

	Excused					Unexcused				
	All SIG Schools		Transformation	Turnaround		All SIG Schools		Transformation	Turnaround	
	Model 1	Model 2	Model 1	Model 1	χ^{2a}	Model 1	Model 2	Model 1	Model 1	χ^{2a}
SIG effect in 2011	1.129	1.14	1.141	1.093	0.07	0.819	0.792*	0.792	0.897	0.30
	(0.123)	(0.118)	(0.162)	(0.108)		(0.108)	(0.089)	(0.135)	(0.146)	
SIG effect in 2012	1.182	1.192	1.13	1.353	0.50	0.758*	0.747*	0.789†	0.646***	1.92
	(0.146)	(0.134)	(0.141)	(0.311)		(0.090)	(0.087)	(0.111)	(0.059)	
SIG effect in 2013	1.174	1.188	1.202	1.071	0.38	0.882	0.849	0.865	0.941	0.12
	(0.126)	(0.129)	(0.166)	(0.146)		(0.116)	(0.102)	(0.139)	(0.189)	
Student covariates	X	X	X	X		X	X	X	X	
Grade fixed effects	X	X	X	X		X	X	X	X	
School fixed effects	X	X	X	X		X	X	X	X	
School pre-SIG trend		X					X			
N (student-years)	124,714	124,714	124,714			124,714	124,714	124,714		

Note. The coefficients are incidence rate ratios (IRR). Cluster-robust standard errors at the school level are included in the parentheses. Model 1 includes student covariates and school fixed effects. Student covariates include their race-ethnicity, gender, disability, ELL, whether either parent has a BA, distance from home to school (in miles), and the log of median household income of students' neighborhood averaged across 5 years (from 2007 through 2012). Model 1 also includes student attributes prior to the reform, such as average days of absences and average math and ELA achievement prior to 2010–2011. Model 2 also includes school-specific pre-SIG trends.

[a]χ^2 statistics testing the null hypothesis that the SIG effects between transformation and turnaround schools are the same.

†$p \leq .1$. *$p \leq .05$. **$p \leq .01$. ***$p \leq .001$.

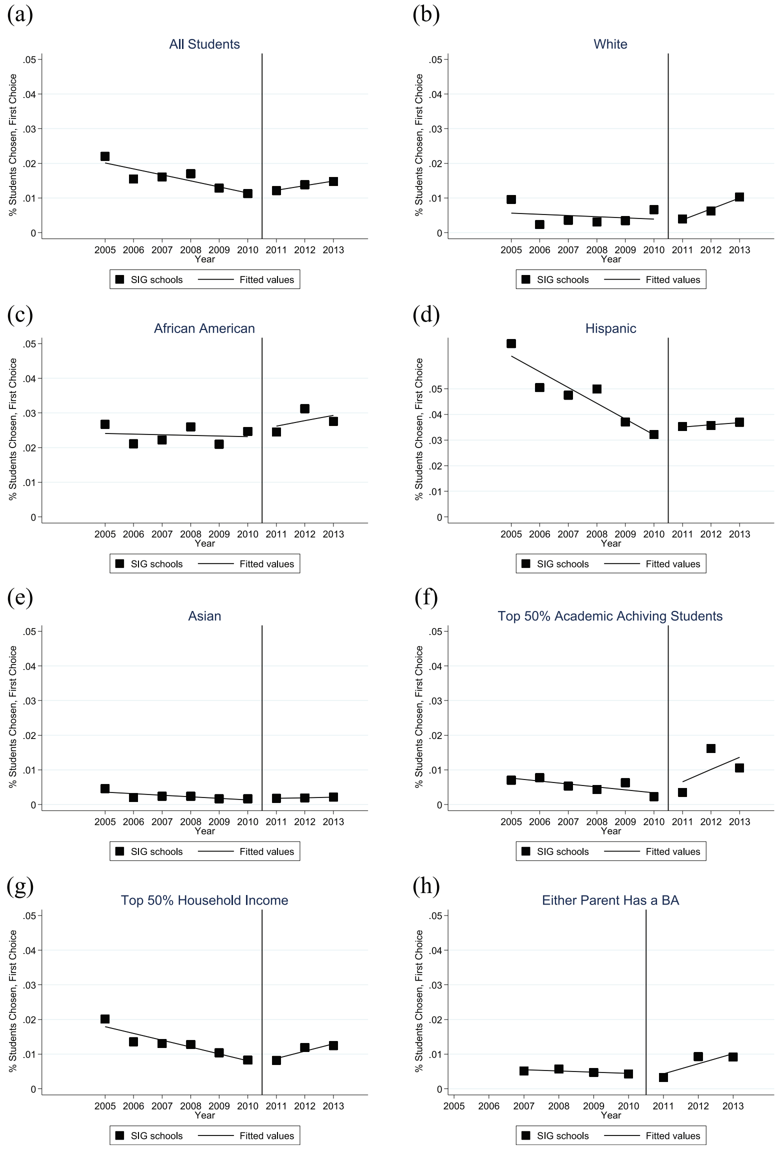

Figure 3. **Percentage of students listing a SIG school as their first choice.**

Changes in the popularity of SIG schools varied by subgroup. Figures 3b and 3c show increasing trends during the intervention period among White and African American students in particular. In Year 3, the odds that both

Table 4

Estimated SIG Effects on Students' School Choices

	All Students	White	Black	Hispanic	Asian	Above-Average Achieving	Above-Average Income	Either Parent Has a BA	Transformation	Turnaround	χ^{2a}
SIG effect in 2011	1.310†	0.737	1.423†	0.749	0.806	0.744†	0.609**	0.606	1.305*	1.329	0.01
	(0.189)	(0.306)	(0.270)	(0.189)	(0.188)	(0.129)	(0.105)	(0.190)	(0.168)	(0.345)	
SIG effect in 2012	1.645	1.543	1.857	0.649†	0.954	1.821***	1.000	1.741	1.691	1.484	0.10
	(0.634)	(0.723)	(0.719)	(0.153)	(0.211)	(0.198)	(0.260)	(0.624)	(0.740)	(0.443)	
SIG effect in 2013	2.168*	2.059***	1.978***	0.572**	0.758	1.215	0.933	1.984*	2.01	2.765**	0.53
	(0.818)	(0.352)	(0.384)	(0.110)	(0.285)	(0.274)	(0.085)	(0.539)	(0.909)	(0.970)	
Student covariates	X	X	X	X	X	X	X	X	X	X	
School fixed effects	X	X	X	X	X	X	X	X	X	X	

Note. N = 138,462. The coefficients are odds ratios. All specifications include school fixed effects and student covariates. Student covariates include race-ethnicity, gender, special education programs, ELL, gifted, and distance from home to school (in miles), as well as the log of median household income of students' neighborhood and whether either parent has a bachelor's degree. Cluster robust standard errors at school level are reported in parentheses.

[a] χ^2 statistics for testing the null hypothesis that the SIG effects between transformation and turnaround schools are the same.

$^{\dagger}p \leq .1$. $^{*}p \leq .05$. $^{**}p \leq .01$. $^{***}p \leq .001$.

African American and White students chose SIG schools were about twice as large as in the pre-SIG period. In contrast, Hispanic students became less likely to choose SIG schools in Year 3 and Asian students' choices were not influenced by SIG designations.

Figure 3f shows that high-achieving students (i.e., those scoring in the top 50% of the distribution of the prior-year average math and ELA scores in the district) became increasingly more likely to choose SIG schools during the reform period, as did students with at least one parent with a bachelor's degree (Figure 3h). Specifically, the odds for high-achieving students listing a SIG school as their first choice significantly increased by 82% in Year 2, relative to the odds that this group listed a SIG school first before the intervention. The odds that students from highly educated families ranked SIG schools as their first choice grew, on average, by 98% in Year 3.

We find no significant differences in desirability in the first 2 years between transformation and turnaround schools, as shown in the last three columns in Table 4.

Teacher Retention

A key piece of the SIG reforms involved staff reconstitution to improve its effectiveness. As shown in Table 5, the estimated SIG effects on retaining math teachers with higher value-added are positive in all intervention years and statistically significant in both Years 1 and 3. With a 1 *SD* increase in a typical teacher's value-added, the odds that this teacher remains in a SIG school significantly increased by 2.68 times in Year 1 and 1.78 times in Year 3, relative to the odds for similarly effective counterparts in non-SIG schools compared with the prereform years. We observe similar positive SIG effects for ELA teachers in these 3 years, with a particularly large effect in Year 2. Taken together, the results in both subjects provide evidence that the SIG schools were able to retain more effective teachers in the reform years than they had been able to do in prior years.

Replacing effectiveness with teacher experience in Equation 4, we observe the opposite pattern. The odds that an experienced teacher stayed in a SIG school declined by 87% in Year 1, compared with the odds of turnover for experienced teachers in non-SIG schools, after accounting for teacher value-added and other controls. This declining trend continued in Years 2 and 3. These findings indicate that during the reform, SIG schools became more likely to retain teachers based on their effectiveness and less likely to retain teachers based on seniority.

Teacher Supports

We document the ways in which SIG schools improved teacher capacity and instructional leadership by examining teacher reports of support for teaching. As shown in the last column in Table 5, there was no significant

Table 5

Estimates of SIG Effects on Teacher Retention and Support

	Retention of Teachers by Math Value-Added	Retention of Teachers by ELA Value-Added	Retention of Experienced Teachers	Teacher-Reported Support
SIG effect in 2011	3.678**	1.98	0.129*	–0.104
	(1.802)	(1.046)	(0.120)	(0.223)
SIG effect in 2012	2.173	4.270*	0.159	0.502*
	(1.666)	(2.859)	(0.190)	(0.226)
SIG effect in 2013	2.777*	1.501	0.327	0.806**
	(1.264)	(1.048)	(0.274)	(0.215)
Teacher covariates	X	X	X	X
School fixed effects	X	X	X	X
School time-varying covariates	X	X	X	X
χ^2	187.25	148.71	114.81	
Adjusted R^2				.122
N (teacher-years)	6,439	6,833	8,898	4,994

Note. Teacher covariates include their demographics and professional backgrounds (e.g., having a master's degree, majored in education in the highest degree). School time-varying characteristics include percentage of White, Black, Hispanic, or Asian; average school-level days of suspension; percentage of novice teachers; and averages of students' socioeconomic backgrounds. For the first two columns, teacher effectiveness in each subject is indicated by the 3-year average value-added in that subject. For the third column, we use an indicator for "experienced teachers" (i.e., >3 years of teaching experience in the district). Because the average of teacher experience increased over time, to avoid the strong correlation between year and experience, we use this dummy indicator. The coefficients are odds ratios in the first three columns and rating points in the last column. Cluster-robust standard errors at the school level are reported in parentheses.

$^{\dagger}p \leq .1$. $^{*}p \leq .05$. $^{**}p \leq .01$. $^{***}p \leq .001$.

difference in teacher-reported support after Year 1 of the SIG award. However, by the second year, SIG teachers reported a level of teacher support that was 0.50 points higher than the level reported in non-SIG schools (an increase of 0.26 *SD*). By the spring of 2013, this difference was 0.81 points higher (an increase of 0.41 *SD*).

Robustness and Falsification Analysis

The key assumption of the analytic approach used to analyze student achievement and absences is that the changes from pre- to postintervention periods 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-SIG trends to assess a possible violation. As shown in Figure 1, the pre-SIG trends in achievement measures were almost parallel between SIG and non-SIG schools, suggesting common pre-SIG trends.[14] As shown in Figure 3 for families' school choices, the general trend was consistently decreasing in the prereform period and then sharply increased after the reform started in 2011. There is no significant sign of discontinuity in the pre-SIG trend. Only for absences do we find some cause for concern. Figure B3 shows parallel trends for the "stayers" sample. Figure 2 shows some closing of the gap in unexcused absences between SIG and non-SIG schools in the "all starters" sample prior to the reforms. We statistically test this threat to the common trends assumption for absences by adding pretreatment, school-specific trends to Equation 1. Results are included in the Model 2 values in Table 3. The estimated SIG effects are largely consistent with our main models—Model 1 values in Table 3.

A second threat to the internal validity is the plausibility of other concurrent events. That is, SIG effects could be invalidated if there were unobserved determinants of our outcome measures that varied both contemporaneously with the onset of SIG interventions *and* uniquely occurred in SIG schools. One such plausible event would be the Quality Teacher and Education Act (QTEA) in June 2008, which authorized SFUSD to collect \$198 per parcel of taxable property annually for 20 years to fund a general increase in teacher salaries and support for school improvement initiatives. A vast majority of the funds were applied to all schools, except for 5% of the funds, which were used to provide \$2,000 for teachers working in designated hard-to-staff schools. Although hard-to-staff schools under QTEA change over time, some of them also receive SIG awards. If there were no systematic differences in changes in student outcomes between SIG schools and other QTEA hard-to-staff schools, we would suspect that the observed SIG effects might be part of the QTEA effects, rather than due to SIG interventions. We compare SIG with non-SIG QTEA schools using Equation 1 and include the results in Panel A of Tables B9 and B10 in the online appendix. Alternatively, in Panel B of both tables, we exclude all

QTEA schools from the analysis and compare non-QTEA SIG schools with non-QTEA non-SIG schools. In both analyses, SIG schools experienced larger math and ELA achievement gains than non-SIG schools, greater reductions in excused and unexcused absences, and increased popularity among parents. These findings provide evidence that QTEA is not a major threat to the inferences of identified SIG effects.

A final alternative explanation for the positive trend would be that the gains made by SIG schools were largely due to mean reversion. *Mean reversion* describes the phenomenon that the lowest achieving schools were likely to experience larger than average gains in subsequent years (Figlio & Rouse, 2006). Were this true, the large test-score gains in SIG schools would not be the result of SIG interventions but rather would have occurred anyway because the lowest performing schools are likely to improve. To test this possibility, we created 10 pseudo-SIG schools using schools' average proficiency in both math and ELA from 2005 to 2007—the 3 years before the school performance data were used for identifying the actual SIG eligible schools. These 10 pseudo-SIG schools were the lowest performing schools during that time interval and did not have a net gain of 50 points or more on Academic Progress Index (API) scores from 2004 to 2007, nor did they meet the statewide goals of 800 API in 2006–2007. In other words, the identification of pseudo-SIG schools mimics the criteria used to identify SIG-eligible schools. To mimic the main analysis for the actual SIG schools, we created "pseudo" "stayers" and "all starters" samples.

If mean-reversion errors explain the test score gains following the SIG reform, then one should also observe such an increase for pseudo-SIG schools from 2008 to 2010—the 3 pseudo years of intervention. The results are presented in Panel B of Tables B11 (achievement) and B12 (absences and school choices) in the online appendix. The estimated pseudo-SIG effects are either in the opposite direction of the actual estimates of SIG effects or statistically insignificant. This suggests that mean reversion is not the explanation for the identified gains in student achievement and desirability or the reduction in unexcused absences in actual SIG schools during SIG reform years.

We tested the degree to which our estimated SIG effects are robust to several other mechanisms and model specifications. In our sample, 6.6% of students repeated a grade. Controlling for grade repeaters did not change our results (Table B14 in the online appendix). Another alternate specification uses student fixed effects instead of controlling for students' prior characteristics and achievement (Table B15), and results are generally consistent with our main estimates in Table 2. Last, we conducted the analyses by instrumenting the actual number of years enrolled in SIG schools using an intent-to-treat definition—the number of years students should be expected to be in SIG schools based on students' initial enrollment in 2010 fall as the instrument. The estimated SIG effects, as shown in Table B17, remain positive and significant.

Discussion and Policy Implications

School Improvement Grants highlight a national focus on improving lowest-performing schools through competitive incentives and highly prescriptive school reform frameworks. Since the SIG program began in 2009, more than 1,500 schools across the country have undertaken one of four interventions that require schools to institute specific changes aimed at raising student outcomes. Although using comprehensive strategies to transform persistently lowest performing schools is not new, the scope of the SIG program and its highly prescriptive models distinguish it from earlier reforms. Rigorous evidence on SIG impacts on a variety of student outcomes and potential mechanisms of change can shed light on the next wave of school improvement efforts under the Every Student Succeeds Act (ESSA), which continues to use research-based evidence and dramatic strategies to transform low-performing schools.

This study provides new evidence based on unique longitudinal data from SFUSD and includes a richer set of measures on both outcomes and mechanisms than those examined by earlier evaluations of SIGs and comprehensive school reforms, more generally. We find that SIG reforms in SFUSD resulted in gradual improvements in the first 2 years, and significant positive changes on several measures of school performance by the third year of the grants. Specifically, SIG reforms narrowed the achievement gap between these lowest performing schools and the rest of the schools in the district from 0.80 *SD* in spring 2010 (right before the reform) to 0.50 *SD* in the third year of SIG. Equally important, SIG reforms reduced the odds of unexcused absences by 24% in Year 2 and improved school desirability among families, indicated by an increase in the odds of being families' first choice by 117% in Year 3 relative to pre-SIG years. These positive effects that emerge during the course of the intervention are robust to a variety of alternative explanations, such as student attrition, concurrent policies, and mean reversion.

Several findings are consistent with prior studies on the SIG program and comprehensive school reform. The positive effects on student achievement mirror Dee (2012) and Papay's (2015) findings from other evaluations of SIG reforms. The larger positive effects in the third year relative to the first year echo earlier findings that comprehensive programs take time to yield impact (G. D. Borman et al., 2003; Bryk et al., 2010; de la Torre et al., 2013). We find some evidence that the impacts in turnaround schools were more pronounced than those in transformation schools on raising year-to-year achievement and increasing popularity among families. Dee (2012) provides similar evidence of greater achievement gains in turnaround than transformation schools across California, and Dragoset et al. (2017) identify a more pronounced improvement in turnaround schools in secondary grades. Moreover, Ahn and Vigdor (2014) find larger improvements among schools that underwent similar restructuring processes under NCLB. Recent work by Strunk and colleagues shows that the use of dramatic turnaround methods (reconstitution and restart), as opposed to softer reform methods (transformation), produced larger positive improvement on student achievement (Strunk, Marsh, Hashim, Bush-Mecenas, & Weinstein, 2016).

Our study adds new evidence to the literature on SIGs and whole-school reform. Ours is one of the few large-scale studies to examine the SIG effects as they unfold during all 3 years of reform. Different from prior studies that included schools that adapt reform strategies with a wide range of rigor (Dragoset et al., 2017), the SFUSD's reform plan was closely based on Bryk and colleagues' (2010) research-based guidelines for successful school improvement. The overall positive findings in this study illustrate a case that uses rigorous evidence to inform the development of a theory of change and its implementation. Moreover, we employ multiple outcome measures that examine specific elements of SFUSD's theory of action. Besides documenting increased student achievement, reduced unexcused absences, and increased popularity among parents, we provide evidence on educator capacity building. SIG schools became more likely to retain more effective teachers and improved teacher-reported professional support. The evidence on the multidimensional SIG reforms in SFUSD shows that comprehensive school transformation can succeed in a complex system.

Accompanied with the novelty of this study, it has several caveats. The SIG interventions include two major components: evidence-based interventions and substantial financial investment. Our data cannot disentangle the program effect from the financial effect. Although it is desirable to know which components of the SIG interventions are most likely to contribute to the positive outcomes, we cannot separate the unique contribution of each component, because given the nature of the whole-school reform, all components are mingled together and implemented concurrently. Additionally, our data are drawn from one school district. Although it is demographically heterogeneous, it may represent a unique case where schools carried out a successful implementation of an evidence-based reform. We cannot be certain about the generalizability of these positive impacts on SIG programs elsewhere. However, it is worth emphasizing that this successful case sheds light on the promises of transforming persistently low-achieving schools and closing achievement gaps between schools.

Despite the caveats, the findings of this study have timely policy implications. ESSA continues to prioritize turning around persistently low-performing schools on the nation's education reform agenda. As opposed to interventions driven by federal mandate, ESSA gives states and districts much more flexibility in which actions they take to support struggling schools (Sun, Saultz, & Ye, 2016). It is then all the more important to provide states and districts with guidance for choosing and implementing effective reforms. The positive impacts of SIG reform in SFUSD add growing evidence in support of school transformation guided by evidence-based frameworks. Last, because comprehensive school reforms take time to implement, an important design feature to underscore is the gradual emergence and intensification of reform impacts, suggesting that such efforts should be given time to come to fruition.

Notes

We thank San Francisco Unified School District and, particularly, Ritu Khanna, Daniel Menezes Monica Vasquez, and Laura Wentworth, for their continued support. We also acknowledge helpful comments and feedback from seminar participants of the Center for Education Policy Analysis at Stanford University and the Labor and Development Economics at the University of Washington, Seattle.

Min Sun's contribution to this study is supported by a grant from the National Science Foundation under Grant No. DRL-1506494. 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 National Science Foundation. Emily K. Penner's contribution is supported by the Institute of Education Sciences Postdoctoral Training Fellowship Award (R305B130017).

[1]Based on the Academic Performance Index (API) and proficiency rate from state standardized tests.

[2]This school closed at the end of 2011—1 year after receiving $50,000 to support a parent-community outreach coordinator to assist students in transitioning to new schools.

[3]We conduct further robustness checks to see whether SIGs have a larger or smaller impact on students who later transferred than students who stayed. If SIGs have had a significantly larger (or smaller) effect on students who transferred, the estimates from the "stayers" sample in Year 2 and 3 would have been underestimated (or overestimated). To understand the impact of excluding students who transferred, similar to a conventional difference-in-differences-in-differences (DDD) framework, we included interaction terms (e.g., *treatment year 1 indicator*SIG school indicator*indicator of whether the student later transferred out; treatment year 2 indicator*SIG school indicator*indicator of whether the student later transferred out; treatment year 3 indicator*SIG school indicator*indicator of whether the student later transferred out*) to Equation 1, as well as relevant two-way interactions. Results are included in online appendix OB-18. While there seems to be some evidence of differential effects in Model 1, after controlling for students' characteristics and performance prior to the SIG reform in Model 2 and Model 3, we do not see any differential effects of SIGs between students who stayed and those who later transferred. The SIG effect estimates remain positive and significant for math in Year 3.

[4]The choice of comparison group may also influence the SIG effect estimates. To examine the degree to which our results are sensitive to the choice of comparison groups, we used the state criteria for defining SIG-eligible schools to choose an alternative plausible comparison group. The state of California published a list of all eligible schools. For each school, the state also published eligibility measures, including their Academic Progress Index (API) scores in the prior 3 years, graduation rates in each of the prior 5 years, tier level, and so on. Using the list and data on the state-identified eligible schools, we constructed one plausible comparison group that includes 19 SFUSD schools that were similar to these nine SIG awardee schools in terms of API or graduation rates. As indicated in online appendix Table B13, the results are very much consistent with the SIG effect estimates in the main analysis in that SIG effects are largely positive and are particularly large in Year 3 of the intervention.

[5]We also specified comparative interrupted time series models to estimate both level and trajectory changes. The results in online Table B5 indicate positive level change in SIG schools in some model specifications and consistently positive effects on the slope (e.g., trajectory) change, although only a few are statistically significant. Our data do not have enough power to simultaneously estimate several school-level treatment parameters.

[6]Starting in Grade 8, students in the same grade started to take different math courses and math examinations, such as Algebra I, Algebra II, or Geometry. To account for this, we include dummy indicators for the types of math examinations that a student took in Equation 1. When modeling math achievement in Grade 8 and above, we control for students' prior test scores in seventh grade when all students took the same examination. Coefficient estimates are consistent with those calculated from Equation 1 at two decimal places, and statistical inferences are the same. Results are available upon request from the authors. In ELA, in contrast, students take a grade-specific examination in all tested grade levels, regardless of course content.

[7]Cluster-robust standard errors account for correlations among observations within clusters and are more conservative than ordinary least squares (OLS) standard errors when fewer clusters are present (Cameron & Miller, 2015, p. 340). We also used cluster bootstrapping with 400 replications, as recommended by Cameron and Miller when clustered fixed effects are included (Cameron & Miller, 2015, p. 331). These two procedures yield similar standard errors in almost all of our model specifications of student achievement. We include bootstrapped estimates in the online appendix Tables B6 and B7.

[8]We used both Poisson and negative binomial models to fit the count data—days of student absences. After comparing the model fit, we concluded that the negative binomial models fit the count data better because negative binomial models account for the overdispersion of the data.

[9]During the school years we examined, teachers used a paper Scantron to mark a student as absent or present in each class. For an absent student, a clerk in the school office would mark the student as excused absent if the clerk received a phone call from a parent or guardian providing reasons for absence; otherwise, the student was identified as unexcused absent for that class. According to our interviews with several administrators in the district, attendance records may bias toward presence due to the funding of Average Daily Attendance (ADA), but this measurement error is on our dependent variable and should not bias our results.

[10]We also model changes among families' top five choices. The estimates of SIG effects on the top five choices are generally similar to those on the first-choice schools. Results are available upon request from the authors.

[11]We also estimate our models using the last 3 years of value-added, the last 4 years, and as many years of prior value-added measures available for a teacher. The results are qualitatively consistent with those presented below. Results are available upon request from the authors. Our primary analyses use the three most recent years of data, because we expect principals or district leaders to rely most heavily on recent data to make staffing decisions.

[12]Across each survey year, well over 1,200 teachers responded, with response rates ranging from 36% to 54%. Notably, average 4-year response rates in SIG schools and non-SIG schools were very similar, at 39.4% and 39.3%, respectively, suggesting that differences in survey participation would not drive differences in teachers' average responses.

[13]Exploratory factor analysis results indicated one underlying factor (eigenvalue ≥ 2.18) of teacher support and Cronbach's α = 0.73–0.76 of all the items across all 4 years.

[14]We included a full set of year fixed effects and their interactions with SIG treatment, which accounts for possible differential pre-SIG trends between SIG and non-SIG schools. The estimates of SIG effects in this alternative model specification, as shown in online appendix Table B8, are very consistent with our main model specifications.

References

Ahn, T., & Vigdor, J. (2014). *The impact of No Child Left Behind's accountability sanctions on school performance: Regression discontinuity evidence from North Carolina* (No. 20511). National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w20511

Aladjem, D. K., LeFloch, K. C., Zhang, Y., Kurki, A., Boyle, A., Taylor, J. E., . . . Fashola, O. (2006). *Models matter—The final report of the National*

Longitudinal Evaluation of Comprehensive School Reform (ERIC Document Reproduction Service No. ED499198). Washington, DC: American Institutes for Research.

Barrett, N., & Harris, D. (2015). *Significant changes in the New Orleans teacher workforce*. New Orleans, LA: Education Research Alliance for New Orleans. Retrieved from http://educationresearchalliancenola.org/files/publications/ERA-Policy-Brief-Changes-in-the-New-Orleans-Teacher-Workforce.pdf

Berends, M. (2000). Teacher-reported effects of New American School designs: Exploring relationships to teacher background and school context. *Educational Evaluation and Policy Analysis*, *22*(1), 65–82.

Berends, M., Bodilly, S. J., & Kirby, S. N. (2002). *Facing the challenges of whole-school reform: New American Schools after a decade*. Santa Monica, CA: Rand Corporation.

Bifulco, R., Duncombe, W., & Yinger, J. (2005). Does whole-school reform boost student performance? The case of New York City. *Journal of Policy Analysis and Management*, *24*(1), 47–72.

Bloom, H. S., Ham, S., Melton, L., & O'Brien, J. (2001). *Evaluating the accelerated schools approach: A look at early implementation and impacts on student achievement in eight elementary schools*. Manpower Demonstration Research Corporation. Retrieved from http://www.mdrc.org/sites/default/files/full_94.pdf

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.

Borman, K. M., Carter, K., Aladjem, D. K., & LeFloch, K. C. (2004). Challenges for the future of comprehensive school reform. In C. T. Cross (Ed.), *Putting the pieces together: Lessons from comprehensive school reform research* (pp. 109–150). Washington, DC: National Clearinghouse for Comprehensive School Reform.

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

Burgess, S., Greaves, E., Vignoles, A., & Wilson, D. (2014). What parents want: school preferences and school choice. *The Economic Journal*, *125*(587), 1262–1289.

California Department of Education. (2010, March 11). *Persistently lowest-achieving schools—Accountability*. Retrieved from http://www.cde.ca.gov/ta/ac/pl/

California Department of Education. (2015, September 21). *Largest & smallest public school districts—CalEdFacts*. Retrieved from http://www.cde.ca.gov/ds/sd/cb/ceflargesmalldist.asp

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

Coleman, J. S., Campbell, E. Q., Hobson, C. F., McPartland, J. M., Mood, A. M., Weinfeld, F. D., & York, R. L. (1966). *Equal educational opportunity*. Washington, DC: U.S. Department of Health, Education, and Welfare, Office of Education.

Cook, T. D., Habib, F.-N., Phillips, M., Settersten, R. A., Shagle, S. C., & Degirmencioglu, S. M. (1999). Comer's school development program in Prince George's County, Maryland: A theory-based evaluation. *American Educational Research Journal*, *36*(3), 543–597.

Council of the Great City Schools. (2015). *School Improvement Grants: Progress report from America's Great City Schools*. Retrieved from http://www.cgcs.org/cms/lib/DC00001581/Centricity/Domain/87/SIG%20Report%202015.pdf de la Torre, M., Allensworth, E., Jagesic, S., Sebastian, J., Salmonowicz, M., Meyers, C., & Gerdeman, R. D. (2013). *Turning around low-performing schools in Chicago: Research report*. Chicago, IL: The University of Chicago Consortium on Chicago School Research.

Dee, T. (2012). *School turnarounds: Evidence from the 2009 stimulus* (Working Paper No. 17990). National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w17990

Desimone, L. (2002). How can comprehensive school reform models be successfully implemented? *Review of Educational Research*, *72*(3), 433–479.

Dickey-Griffith, D. (2013). Preliminary effects of the school improvement grant program on student achievement in Texas. *The Georgetown Public Policy Review*, 21–39. Retrieved from http://gppreview.com/wp-content/uploads/2014/02/Dickey-Griffith-D.pdf

Dragoset, L., Thomas, J., Herrmann, M., Deke, J., James-Burdumy, S., Graczewski, C., . . . Giffin, J. (2017). *School Improvement Grants: Implementation and effectiveness* (NCEE 2017-4012). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Figlio, D. N., & Rouse, C. E. (2006). Do accountability and voucher threats improve low-performing schools? *Journal of Public Economics*, *90*(1), 239–255.

Gross, B., Booker, T. K., & Goldhaber, D. (2009). Boosting student achievement: The effect of comprehensive school reform on student achievement. *Educational Evaluation and Policy Analysis*, *31*(2), 111–126.

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

Harris, D. N., & Larsen, M. F. (2016). *The effects of the New Orleans post-Katrina school reforms on student academic outcomes*. Education Research Alliance for New Orleans. Retrieved from http://educationresearchalliancenola.org/files/publications/The-Effects-of-the-New-Orleans-Post-Katrina-School-Reforms-on-Student-Academic-Outcomes.pdf

Hastings, J. S., Kane, T. J., & Staiger, D. O. (2005). *Parental preferences and school competition: Evidence from a public school choice program* (Working Paper 11805). National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w11805

Hastings, J. S., & Weinstein, J. M. (2008). Information, school choice, and academic achievement: Evidence from two experiments. *The Quarterly Journal of Economics*, *123*(4), 1373–1414.

Heissel, J. A., & Ladd, H. F. (2016). *School turnaround in North Carolina: A regression discontinuity analysis* (Working Paper No. 156). Washington, DC: CALDER.

Herman, R., Dawson, P., Dee, T., Greene, J., Maynard, R., Redding, S., & Darwin, M. (2008). *Turning around chronically low-performing schools. IES Practice Guide* (NCEE 2008-4020). Washington, DC: National Center for Education Evaluation and Regional Assistance. Retrieved from https://ies.ed.gov/ncee/wwc/Docs/PracticeGuide/Turnaround_pg_04181.pdf

Hess, G. A. (1999). Expectations, opportunity, capacity, and will: The four essential components of Chicago school reform. *Educational Policy*, *13*(4), 494–517.

Hill, P. T. (2006). *Put learning first: A portfolio approach to public schools* (PPI Policy Report). Washington, DC: Progressive Policy Institute. Retrieved from http://www.crpe.org/sites/default/files/pub_portfolio_putlearningfirst_feb06.pdf

Jackson, C. K. (2012). *Non-cognitive ability, test scores, and teacher quality: Evidence from ninth grade teachers in North Carolina* (Working Paper No. 18624). Cambridge, MA: National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w18624

Kantor, H., & Lowe, R. (1995). Class, race, and the emergence of federal education policy: From the New Deal to the Great Society. *Educational Researcher*, *24*(3), 4–21.

Lachlan-Haché, J., Naik, M., & Casserly, M. (2012). *The School Improvement Grant rollout in America's Great City Schools: School Improvement Grants*. Council of the Great City Schools. Retrieved from http://eric.ed.gov/?id=ED530694

Loeb, S., & Candelaria, C. (2012). *How stable are value-added estimates across years, subjects, and student groups?* Carnegie Knowledge Network. Retrieved from http://www.carnegieknowledgenetwork.org/briefs/value-added/value-added-stability/

Malen, B., & Rice, J. K. (in press). School reconstitution as a turnaround strategy: An analysis of the evidence. In J. Mathis & T. Trujillo (Eds.), *Test-based education reforms: Lessons from a failed agenda*. Greenwich, CT: Information Age.

Malen, B., & Rice, J. K. (2004). A framework for assessing the impact of education reforms on school capacity: Insights from studies of high-stakes accountability initiatives. *Educational Policy*, *18*(5), 631–660.

McEachin, A. J., Welsh, R. O., & Brewer, D. J. (2016). The variation in student achievement and behavior within a portfolio management model: Early results from New Orleans. *Educational Evaluation and Policy Analysis*, *38*(4), 162373 716659928.

Mintrop, H., & Trujillo, T. (2005). Corrective action in low performing schools: Lessons for NCLB implementation from first-generation accountability systems. *Education Policy Analysis Archives*, *13*(48).

National Center for Education Statistics. (2015). The *NCES Fast Facts Tool provides quick answers to many education questions*. Retrieved from https://nces.ed.gov/fastfacts/display.asp?id=55

National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform: An open letter to the American people: A report to the nation and the Secretary of Education*. Washington, DC: National Commission on Excellence in Education.

Norton, R. (2010). *SIG Improvement Grant (SIG) & SFUSD options*. Retrieved from http://slidegur.com/doc/3160328/sfusd—rachel-norton–sf-board-of-education—a-public

Papay, J. P. (2015, November). *The effects of school turnaround strategies in Massachusetts*. Paper presented at the Association of Public Policy and Management Annual Meeting, Miami, FL.

Player, D., & Katz, V. (2013). *School improvement in Ohio: An evaluation of the School Turnaround Specialist Program* (CEPWC Working Paper Series No. 10). Charlottesville, VA: Center on Education Policy and Workforce Competitiveness. Retrieved from http://curry.virginia.edu/uploads/resourceLibrary/10_Player_SchoolTurnaround.pdf

Rowan, B., & Miller, R. J. (2007). Organizational strategies for promoting instructional change: Implementation dynamics in schools working with comprehensive school reform providers. *American Educational Research Journal*, *44*(2), 252–297.

San Francisco Unified School District (SFUSD). (2010). *SFUSD School Improvement Grant (SIG) application executive summary*. San Francisco, CA: SFUSD. Retrieved from www.nctq.org/docs/sig-executive-summary.pdf

San Francisco Unified School District (SFUSD). (2012). *SFUSD: SF students show academic achievement gains* (press release). San Francisco, CA: SFUSD. Retrieved from http://www.sfusd.edu/en/news/current-news/2012-news-archive/08/sf-students-show-academic-achievement-gains.html

San Francisco Unified School District (SFUSD). (2015). *SFUSD facts at a glance 2015*. San Francisco, CA: SFUSD. Retrieved from http://www.sfusd.edu/en/assets/sfusd-staff/about-SFUSD/files/SFUSD%20Facts%20at%20a%20Glance%20-%20June%202015.pdf

San Francisco Unified School District (SFUSD). (2016). *SFUSD: Our district in a snapshot*. San Francisco, CA: SFUSD. Retrieved from http://www.sfusd.edu/en/employment/certificated-careers/teaching-careers/why-teach-with-sfusd/our-district-snapshot.html

Scott, C., Krasnoff, B., Davis, D., & Northwest, E. (2014, October). *Digging into transformation: Implementation of federal school improvement grants in Oregon*. Paper presented at the Annual Meeting of the American Evaluation Research Association, Denver, CO.

Scott, C., & McMurrer, J. (2015). Mission impossible? What states with large percentages of rural schools tell us about federal school improvement grants. *Peabody Journal of Education*, *90*(2), 211–223.

Spillane, J. P., & Thompson, C. L. (1997). Reconstructing conceptions of local capacity: The local education agency's capacity for ambitious instructional reform. *Educational Evaluation and Policy Analysis*, *19*(2), 185–203.

Strunk, K. O., Marsh, J. A., Bush-Mecenas, S. C., & Duque, M. R. (2015). The best laid plans: An examination of school plan quality and implementation in a school improvement initiative. *Educational Administration Quarterly*, *52*(2), 259–309.

Strunk, K. O., Marsh, J. A., Hashim, A. K., & Bush-Mecenas, S. (2016). Innovation and a return to the status quo: A mixed-methods study of school reconstitution. *Educational Evaluation and Policy Analysis*, *38*(3), 549–577.

Strunk, K. O., Marsh, J. A., Hashim, A. K., Bush-Mecenas, 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*(3), 251–282.

Sun, M., Saultz, A., & Ye, Y. (2016). Federal policy and the teacher labor market: Exploring the effects of NCLB on teacher turnover. *School Effectiveness and School Improvement*, *28*(1), 102–122.

Sunderman, G. L. (2001). Accountability mandates and the implementation of Title I schoolwide programs: A comparison of three urban districts. *Educational Administration Quarterly*, *37*(4), 503–532.

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

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

Wang, M. C., Wong, K. K., & Kim, J.-R. (1999). *A national study of Title I schoolwide programs: A synopsis of interim findings*. Laboratory for Student Success, the

Mid-Atlantic Regional Educational Laboratory at Temple University, Center for Research in Human Development and Education. Retrieved from https://archive.org/stream/ERIC_ED436596#page/n5/mode/2up

Welsh, R. O., Duque, M., & McEachin, A. (2016). School choice, student mobility, and school quality: Evidence from post-Katrina New Orleans. *Education Finance and Policy*, *11*(2), 150–176.

Wentworth, L., Khanna, R., & Piper, R. (2016). Vision to action: supporting the most underserved schools. *The Leadership Magazine*, pp. 34–37. Retrieved from https://www.joomag.com/magazine/leadership-magazine-jan-feb-2016-v45-no-3/0067286001452036965?page=34

Wong, K. K., & Meyer, S. J. (1998). Schoolwide programs: A synthesis of findings from recent evaluation. *Educational Evaluation and Policy Analysis*, *20*(2), 115–136.

Zimmer, R., Henry, G., & Kho, A. (2016, November). *The effects of school turnaround in Tennessee's achievement school district and innovation zones*. Paper presented at the annual meeting of the Association for Public Policy Analysis and Management, Washington, DC.

Manuscript received April 20, 2016
Final revision received January 29, 2017
Accepted January 30, 2017