University of Massachusetts Amherst

ScholarWorks@UMass Amherst

2019

The effects of school reform under NCLB waivers: Evidence from focus schools in Kentucky.

Sade Bonilla

Thomas S. Dee

The Effects of School Reform Under NCLB Waivers:
Evidence from Focus Schools in Kentucky

Sade Bonilla
Thomas S. Dee
Stanford University

May 2017

Abstract - Under waivers to the No Child Left Behind (NCLB) Act, the federal government required states to identify schools where targeted subgroups of students have the lowest achievement and to implement reforms in these "Focus Schools." In this study, we examine the Focus School reforms in the state of Kentucky. The reforms in this state are uniquely interesting for several reasons. One is that the state developed unusually explicit guidance for Focus Schools centered on a comprehensive school-planning process. Second, the state identified Focus Schools using a "super subgroup" measure that combined traditionally low-performing subgroups into an umbrella group. This design feature may have catalyzed broader whole-school reforms and attenuated the incentives to target reform efforts narrowly. Using regression discontinuity designs, we find that these reforms led to substantial improvements in school performance, raising math achievement by 17 percent and reading achievement by 9 percent.

1. Introduction

In the United States, the last 15 years have marked an era of exceptional federal activism focused on the performance of public schools, particularly with regard to historically underserved students. This period began with the No Child Left Behind Act (NCLB), which brought test-based school accountability to scale in the U.S. and highlighted stark and enduring achievement gaps by race, ethnicity, and income. NCLB also set a 2014 deadline for all students to achieve proficiency. The broad expectation was that NCLB would be redesigned and reauthorized in advance of that deadline. However, because partisan gridlock in Congress stalled this reauthorization, U.S. Secretary of Education Arne Duncan articulated, in 2011, conditions under which states could receive NCLB waivers. These waivers reflected the Obama administration's vision for the reauthorization of NCLB but proved controversial, generating pushback from both parties in Congress, as well as teachers' unions and state education agencies (Klein 2015). The goals of these waivers included increased flexibility in school-accountability systems, a focus on college and career-ready standards and new systems for evaluating the effectiveness of teachers and principals.

However, a defining feature of these waiver-driven reforms was also the emphasis on "differentiated accountability" for schools. In particular, states that received waivers were required to categorize the 10 percent of their schools with the largest achievement gaps as "Focus Schools" and the 5 percent of schools with persistently low achievement as "Priority Schools." Both reforms were effectively unfunded but otherwise differed in interesting ways. Like schools that had recently received federal School Improvement Grants, Priority Schools were required to implement a school-turnaround model consistent with federal guidance.[1] In sharp contrast, Focus Schools were given substantial *flexibility* as to how they identified and implemented reforms intended to reduce their achievement gaps. Understanding the effects of these federally catalyzed reforms is particularly salient as these approaches to school

---

[1] See Dee (2012) for an overview of federal turnaround of these models (i.e., transformation, turnaround, restart, and closure) and their features (e.g. hiring a new principal, replacing a large portion of teaching staff, implementing instructional reforms, and social-emotional supports) and their effects.

accountability may continue under the recent reauthorization of NCLB known as the "Every Student Succeeds Act" (ESSA) (Klein 2016).

In this study, we examine the implementation and impact of Focus School reforms in the state of Kentucky.[2] Kentucky is a particularly interesting setting in which to examine the effects of these reforms for two broad reasons. One is that, during this period, Kentucky established a reputation for their early adoption and high-fidelity implementation of new reforms (e.g., the Common Core State Standards, Race to the Top). Therefore, they are likely to constitute a strong test of whether these federally motivated reforms can drive meaningful improvements in school-level outcomes. Interestingly, Kentucky chose a relatively prescriptive approach to its Focus Schools that included an emphasis on school-improvement planning and teacher professional development. Second, along with a few other states, Kentucky implemented an interesting and fairly unique design feature. They identified Focus Schools using the performance of a "super subgroup" of traditionally low-performing subgroups (i.e., free/reduced price lunch, students with disabilities, black, Hispanic, American Indian, and limited English proficient) for two reasons. Firstly, this approach increased the number of smaller rural schools held accountable under the new system.[3] Secondly, this inclusive design ensured that small subgroups that might not meet size requirements within a school would receive attention. This may be relevant given that accountability systems that target narrow subgroups can induce schools to engage in strategic behaviors like targeting efforts among specific populations of students (Figlio and Loeb 2011). Kentucky's unique omnibus approach to identifying Focus Schools could sharply attenuate the incentives for such strategic behaviors.

To credibly identify the causal effects of these reforms, we rely on regression discontinuity (RD) designs based on the discontinuous assignment of schools to Focus School status. Our RD results indicate that Kentucky's Focus School reforms led to substantial gains in

---

[2] We do not examine the effects of the Priority School reforms because the small number of Priority Schools (and their overlap with earlier SIG reforms) implies attenuated statistical power.

[3] In Kentucky, subgroups with at least 25 students count for accountability purposes. By combining traditionally low-performing subgroups into an umbrella group, more schools receive an accountability rating.

students' math and reading proficiency rates (i.e., a 17 percent improvement in mathematics proficiency and 9 percent improvement in reading proficiency). Furthermore, we find evidence that these gains occurred throughout the distributions of math and reading achievement (i.e., not just proximate to the proficiency threshold). We also find that these results are robust to a variety of checks for internal validity including possible confounds related to the non-random, post-reform sorting of students. We also present evidence suggesting that comprehensive school planning and effective teacher professional development were important mediators of the Focus School reform's impact. We organize our study as follows. In section 2, we discuss the policy context in Kentucky and provide a brief review of the relevant literature and theoretical considerations. In section 3, we characterize the treatment contrast for Focus Schools. In section 4, we describe the data and the construction of our analytic samples. In section 5, we present our RD estimation strategy and related robustness checks. We discuss our results in section 6 and conclude in section 7.

2. Background and Prior Literature

The Focus School reforms examined in this study were implemented under the recent federal waivers to the No Child Left Behind (NCLB) Act. NCLB was the influential and controversial federal legislation that brought test-based school accountability to scale in the United States and required all students to be proficient by 2014. Anticipating that states would not be able to meet this proficiency requirement, the U.S. Department of Education began offering waivers in 2011 to states that developed more flexible systems of school accountability consistent with new guidelines. Among other things, these waivers mandated that states identify and target reforms in schools that contributed the most to achievement gaps (i.e., "Focus Schools").[4] Learning from these early Focus School reforms is highly relevant as they will continue nationwide under NCLB's recent reauthorization as the "Every Student Succeeds Act" (ESSA). This current policy emphasis on a more targeted "differentiated" approach to school

---

[4] We describe the federal NCLB waivers and Focus School reforms in more detail in Section 3.

accountability represents the most recent iteration of a quarter-century of school accountability policies. In general, these school accountability reforms articulated consequential sanctions and supports (e.g., labeling, staff or school reconstitution, technical assistance) for schools that failed to meet test-based performance goals.

The implied theoretical motivations for school accountability situate the chronic underperformance of some schools with respect to information asymmetries and the collective-action challenges of organizational effectiveness. More specifically, the logic of these reforms is that the public-sector agents in districts and schools may lack knowledge of their performance (e.g., the existence and character of achievement gaps) or a willingness to focus on them instead of other goals without more oversight from key stakeholders (e.g., parents and tax payers). Also, the external incentives created by school accountability may also encourage schools to overcome the collection-action challenges that can inhibit school effectiveness (e.g., school culture, cooperative management, and instructional practices).

The empirical literature on school accountability generally suggests that these reforms generated meaningful, though not transformational, improvements in student learning as measured by low-stakes tests unrelated to the accountability incentives (National Research Council, 2011, Figlio and Loeb 2011). This quasi-experimental literature includes studies of the state-level reforms that preceded NCLB (Carnoy & Loeb, 2002; Hanushek & Raymond, 2005) as well as studies of NCLB's impact (Dee and Jacob 2011; Wong, Steiner, and Cook 2013). There is also some evidence that schools and teachers may respond strategically in certain ways to the presence of accountability incentives (Figlio and Loeb 2011). For example, Dee, Jacob, and Schwartz (2012) find evidence that NCLB led to a modest reallocation of instructional time from non-tested to tested subjects.

The Focus School reforms are also closely related to "whole-school" or comprehensive school reforms (CSR) that call for the simultaneous and multi-faceted overhaul of an entire school to refocus efforts to raise student-achievement (Borman et al., 2003). Between 1998 and 2005, the U.S. Department of Education disseminated $1.8 billion in grants for low-performing

schools to participate in a CSR program. The federal CSR program articulated eleven broad reform strategies required of participating schools. These included a focus on strong academic content, articulating measurable goals, professional development, attention to the allocation of resources, and outside technical assistance. Studies of the federal CSR initiative found little evidence that they were successful in raising student achievement (U.S. Department of Education, 2010; Gross, Booker, and Goldhaber 2009). Other evidence suggests that the school-level implementation of the federal CSR model was uneven in ways that mattered for sustaining school improvement (Desimone 2002, Bifulco, Duncombe, and Yinger 2005, U.S. Department of Education 2010). A meta-analysis by Borman et al. 2003 concludes that, in contrast to the federal CSR effort, other specific CSR models (i.e., Direct Instruction, Success for All, and the School Development Program) appear to have had meaningful impacts on student achievement, particularly when implemented over several years.

These prior studies have several implications for the Focus School reforms studied here. First, the prior evidence suggests some optimism that the latest incarnation of consequential school accountability (i.e., targeting particular schools for improvement and supports) may be effective in improving student achievement. Second, this literature also implies that attention to the potential unintended consequences of such reforms is warranted. Third, the evidence from successful CSR models suggests that Focus School reforms may be more effective to the extent that they employ similar strategies such as formative assessment and data-driven instruction (e.g., Success for All), school-wide planning and community engagement (e.g., the School Development Program), and differentiated instruction (e.g., Direct Instruction). Finally, the experience with the federal CSR program implies that the efficacy of Focus School reforms will turn critically on the quality of the implementation at the state and local level. In the next section, we turn to such evidence by characterizing Kentucky's Focus School reforms as well as the evidence on their implementation.

3. Focus Schools in Kentucky

For the past 25 years, Kentucky has situated itself as a consistently early and enthusiastic adopter of education reforms. For example, in 1990, the Kentucky Education Reform Act (KERA) reformed the state's education finance system, increased school accountability, and introduced new educational standards and assessments. More recently, the state was the first to adopt the Common Core State Standards (CCSS), doing so before the standards were complete (Governor's Task Force, 2011; Ujifusa, 2013). Kentucky was also viewed as a leader in supporting the classroom implementation of the CCSS (e.g., Gewertz 2011). The state received $17 million in the federal Race to the Top (RttT) competition and was among the first group of states to apply for a federal waiver from NCLB. Furthermore, the U.S. Department of Education later invited Kentucky to apply for a "fast-track" waiver renewal of its NCLB waiver (Klein, 2015). Kentucky's recent record of adopting and implementing leading K-12 policy reforms suggests that it is a particularly propitious setting in which to study the effects of the school reforms required under NCLB waivers.

3.1 NCLB Waivers

The waiver process had its genesis in both the design of NCLB and federal politics. NCLB required that all students be deemed proficient in their state accountability system by 2014. The widespread anticipation that this ambitious goal would not be met motivated extensive discussion of reauthorizing NCLB with updated requirements. However, the legislative failure to advance any such legislation created the opportunity for the U.S. Department of Education to offer states waivers from these requirements. In 2011, under Secretary Arne Duncan, states were encouraged to apply for such waivers by outlining new accountability systems consistent with the reform principles favored by the Obama administration. Central among these was the encouragement to adopt "differentiated recognition, accountability, and support systems." States

receiving waivers were required to identify schools that were persistently lowest performing schools (i.e., Priority Schools) and have them implement federally prescribed reforms.[5]

They were also required to designate as "Focus Schools" those that had the lowest performance for designated subgroups or the largest performance gaps between subgroups. A minimum of 10 percent of Title I schools in a state needed to be identified as Focus Schools. States were required to field interventions in these schools though the federal waivers allowed flexibility in contrast to the prescriptive requirements associated with Priority Schools. While Priority Schools were required to implement one of the federal School Improvement Grant (SIG) turnaround strategies, Focus Schools were asked to implement interventions that were "consistent with" federal turnaround principles or any other "research-based" interventions to meet the needs of students at the school (U.S. Department of Education 2012). We describe Kentucky's Focus School initiative in more detail below.

Kentucky submitted its application for an NCLB waiver in November of 2011. The U.S. Department of Education approved the waiver in February of 2012 and implementation was set to begin in the 2012-13 school year. However, a vendor delay slowed implementation of the identification of schools and, as a result, the Kentucky Department of Education (KDE) notified schools of their Focus School status in late October of 2012. Schools then had 90 days to develop their Comprehensive School Improvement Plan (CSIP), a central element of the waiver and Focus School reforms that we describe in more detail below. Given the delayed timeline, schools did not have a full school year to implement reforms. Therefore, in this paper we focus our analyses on outcomes from the 2013-14 school year, the first *full* year in which Focus Schools implemented their chosen reforms.

We later describe in detail Kentucky's procedure for identifying Focus Schools. However, one distinctive feature merits special attention. As a predominantly rural state, Kentucky has many small schools that often do not enroll enough students to meet the minimum subgroup

---

[5] The number of Priority Schools is required to be equal to 5% of the number of Title I schools in the state. We do not study these reforms in Kentucky because there are two few of such schools to support reasonable statistical power.

size.[6] In order to have appropriate representation of schools serving targeted students among designated Focus Schools, Kentucky introduced a "super-subgroup" measure, combining unduplicated counts of students from traditionally low-performing subgroups (i.e. free/reduced price lunch, students with disabilities, black, Hispanic, American Indian, and limited English proficient) into an umbrella group. To mitigate any concerns that the larger group would mask low individual subgroup achievement, Kentucky also used a second method for Focus School identification that considered subgroup performance of traditionally low-performing subgroups separately. A seemingly unintended but possibly beneficial feature of the "super-subgroup" measure is that it attenuates any incentives school may have to focus on a particular subgroup of students, possibly at the expense of others (i.e., "triage" behavior).

3.2 Focus School Reforms in Kentucky

Kentucky's leading transition to the CCSS also provided the state with motivation to redesign its school accountability system. In 2009, bipartisan legislation (Senate Bill 1) called for an overhaul of previous reform efforts and a unified focus on college and career-ready standards and assessments to be implemented in the 2011-12 school year. In its 2011 waiver application, the state noted the strong alignment between these in-progress reforms, the corresponding changes to their accountability system, and the principles behind NCLB waivers. Kentucky's newly designed accountability system focused on holding all schools accountable for student achievement.[7] In particular, the "central tenet" of Kentucky's accountability system (U.S. Department of Education 2011) focused on the design, revision, and monitoring of Comprehensive School Improvement Plans (CSIP). A school committee that includes representation from parents, students, and the community would begin the state-designed CSIP process with a school-needs assessment linked to data, explicit standards, and indicators for school improvement. The CSIP guides these committees to identify and prioritize their school's

---

[6] The minimum subgroup size in Kentucky is 25. A subgroup that meets the minimum subgroup threshold has their achievement factored into a school's accountability rating and reports their achievement publicly.

[7] As such, all schools - including those ineligible for Title I support - were eligible for Focus School status if they met the selection criteria.

needs. They then identify strategies and activities designed to address the high-priority needs, a person responsible for that task, a corresponding timeline, and a funding source.

The CSIP process is critical for understanding the treatment contrast created by Focus School eligibility. In particular, as articulated in Kentucky's original waiver application, designation as a Focus School implied several *additional* CSIP requirements related to gap issues and the school's plan to address them. For example, districts were required to assist their Focus Schools with their needs assessment, using guidance from the Commissioner's Raising Achievement/Closing Gaps Council (CRACGC). The CSIPs for Focus Schools also had a unique requirement to identify specific strategies for closing gaps in student achievement and graduation rates. The state also required the Focus School CSIPs to discuss curricular alignment that would ensure a rigorous instructional program aligned to student needs and the Common Core. Other unique requirements of Focus School CSIPs were the provision of time for collaborating on the use of data to inform instructional strategies, aligned professional development, and strategies for school safety and discipline. All schools were required to post their CSIPs on their web sites so that they would be available for public inspection.

The KDE also established units dedicated to supporting school capacity to implement these plans. The Office of Next Generation Schools and Districts orchestrated technical assistance to schools statewide. However, a specialized KDE unit, District 180, was specifically tasked with providing Focus and Priority Schools technical assistance in developing and implementing their CSIPs (e.g., identifying research-based interventions). The state also sought to publicize school improvement efforts. The KDE encouraged Focus Schools to submit potential best practices, which they graded according to a standard rubric and posted these vetted practices in an online database for other schools to access. The KDE gave schools with assessed best practices a $500 reward and publicized their name on the department website and in media press releases. In terms of monitoring Focus School improvement, the KDE viewed local districts as the first level of oversight while KDE staff provided additional oversight when schools were not on-track to exit Focus School status (Kentucky Department of Education 2011). Focus Schools that failed to

exit Focus School status after several years could also be redesignated as Priority Schools and compelled to undertake more explicit reforms.

Kentucky's original waiver application explicitly indicates that the CSIP requirements unique to Focus Schools are the defining characteristic of their intended intervention. However, apart from the CSIP process, there were at least two other candidate mediators that may characterize the Focus School intervention. One is simply the stigma that might be associated with being labeled a Focus School and that could spur school improvement independent of the CSIP process. Another is that Kentucky's waiver reforms provided Focus Schools with enhanced financial flexibility to combine and redirect multiple Title I funding streams. It also expanded eligibility for the Title I school-wide program (SWP) to all Focus Schools. As a Title I SWP, Focus Schools were also eligible to receive additional support from the state to implement the reforms articulated in their CSIPs.[8]

After the first full year of Focus School implementation (i.e., in August of 2014), Kentucky submitted its successful application for an extension to its NCLB waiver. This application indicates that a key treatment contrast between Focus Schools and other schools was attenuated by the 2014-15 school year. That is, the application describes a reinvigorated emphasis on supporting and monitoring CSIPs through a new review rubric that applies to *all* Focus and Title I schools. We examined this rubric and found that it contained the CSIP elements that had been required only of Focus Schools through the 2013-14 school year. Because of this extension of the key intervention elements to all Title I schools (i.e., the overwhelming majority of our analytical sample), we focus our analysis on outcomes from the 2013-14 school year (i.e., the first full year of Focus School implementation) in which the treatment contrast was crisp.[9] Nonetheless, we also report the results from the 2014-15 school year in an appendix.[10]

---

[8] As a practical matter, this funding flexibility has attenuated empirical relevance for the treatment contrast we study. Specifically, among the 187 Focus Schools in our analytical sample, only 42 obtained SWP status as a result of the waiver reforms.

[9] The treatment contrast may have also been attenuated in subsequent years by a growing awareness of the threat of Focus School status among schools above but close to the eligibility threshold.

3.3 Focus School Implementation

The activities unique to Focus Schools featured school-improvement planning with an emphasis on state-approved strategies (e.g., gap-closing programs and activities, instructional reforms, and professional development) coupled with local and state monitoring as well as financial flexibility. However, the actual implementation of these reforms may have differed from the stated policy intent. While we lack comprehensive information on the Focus School implementation in Kentucky, we look to several sources of information. These include federal monitoring reports of Kentucky's waiver activities, evidence from media accounts, our own interviews with KDE officials responsible for supporting Focus Schools, our own review of school improvement plans, and results from a statewide teacher survey.

A federal review of Kentucky's waiver reforms occurred during a week in late August 2013 (i.e., just *before* the first full year of Focus School implementation). Federal officials reviewed state documents and conducted interviews with several state officials before issuing annual monitoring reports summarizing implementation and listing next steps when implementation did not comply with the approved waiver. This report characterized Kentucky's Focus School implementation as "Not Meeting Expectations." Specifically, the federal review called for the state to increase their oversight and monitoring in concordance with their approved waiver (U.S. Department of Education 2013). Kentucky's successful 2014 application for a waiver extension indicated that the U.S. Department of Education was satisfied the state addressed the areas of concern from the monitoring report (Delisle, 2014).

We now turn to implementation evidence about the experience of Focus Schools from local media coverage that includes interviews with school principals as well as the authors' personal communications with KDE officials. Media coverage of Focus Schools in Fayette County Public Schools (FCPS), the second largest school district in the state, indicates that Focus

---

[10] These results in Appendix Table 1 suggest similarly positive effects of the Focus School status. However, the point estimates are somewhat smaller and statistically insignificant. This pattern is consistent with the relevance of the school-planning process as a mediator and the clear dampening of the treatment contrast described here. However, we acknowledge the possibility that the intervention's effects faded.

School principals were required to present their school improvement plans to the school board. District officials also assigned principals to work with a mentor, usually a retired principal, to help implement the plans (Spears, 2015). Other coverage indicates that staff and the local community were aware of the label and highly motivated to exit Focus School status. A local media outlet covered the celebration of one high school that exited Focus School status: the principal invited local media to cover the outdoor carnival-like event with activities for students and teachers (Foley, 2015).

Personal communication with state officials on technical assistance and state monitoring revealed that Focus Schools may not have received the high level of state support and oversight articulated in the approved waiver (i.e., a finding also suggested by the federal monitoring report from 2013). A KDE official disclosed to us that the state office charged with providing technical assistance primarily served Priority Schools although services were accessible to Focus Schools. We also learned that the staff at the KDE technical assistance office, District 180, interacted with Focus Schools mainly by reviewing their CSIP. Unfortunately, District 180 staff did not maintain records of their interactions with schools so it is unknown how many Focus Schools sought out technical assistance nor the intensity of such. Moreover, few schools were subject to the more extensive management audits that examined their activities and only one of these schools was a Focus School.[11] Although the state articulated an exhaustive protocol for the monitoring and oversight of Focus Schools, the onus appears to have been on local districts to provide oversight and guidance to treatment schools.

With this in mind, we examined school improvement plans to gain additional insight into the character of the Focus School activities. We conducted a review of CSIPs for all schools in the two largest school districts in the state, Jefferson County Public Schools (JCPS) and Fayette County Public Schools (FCPS) that had plans readily available for review. We obtained CSIPs

---

[11] A KDE official also revealed that four schools ultimately underwent a management audit: one Focus school, one Priority, and two additional schools with no accountability classification. For confidentiality reasons, we suppress here the names of the KDE officials with whom we spoke.

for all 107 JCPS and 47 FCPS schools in our sample for the 2013-14 school year.[12] In our review, we coded whether schools mentioned the major whole-school reforms recommended by the state (e.g. extended learning time, teacher collaboration, professional development, data-driven instructional practices, and computer-aided instruction for low-performing students) and find nearly universal compliance. However, schools were somewhat less likely to follow KDE guidance in listing specific strategies for students with disabilities and/or non-English proficient students (i.e., 61 percent). Our review of CSIP plans suggests that Focus Schools followed the recommended state guidelines in developing their plans but provides little insight into whether Focus Schools carried out their plans differently from those of non-treatment schools.

To this end, we turn to our final source of information from state surveys of teachers that provides information on the experience of teachers rather than intended activities. The Teaching, Empowering, Leading and Learning (TELL) Kentucky survey, administered in 2015, covered salient topics on teacher working conditions including professional development, use of time, instructional practices and support. The high average response rate (i.e., 92 percent) allows us to examine whether teachers at Focus Schools report differences compared to those at non-treatment sites.[13] Our Appendix Table 2 presents results from auxiliary RD regressions that examine the effects of Focus School reforms on teachers' perceptions of the quality of their professional development (i.e., 10 distinct items and an overall composite). The point estimates are uniformly positive, indicating that Focus School reforms led to broad improvements in the quality of teacher professional development. The gains were particularly large (and statistically significant) for teacher reports of whether their professional development included collaboration and follow-up as well as whether it supported teachers in meeting student needs and promoting learning.

---

[12] Nearly half of JCPS middle and elementary schools were classified as Focus Schools (n=51). Fayette County Public Schools have 17 Focus Schools in the sample. Together these two large districts host 68 Focus Schools or 36 percent of the Focus Schools in the sample.

[13] We run auxiliary regressions to check for continuity of response rate and likelihood of any responses at the Focus School assignment threshold and find no discontinuous jumps.

In summary, the available evidence suggests that Focus Schools followed state guidance in developing key operational details of their plans for school improvement. This activity is reflected in teacher reports of improved teacher professional development at treatment schools. The available media accounts also provide evidence that Focus School status captured the attention of key stakeholders in the school improvement process. However, our interviews with KDE officials suggested that the quality of state technical assistance and oversight is probably best characterized as uncertain, particularly for our key outcome year (i.e., 2013-14).

4. Data and Sample Construction

We use cross-sections of school-level data with information on relevant subgroup performance to examine the effect of the Focus School reform. Our data come from the annual School Report Cards (SRC) published by the Kentucky Department of Education (KDE) over several years. Our core outcomes are school-level measures of student test performance (e.g., percent proficient) by subgroup for the first full year of implementation (i.e., 2013-14). The SRC also provides school-year data on the baseline assignment variables, Focus School status, and other characteristics of schools, students and teachers. We also add school level data from the Common Core of Data (CCD) on eligibility for and participation in federal Title I programs as well as school directory information.

To construct the analytic sample, we include all schools in the risk pool for Focus School status following the multiple rating assignment procedure. KDE provided 1,296 schools with a differentiated accountability rating but we restrict the analytic sample in several important ways. First, we exclude all 230 high schools because the end-of-course exams administered to these students are not part of the state's K-PREP testing system and separate analyses of high schools are uninformative due to insufficient power. Second, we exclude 12 schools for one of three reasons. The state identified nine of these schools as Priority Schools, implying ineligibility for

Focus School status.[14] We eliminate a single school because it did not meet the minimum student group size to receive an assignment score.[15] We also exclude two schools, which are part of a PK-12 laboratory school situated on Eastern Kentucky University's campus and require an application for admission and charge tuition and fees. These edits resulted in a sample of 1,054 elementary and middle schools eligible for Focus School status.

Our sample is limited further by the exclusion of schools that have a valid assignment variable but lack outcome data (i.e., test scores). Schools lack valid outcome data for one of two reasons: school closure (n=27) or a grade reconfiguration such that the school no longer has assessment results for the content level in which they received their original accountability assignment (n=11). The exclusion of these 38 schools, which are located in small, rural districts, reduces our sample further to 1,016 schools. Though these school changes are uncommon, we find that they are concentrated among schools that were *not* designated as Focus Schools (Appendix Figure 1). In fact, parametric RD estimates indicate that Focus School status significantly reduced the probability of school-closure or reconstitution.[16] Given the policy debates around school closures, we view this as a substantively interesting finding (i.e., that an energetic school-reform effort reduced the risk of a school being closed or reorganized).

However, in terms of evaluating the impact of Focus School reforms, these school changes, though uncommon, may still constitute an internal-validity threat. Specifically, it is reasonable to suppose that the schools that were closed or reorganized would have had comparatively poor student outcomes if they had remained structured as they were. The systematic removal of such struggling schools from the "control" side of the threshold that determined Focus School status would then imply an upward bias in the estimated effects of these reforms. To address this

---

[14] Kentucky identified 41 Priority Schools, 32 were high schools and nine middle schools. No elementary schools were given Priority School status.

[15] Anchorage Independent Public School is a K-8 school that is the sole school in the school district. The elementary level classification was "High performing" while the state classified the middle school as a reward school. The school did not have enough students in the gap group (or any of the individual subgroups) to present results. They are the only school in KY that did not report information on gap group students at baseline.

[16] We look both parametrically and non-parametrically for discontinuities at the assignment threshold for school closure and grade reconstitution (both combined and separately). Our estimates confirm the statistical significance of the discontinuities.

concern, we identified the 18 small, rural school districts in which these school changes occurred and eliminated *all* of the schools from these districts. This implied an exclusion of only 96 additional schools and reduced our final analytical sample to 920 schools.[17] We focus on this sample because of the clear integrity of the intent-to-treat population. However, we also note that the fuller sample implies similar results suggesting that the external-validity caveats associated with deleting these rural districts is not empirically consequential.[18]

In Table 1, we present descriptive statistics for this sample of 920 schools. Our core outcome measures are the reading and math proficiency rates for the targeted umbrella "Gap Group" in 2013-14. In the publicly reported data, there is some modest suppression of these test score results because of privacy concerns related to small cell sizes and the restrictions of the Family Educational Rights and Privacy Act (FERPA). For our core test-score outcomes, only two schools suppress their reading performance and four schools suppress their math performance. Given how uncommon this suppression is (and the procedure for being allowed to do so), there is little reason to suspect that it constitutes an internal-validity threat. Nonetheless, for every test-score outcome and specification (e.g., fuzzy and frontier RD) we report, we examined auxiliary RD models that examined whether test-score suppression was balanced around the threshold. We consistently found no evidence that test suppression changed significantly at the Focus School threshold.

Kentucky's SRC contain a rich set of covariates at the school level including content level, (e.g. elementary, middle, high), total enrollment, student-teacher ratio, attendance and teacher qualifications (Table 1). We supplement SRC data with additional measures from the Common Core of Data including school urbanicity, Title I eligibility and Title I program type. We utilize performance level assessment data from the Kentucky Performance Rating of Education

---

[17] We adopted a parallel sample-construction protocol for the 2014-15 results we report in an appendix. This results in a sample of 852 schools.

[18] We show these results in Appendix Table 1. The first column shows results from our preferred sample (i.e., including only schools in districts without school closures or grade configuration changes). The second column includes all schools with available results and, finally, the third column uses the Last Observation Carry Forward (LOCF) imputation procedure. LOCF carries forward the last available proficiency rate for schools in the sample with missing outcomes.

Progress (K-PREP). These K-PREP assessment data are aligned with the Common Core State Standards (CCSS) and available in mathematics and reading for grades 3-8 aggregated at the school level for different subgroups. Performance data reports the percent of students achieving a particular performance level: novice, apprentice, proficient and distinguished. While we focus on the proficiency rates of "gap group" students (i.e. free/reduced price lunch, special education, black, Hispanic, American Indian, and limited English proficient), we also explore treatment heterogeneity at different performance thresholds and for the available subgroups. The K-PREP assessment program was first fielded in the 2011-12 school year (i.e., the first state to administer CCSS-aligned assessments). The identification of Focus Schools under Kentucky's NCLB waiver, which we describe in more detail below, is from this first year of K-PREP assessments.

Table 1 lists baseline characteristics for the 920 schools in the analytic sample that are eligible for Focus School status. About one-fifth of the analytic sample, 187 schools, have Focus School status. The Student Gap Group (SGG) score, based on 2011-12 K-PREP results, is one of the assignment variables that determines Focus School status. We center the SGG score so that schools with a value less than zero are eligible for Focus School status (i.e., 9.2 percent of the sample).

Student characteristics include the percent of male students, percent of students qualifying for free/reduced price lunch, racial/ethnic composition (e.g. white, black, Hispanic etc.) and attendance rates. Kentucky schools largely enroll white students; the average school enrolls 82.7 percent white students though some regions enroll relatively small shares of whites and larger proportions of blacks and Hispanics. Asians, American Indian/Alaskan Natives, and other groups comprise a small proportion of Kentucky students. Due to the overall low enrollment numbers of blacks and Hispanics and their concentration in a minority of schools, we do not have statistical power to detect treatment effects for these subgroups and instead focus on the more inclusive gap group. Students in Kentucky are relatively socioeconomically disadvantaged; the average school has 63 percent of students eligible for free or reduced priced meals. Kentucky is a predominantly rural state; half of the schools in the analytic sample are located in rural areas while less than one

in five are located in urban areas. Although information on limited English proficiency (LEP) and special education program participation is not available at baseline, they are available for the post-treatment period and we use these data to conduct additional robustness checks.

Almost 80 percent of schools administer a Title I school wide program (SWP) at baseline, a designation that prior to the ESEA waiver required schools to enroll at least 40 percent free/reduced price lunch eligible students.[19] Teacher characteristics include teachers' mean years of experience and the proportion with an advanced degree. While the average school has a teaching staff with over 10 years of experience and just over half have advanced degrees, this varies dramatically across schools. Average teacher experience ranges from 2 to 20 years and the proportion of teaching staff with an advanced degree ranges from 10 to 90 percent.

5. Regression Discontinuity (RD) Design & Estimation

5.1. Assignment mechanisms

In this study, we largely focus on one of the two assignment criteria for Focus School classification. The Student Gap Group (SGG) score is a weighted average of the percent of proficient or above for "gap group" students across five subject areas in the 2011-12 K-PREP assessments.[20] Gap group students with membership in multiple subgroups are counted once in the formula. Therefore, the group is based on the unduplicated count of students with membership in any gap group (i.e. free/reduced price lunch, special education, black, Hispanic, American Indian and/or limited English proficient). The threshold for Focus School status is specific to each school level; schools with an SGG score in the bottom 10 percent of elementary, middle or high schools are eligible for Focus School classification.

---

[19] As noted earlier, 42 of the 187 Focus Schools in our sample obtained SWP status as part of the reforms, implying that funding flexibility is part of the treatment contrast. Visual inspection of SWP status suggests a modest increase at the eligibility threshold. However, auxiliary RD regressions that condition on the control variables indicate that SWP status appears continuous through the Focus School threshold, suggesting it is not a salient part of the contrast we study.

[20] The five subject areas consist of reading, math, science, social studies, and writing/language mechanics. We focus on reading and math because these are the only subjects in which students are tested annually in grades 3 through 8. Students are tested in science in grades 4 and 7; social studies in grades 5 and 8; writing in grades 5, 6 and 8; and language mechanics in grades 4 and 6.

A second assignment criteria, the Third Standard Deviation (TSD) method, targets schools with low subgroup performance relative to the statewide proficiency rate in any of the five designated subjects. Any school with a minimum of 25 tested students in a given subgroup would receive the Focus School designation if a subgroup-by-subject performance met the TSD assignment criteria. For each level (i.e. elementary, middle and high school), the mean and standard deviation of the statewide proficiency rate for all students was calculated in each of the five subjects. Any school with a subgroup performing three standard deviations below the statewide mean in a subject was eligible for Focus School status. This implies that, in theory, a school could have as many as 30 TSD assignment variables (i.e., scores for 6 subgroups by 5 subjects).[21] However, we define each school's relevant TSD assignment variable as its *lowest* subgroup-by-subject performance measure with respect to the relevant TSD threshold. As a practical matter, the lowest TSD score among Focus-eligible schools was often in reading (76 percent) and the lowest-performing subgroup was, typically, special education (86 percent).

In sum, the state intended to designate a school as a Focus School if one of two conditions were met. One is when gap group students performed below the 10-percent threshold for its school level. Alternatively, a school would receive Focus status if, for any designated subject, a targeted subgroup of 25 or more students within the school performed 3 standard deviations below the state mean for the group. Because schools receive a rating along both dimensions, they can fall into one of four quadrants that relate to their intent-to-treat status as illustrated in Figure 1. The upper right quadrant contains schools with SGG scores and their lowest TSD scores above the assignment thresholds (n=720). These schools are ineligible for Focus School status (i.e., the intent to treat is zero). The 200 schools spread across the remaining three quadrants qualify for Focus School status based on their SGG score; their TSD score or both SGG and TSD scores below the assignment threshold (i.e., the intent to treat is 1).

---

[21] This was not true in practice because the 3 standard-deviation threshold was below zero in two subjects (i.e., math and writing). So, the maximum number of TSD assignment variables per school was 18 (i.e., 6 subgroups by 3 subjects).

Our analytical approach, described in detail below, relies almost exclusively on the Focus School assignment generated by SGG scores (rather than the TSD scores) as a source of identifying variation. Our motivation for this emphasis is that the Focus School uptake generated by SGG scores is particularly interesting because it is likely to motivate a broader and more impactful reform than those motivated more narrowly by the low performance of just one subgroup and subject. In our concluding remarks, we underscore that the gap-group policy design we leverage (and the particular treatment contrast it created) is among several important issues that influence the external validity of our findings. However, for completeness, we also show results based on the TSD assignment variable we constructed (Appendix Table 3) where we find similar school-level improvements in students' reading performance, the subject most targeted by this narrow assignment rule.

Our preferred RD designs leverage the credibly quasi-random assignment to Focus reforms generated by the proximity of a school's SGG score to the arbitrary threshold that determined treatment. Specifically, we adopt two approaches articulated in the recent literature on multivariate RD designs (e.g., Reardon & Robinson & 2012; Papay et al. 2011; Wong et al. 2013). One is a straightforward "fuzzy" RD design based on the full sample of n=920 schools. In this sample, all schools with SGG scores below the threshold are subject to Focus School reforms. However, as we show below, a fraction of schools with SGG scores above the threshold are Focus Schools because they are assigned under the TSD rule (hence, the "fuzziness" in the first stage assignment). This approach is akin to comparing the schools on either side of the entire vertical axis in Figure 1. Our second approach is a "frontier" RD design based on eliminating schools that were already treatment-eligible under the TSD rule (i.e., the schools in the bottom two quadrants of Figure 1). This approach gives us a virtually "sharp" first-stage relationship focusing on schools that were only eligible because their SGG score fell below the threshold.[22] We report both results because each has potentially compelling properties. The fuzzy

---

[22] The frontier RD is not a fully sharp design in this application because two ineligible schools mistakenly received Focus School status.

RD approach leverages the statistical power of the full sample. However, the frontier RD approach does not rely on the assumption that potential outcomes are continuous along a longer frontier that may be characterized by heterogeneous treatments (Wong et al. 2013). We also examined a "binding score" RD approach in which both assignment variables were converted to a single variable (i.e., the minimum of both). We find that this approach generates results similar to those we report but believe its implicit assumption (namely, that of homogeneous treatments along the entire response surface) is untenable given the more targeted treatment we expect in schools eligible *only* under the TSD rule.

5.2. Estimation

In this study, we report reduced-form estimates of the effect of Focus School eligibility (i.e., the intent to treat or ITT) based on specifications with the following general form:

$$Y_i = \delta\, I(SGG\ Score_i < 0) + f(SGG\ Score_i) + X_i + \varepsilon_i \tag{1}$$

In this specification, $I(SGG\ Score_i < 0)$ is an indicator for whether the centered SGG score for school i is below the threshold that implies Focus School eligibility. The variable, X_i, represents school-level baseline covariates and ε_i is the error term. The parameter of interest, δ, represents the level change in student proficiency at the Focus School assignment threshold conditional on a flexible function of the assignment variable, $f(SGG\ Score_i)$. Our preferred models impose a linear relationship between the SGG score and outcomes but allow for different slopes on either side of the threshold. We chose our preferred functional form using both graphical evidence and Akaike's Information Criteria (Schochet et al. 2010). However, we also estimate outcomes using higher order polynomials. Furthermore, we also present results based on local linear regressions based on only those observations in increasingly tight bandwidths around the assignment threshold (Lee and Lemieux 2010). These bandwidths include the optimal bandwidth chosen by the procedure introduced by Imbens and Kalyanaraman (2011), which is generally in the range

of +/- 7 to 9 rating score points. We also present estimates based on triangular kernel weights in addition to unweighted (i.e., rectangular kernel) results.

5.3 Treatment Assignment

Before turning to our main results, we first discuss the first-stage relationship between the assignment variable and schools' participation in Focus School reforms as well as the causal warrant of our RD designs. We begin by illustrating, both graphically and parametrically, the relationship between a school's SGG score and its participation in Focus School reforms. In Figure 2(a), we show the change in probability of Focus School status for the SGG assignment variable for the Fuzzy RD sample. All schools with an SGG score in the bottom 10 percent are classified as Focus Schools while the probability drops approximately 70 percentage points to the right of the threshold where some schools are Focus Schools because of the secondary assignment mechanism (i.e., TSD score). In Figure 2(b), the frontier RD changes in probability of treatment from one to nearly zero, a virtually sharp contrast.[23]

In Table 2 we show the first-stage RD estimates based on the SGG assignment variable for the fuzzy and frontier RD. In the sparest specification (i.e., column 1), the RD estimates indicate that the probability of becoming a Focus School jumps by 75 percentage points at the threshold. In the subsequent specifications that introduce additional controls (e.g., quadratic splines and baseline covariates), the point estimate is 68 to 70 percentage points. As expected, in the Frontier RD application, the "jump" in treatment status at the threshold is nearly 1, reflecting the virtually sharp first-stage relationship.

These first-stage relationships clearly show a large and discontinuous jump in a school's treatment status at the arbitrary threshold set by federal policy. Several other ancillary sources of information support our maintained assumption that this RD approach results in credible causal inferences. First, though some schools may feel stigmatized by their Focus status, the multi-tiered assignment protocol based on pre-existing data (i.e., the 2011-12 K-PREP used to generate

---

[23] Two schools that are classified as Focus Schools but do not fall below the assignment threshold for either criteria are included in the analytic sample. When we removed these schools from the analytic sample results were similar. After communicating with KDE officials we were unable to uncover the rationale for Focus School classification.

SGG scores) makes it quite unlikely that schools would have been able to manipulate their treatment assignment. A precise manipulation of the SGG score is also infeasible because it would need to rely on the unduplicated count of students—many of whom could be members of multiple subgroups—that comprise the gap group. Nonetheless, we conduct robustness checks to interrogate this assumption empirically.

For example, following McCrary (2008), we implement a density test to see whether the density of the SGG values jumps discontinuously at the threshold. We fail to reject the null hypotheses that there is no discontinuity at both thresholds for the fuzzy and frontier samples (See Appendix Figure 2). Second, we also examine the density of the assignment variable visually because "heaping" at certain values may fail to be detected by the McCrary density test (Barreca et al. 2016). In our Appendix Figure 3, we present additional histograms to examine potential heaping-induced bias by inspecting the distribution of the forcing variable with the smallest binwidths possible. Although there does appear to be some evidence of heaping at several assignment variable values, we follow the suggestions of Barreca et al. (2016) for detecting non-random heaping and confirmed the robustness of our results.[24]

Third, we also examined the baseline traits of schools listed in Table 1 to assess whether they are continuous across the assignment variable threshold. In Table 3, we present evidence of this covariate balance based on a simple two-stage process. In the first stage, we regressed the outcome of interest, the 2013-14 proficiency rate in math or reading, on all of the baseline covariates. We then obtain predicted proficiency rates based on these regressions. These predicted values provide a single value for each school that represent an index of all the baseline traits that influence outcomes, each weighted by their regression-estimated impacts. We then estimate and report in Table 3 auxiliary RD equations in which these indices are the dependent

---

[24] Specifically, we tested for discontinuous jumps in baseline characteristics at data heaps and did not find statistically significant differences. Nonetheless, we also examined our results, omitting observations at possible heaps, and found similar results.

variables.[25] In a valid RD design, our expectation would that these indices of baseline school traits would be continuous through the threshold that defines treatment status. The results in Table 3 consistently indicate, for different estimation methods (and different samples) that the covariates are balanced around the threshold. Overall, these results clearly suggest the assignment of Kentucky's schools to Focus School reforms functions as a well-behaved regression discontinuity.

6. Results

6.1. Main results

We begin by illustrating our main findings graphically in Figure 3. These 4 panels show the 2013-14 math and reading proficiency rates of gap-group students in schools above and below the treatment-eligibility threshold and do so both for the full "fuzzy" sample and the smaller "frontier" sample. These figures consistently indicate that, after the first full year of implementation, the math and reading proficiency of gap-group students jumped meaningfully at the threshold that influenced Focus School status. For reading, the jumps in both samples appear to be of a slightly smaller magnitude compared to the substantial jump for math. We provide parametric evidence that supports these graphical results in Table 4.

Table 4 reports the estimated reduced-form effect of treatment eligibility on reading and math proficiency across different specifications (i.e., with and without baseline controls and in the fuzzy and frontier samples). The top panel of Table 4 shows the math results, which consistently indicate impacts of roughly 5 percentage points. Because these are intent-to-treat (ITT) estimates and the fuzzy RD has lower first-stage compliance, the "treatment on the treated" (TOT) effects implied by these results is somewhat larger in the fuzzy RD specification. To put estimates of this size in perspective, it should be noted that the math proficiency rate for gap-group students in schools just above the eligibility threshold is approximately 30 percent.

---

[25] The use of these indices attenuates the multiple comparison problems that might plague a plethora of tests based on each baseline covariate.

This implies that a 5 percentage-point increase is equivalent to a 17 percent increase relative to the counterfactual mean.

The bottom panel of Table 4 reports the estimated effect of Focus School eligibility on reading proficiency rates among gap-group students. These ITT estimates indicate that Focus School eligibility increased reading proficiency by 2.3 to 3.6 percentage points.[26] The implied TOT estimates (i.e., the estimated effect of being a Focus School) is between 3 and 3.5 percentage points. Reading proficiency rates are approximately 37 percent for gap-group students in schools just above the assignment threshold. This implies that Focus School reforms increased reading proficiency by roughly 9 percent relative to the counterfactual mean (i.e., 3.5/37).

These results indicate that the Focus School reforms led to substantial gains in math and reading proficiency among the broad gap group. These performance gains are also reflected in subsequent school exits from Focus School status. The KDE used the 2013-14 achievement results to identify (and reclassify) Focus Schools that met performance targets (and to designate new Focus Schools). The treatment effects documented here contributed to the exit of 47 Focus Schools from this status.

6.2. Robustness checks

We perform a number of robustness checks to test the sensitivity of our main results. First, we examine the robustness of our results to functional-form concerns by reporting the results from local linear regressions (LLR). In these LLR specifications, we condition on a linear spline of the assignment variable and the baseline controls but limit the sample to increasingly tight bandwidths of schools proximate to the eligibility threshold. These bandwidth restrictions include the optimal bandwidths (i.e. +/-7 to 9 assignment score points) based on the procedure introduced by Imbens and Kalyanaraman (2011). We also report weighted least-squares

---

[26] In the fuzzy RD specification that omits baseline controls, this estimate is smaller and statistically insignificant. However, we examined Akaike's information criterion (AIC) for this and all other models and found that the AIC privileges those that condition on both linear splines of the assignment variable and the baseline controls.

estimates based on a triangular kernel weight that takes higher values for schools close to either side of the eligibility threshold.

Table 5 reports the key results from these specifications for both math and reading results as well as for the fuzzy and frontier RD samples. The results across these different bandwidths and outcomes consistently indicate that the Focus School reforms had qualitatively large and positive effects on student proficiency in math and reading. In most cases, the estimates based on these smaller samples also remain statistically significant. However, we note two broad exceptions. First, in the smaller frontier RD sample, further bandwidth restrictions imply that the estimated effects on math, though still positive, become smaller and substantially noisier (i.e., estimates in column 3). Second, with regard to the reading results, modest bandwidth restrictions (i.e., within 10 to 20 points of the threshold) imply estimated effects that are smaller and statistically insignificant. However, further bandwidth restrictions and the kernel-weighted results imply larger and statistically significant effects on reading proficiency that are actually somewhat larger than those based on the full samples.

As an additional robustness check, we present several auxiliary RD estimates at "placebo" RD thresholds (see Appendix Table 4). We artificially move the treatment threshold to various *placebo* cut scores to examine whether the positive treatment effects are evident other unimportant thresholds that hold no policy significance for schools. This additional check allows us to increase confidence in our main result estimates, which are significant while the placebo threshold are not. None of the *placebo* estimates are consistently statistically significant and further bolster our main results on the effects of Kentucky's Focus School reforms.

Finally, because our analysis leverages school-level data, we may be concerned that changes in the composition of students enrolled at schools (or those taking the test) drives the positive results. To examine this concern, we first examined auxiliary RD specifications in which the number of students tested (i.e., in the gap group as well as other subgroups) was the dependent variable. These results consistently indicated no change in the size of these test-taking populations at the eligibility threshold. We adopted a similar approach to examine whether non-

random mobility into or out of the school may be a source of bias. First, we estimated a regression of the 2013-14 performance outcomes on a variety of *post-treatment* school covariates reflecting the size and composition of the school (e.g., total enrollment, percent racial/ethnic enrollment, proportion qualifying for free/reduced priced lunch, limited English proficient and special education participation). We then took the predicted values from these regressions and used them as the dependent variables in auxiliary RD specifications. This approach allows us to examine whether an index of these post-treatment school traits, weighted by their relevance for student outcomes, differs above and below the eligibility. The key results reported in Table 6 indicate that these post-treatment school traits do not differ significantly around the threshold. This finding is consistent with the absence of treatment-endogenous student sorting and, by implication, the causal warrant of the RD results.

6.3. Treatment heterogeneity

Another relevant concern with respect to our main findings is that they may be misleading if the state's focus on proficiency encouraged treatment schools to focus their reforms largely on students close to the proficiency threshold (Lazear 2006). In Table 7, we present evidence that speaks to this concern by estimating the effects of Focus School eligibility on alternative test thresholds. Moving a student from the lowest performance level to the apprentice level would not enable a school to exit Focus School status. Similarly, moving a proficient student to distinguished status also does not influence Focus School status. Therefore, if schools were adopting strategic "triage" behavior, we would not expect to see treatment effects at these margins. However, the RD results in Table 7 indicate that Focus School eligibility generated positive and, in most cases, statistically significant effects at each performance threshold. One notable exception is that the estimated effects on distinguished performance in mathematics are much smaller, possibly reflecting the comparative rarity of performance at this level (i.e., 8.6 percent in this sample).

Our analysis so far has focused on the test performance of gap-group students, an amalgamation of traditionally low-performing subgroups and the focal point of these reforms. However, the effects of these reforms on other groups of students are also of interest. For example, it may be that these reforms, which had a schoolwide character (e.g., improvement planning, teacher professional development) had spillover benefits for students outside the gap group. Alternatively, it may be that schools strategically reallocated instructional effort away from such "non-gap" students. Although Kentucky does not specifically report on the performance of non-gap group students, we are able to construct performance level percentages from available information on the performance of all students and gap group students. We also have school performance data on subgroups of students qualifying for free/reduced price lunch and students receiving special education services (i.e., two categories in the gap group). In Table 8, we present RD estimates of the effect of Focus School eligibility on the proficiency rates of these student groups. The results are consistently positive, though not always statistically significant. These results suggest that the performance benefits of the Focus School reforms were broad and, in particular, did not harm (and may have helped) the performance of students outside the gap group.

7. Conclusion

The conduct and performance of public schools in the U.S. have historically been the concern of local and state authorities. However, over the last 15 years, the federal government has been unusually prescriptive in seeking to improve the performance of public schools, particularly with regard to historically marginalized students. The most recent iteration of these controversial, federally catalyzed reforms occurred under waivers to NCLB. A signature feature of these waiver reforms was the push for "differentiated accountability." In order to address achievement gaps, these waivers required states to identify schools where targeted subgroups performed particularly poorly and to support reforms in these designated "Focus Schools." In this study, we examined the causal effects of the Focus School reforms in the state of Kentucky. Our RD results

indicate that these reforms led to substantial improvements in the math and reading performance of the targeted students (and, possibly, among the students who were not targeted as well). We also present evidence that comprehensive school planning, including an emphasis on high-quality teacher professional development, were mediators of these effects.

The effectiveness of Focus School reforms in Kentucky stands in notable contrast to what we have learned about similar reform efforts in other states (Hemelt & Jacob 2017, Dee & Dizon-Ross 2017) where waiver-driven reforms appear to have had no or at best targeted effects. One ready explanation for the comparative success of Focus School reforms in Kentucky involves the state-specific context. The state of Kentucky has established a reputation as an enthusiastic and energetic adopter and implementer of school reforms. The Focus School reforms in Kentucky appeared to reflect this. The federal requirements for Focus School reforms were vague but Kentucky articulated detailed and comprehensive reform activities for Focus Schools to undertake. This engagement and the articulation of "home grown" reforms are both likely to complement a high-fidelity implementation. Furthermore, the Focus School reforms in Kentucky also had a fairly unique design feature: the designation of a large, umbrella group of underserved students for which schools would be held responsible. The use of this larger gap group may have been instrumental in catalyzing school-wide reforms and avoiding incentives for narrowly targeted reform efforts. Learning more about how state and local contexts may channel reform efforts productively is likely to be an increasingly important area of study in education policy.

References

Barreca, A. I., Lindo, J. M., & Waddell, G. R. (2016). Heaping-Induced Bias in Regression-Discontinuity Designs. *Economic Inquiry*, *54*(1), 268-293.

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.

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.

Carnoy, M., & Loeb, S. (2002). Does external accountability affect student outcomes? A cross-state analysis. *Educational evaluation and policy analysis*, *24*(4), 305-331.

Dee, T. (2012). School turnarounds: Evidence from the 2009 stimulus (No. w17990). National Bureau of Economic Research.

Dee, T.S. & Dizon-Ross, E. (2017). School Performance, Accountability, and Waiver Reforms: Evidence from Louisiana. NBER Working Paper.

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

Dee, T. S., Jacob, B.A., & Schwartz, N. L. (2012). The effects of NCLB on school resources and practices. *Educational Evaluation and Policy Analysis*, 0162373712467080.

Delisle, D. S. (2014, August 14). [Letter to Commissioner Terry Holliday]. US Department of Education, Washington, D.C. Retrieved from http://www2.ed.gov/policy/eseaflex/secretary-letters/ky2extltr814.pdf

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

Figlio, D., & Loeb, S. (2011). School accountability. *Handbook of the Economics of Education*, *3*(8), 383-417.

Foley, S. (October 9, 2015) GRC celebrates school of distinction honor. *Central Kentucky News*. Retrieved from http://www.amnews.com/GRG-10092015.html

Gewertz, C. (November 4, 2011). Kentucky to Model Professional Development for New Standards. *Education Week*. Retrieved from http://blogs.edweek.org/edweek/curriculum/2011/11/kentucky_to_model_professional.html.

Governor's Task Force (2011). Breaking new ground: Final report of the Governor's Task Force on transforming education in Kentucky. Retrieved from http://www.schoolinfosystem.org/

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.

Hanushek, E. A., & Raymond, M. E. (2005). Does school accountability lead to improved student performance?. *Journal of policy analysis and management*, *24*(2), 297-327.

Hemelt, S.W. & Jacob, B.A. (2017). Differentiated Accountability and Education Production: Evidence from NCLB Waivers. NBER Working Paper.

Imbens, G., & Kalyanaraman, K. (2011). Optimal bandwidth choice for the regression discontinuity estimator. *The Review of economic studies*, 79, 933-959.

Kentucky Department of Education (2011). Guidelines for closing the gaps for all students. Commissioner's Raising Achievement/Closing Gaps Council. Retrieved from http://education.ky.gov/commofed/adv/pages/cracgc.aspx

Klein, A. (October 2, 2015) Arne Duncan Stepping Down as Education Secretary: Hefty Policy Footprint Over Tumultuous Term. *Education Week*. Retrieved from http://www.edweek.org/ew/articles/2015/10/02/arne-duncan-stepping-down-as-education-secretary.html.

Klein, A. (March 31, 2016) The Every Student Succeeds Act: An ESSA Overview. *Education Week*. Retrieved from http://www.edweek.org/ew/issues/every-student-succeeds-act/index.html

Lazear, E. P. (2006). Speeding, terrorism, and teaching to the test. *The Quarterly Journal of Economics*, 1029-1061

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, *142*(2), 698-714.

National Research Council. (2011). Incentives and test-based accountability in public education (Committee on Incentives and Test-Based Accountability in Public Education, M. Hout, & S. W. Elliott, Eds.). Washington, DC: National Academies Press, Board on Testing and Assessment, Division of Behavioral and Social Sciences and Education

Papay, J. P., Willett, J. B., & Murnane, R. J. (2011). Extending the regression-discontinuity approach to multiple assignment variables. *Journal of Econometrics*, *161*(2), 203-207.

Reardon, S. F., & Robinson, J. P. (2012). Regression discontinuity designs with multiple rating-score variables. *Journal of Research on Educational Effectiveness*, *5*(1), 83-104.

Schochet, P., Cook, T., Deke, J., Imbens, G., Lockwood, J. R., Porter, J., & Smith, J. (2010). Standards for Regression Discontinuity Designs. *What Works Clearinghouse*.

Spears, V.H. (May 5, 2015) William Wells Brown Elementary struggles to move beyond label as state's lowest performer. *Lexington Herald Leader*. Retrieved from http://www.kentucky.com/news/local/education/article44601483.html

Ujifusa, A. (December 31, 2013). Kentucky Chief Holliday is New CCSSO President; Will Focus on Career Readiness. *Education Week*. Retrieved from http://blogs.edweek.org/edweek/state_edwatch/2013/12/kentucky_chief_holliday_is_new_ccsso_president_focus_on_career_readiness.html.

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

U.S. Department of Education. (2011) Kentucky ESEA Flexibility Request: Final Submission. Retrieved from http://www2.ed.gov/policy/elsec/guid/esea-flexibility/map/ky.html

U.S. Department of Education (2012) ESEA Flexibility. Retrieved from http://www2.ed.gov/policy/elsec/guid/esea-flexibility/index.html

U.S. Department of Education (2013) ESEA Flexibility Part B Monitoring Report. Retrieved from https://www2.ed.gov/admins/lead/account/monitoring/reports13/kypartbrpt2014.pdf

Wong, M., Cook, T. D., & Steiner, P. M. (2015). Adding Design Elements to Improve Time Series Designs: No Child Left Behind as an Example of Causal Pattern-Matching. *Journal of Research on Educational Effectiveness*, *8*(2), 245-279.

Wong, V. C., Steiner, P. M., & Cook, T. D. (2013). Analyzing regression-discontinuity designs with multiple assignment variables a comparative study of four estimation methods. *Journal of Educational and Behavioral Statistics*, *38*(2), 107-141.

Table 1. Descriptive Statistics of RD Samples

Variable	Mean	Std. Dev.	Minimum	Maximim
Accountability Characteristics				
Focus School 2013-14	0.203	0.403	0	1.0
SGG Score (centered)	12.5	9.8	-19.8	55.0
I(SGG Score < 0)	0.092	0.290	0	1.0
Reading Proficiency Rate, Gap Group 2013-14	46.2	10.0	17.2	85.7
Math Proficiency Rate, Gap Group 2013-14	38.8	11.5	0.0	78.6
Student Characteristics				
Male	0.512	0.035	0	0.619
White	0.827	0.197	0.073	1.0
Black	0.093	0.144	0	0.767
Hispanic	0.043	0.061	0	0.739
Asian	0.012	0.021	0	0.218
Free/Reduced-Price Lunch	0.630	0.200	0.028	1.0
Attendance	0.952	0.012	0.892	0.979
School Characteristics				
Elementary	0.704	0.457	0	1.0
Middle	0.296	0.457	0	1.0
Enrollment	475	195	79	1,625
Student-teacher ratio	15.5	2.2	10	21
Rural	0.500	0.500	0	1.0
Urban	0.173	0.378	0	1.0
Suburban	0.134	0.341	0	1.0
Reward school	0.053	0.225	0	1.0
Title I-eligible	0.919	0.274	0	1.0
Title I Schoolwide Program	0.799	0.401	0	1.0
Teacher Characteristics				
Teaching experience (years)	11.7	2.5	2.3	20
Advanced degree (MA, PhD)	0.513	0.126	0.071	0.917

SOURCES: Student, school and teacher characteristics are authors' calculations using Kentucky School Report Card (SRC) and Common Core of Data (CCD), 2011-12.

NOTES: The sample includes 920 schools with an SGG (Student Gap Group) assignment variable. The SGG Score refers to the baseline assignment variable based on 2011-12 test scores and is centered on the eligibility threshold. The student-teacher ratio is recoded (winsorized) at the 1st and 99th percentiles.

Table 2. The Effect of Focus School Eligibility on Focus School Status

Independent Variable:	Dependent Variable: Focus School 2013-14							
	Fuzzy RD				Frontier RD			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$I(\text{SGG Score}_i < 0)$	0.746***	0.700***	0.698***	0.684***	0.994***	0.995***	0.995***	0.993***
	(0.025)	(0.038)	(0.038)	(0.047)	(0.004)	(0.004)	(0.005)	(0.005)
R^2	0.434	0.527	0.437	0.529	0.957	0.957	0.957	0.957
Linear spline	yes	yes	yes	yes	yes	yes	yes	yes
Quadratic spline	no	no	yes	yes	no	no	yes	yes
Baseline controls	no	yes	no	yes	no	yes	no	yes

NOTES: *** $p < 0.001$ ** $p < 0.01$ * $p < 0.05$ Robust standard errors in parentheses. Each column contains the first stage, the estimated effect of the intent to treat on treatment status. See Table 1 for additional details on student, school and teacher controls. Fuzzy RD n=920 ; Frontier RD n=767.

Table 3. Auxiliary RD Estimates of Baseline Covariate Balance

Dependent variable:	Math achievement index		Reading achievement index	
Sample	Fuzzy RD	Frontier RD	Fuzzy RD	Frontier RD
Gap Group	0.102	0.493	-1.523	-0.110
	(0.791)	(0.962)	(0.844)	(1.052)
Non-gap group	1.909	0.807	0.080	0.461
	(1.231)	(1.621)	(0.820)	(1.122)
Free/Reduced-Price Lunch	-0.316	0.085	-1.661	0.319
	(0.812)	(0.928)	(0.934)	(1.121)
Special education	-1.218	-0.592	-2.720	-1.395
	(1.660)	(2.290)	(2.042)	(2.876)

NOTES: *** $p < .001$ ** $p < .01$ * $p < .05$. Robust standard errors in parentheses.
Each cell contains the results of a two-stage regression: (1) in the first stage the 2013-14 outcome for the relevant subgroup is regressed on all baseline covariates (see Table 1 for a description of covariates) and a predicted achievement composite is generated. (2) In the second stage the predicted achievement composite is then regressed on I(SGG Score < 0) and a linear spline for the assignment variable.

Table 4. The Effect of Focus School Eligibility on Gap Group Proficiency Rates by Subject

Independent Variable	Math Proficiency Rate			
	Fuzzy RD		Frontier RD	
	(1)	(2)	(3)	(4)
$I(SGG\ Score_i < 0)$	4.361***	5.166***	4.586**	4.754**
	(1.651)	(1.641)	(1.976)	(1.976)
R^2	0.343	0.425	0.304	0.382
n	916	916	765	765
	Reading Proficiency Rate			
Independent Variable	Fuzzy RD		Frontier RD	
$I(SGG\ Score_i < 0)$	0.373	2.349**	2.518*	3.555***
	(1.177)	(1.088)	(1.371)	(1.332)
R^2	0.418	0.487	0.366	0.427
n	918	918	765	765
Baseline controls	no	yes	no	yes

NOTES: *** $p < .01$ ** $p < .05$ * $p < .1$. Robust standard errors in parentheses.
Each cell contains the result of a separate regression of the effect of $I(SGG\ Score_i < 0)$ on 2013-14 proficiency rates for gap group students. All models condition on a linear spline of the assignment variable. Akaike's information criterion (AIC) implied the optimal order of polynomial is linear.

Table 5. The Effect of Focus School Eligibility on Proficiency Rates, by Alternative Bandwidths

Bandwidth Sample	Math Proficiency Rate, Gap Group				Reading Proficiency Rate, Gap Group			
	Fuzzy RD		Frontier RD		Fuzzy RD		Frontier RD	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Estimate	n (Focus)	Estimate	n (Focus)	Estimate	n (Focus)	Estimate	n (Focus)
Full Sample	**5.166***	916	**4.754**	765	**2.349**	918	**3.555***	765
	(1.641)	(187)	**(1.976)**	(49)	**(1.088)**	(187)	**(1.332)**	(49)
$|SGG\ Score_i| \leq 20$	4.854***	715	4.132**	570	1.819	717	2.823**	570
	(1.707)	(182)	(2.066)	(49)	(1.149)	(182)	(1.418)	(49)
$|SGG\ Score_i| \leq 15$	4.135**	574	3.287	435	1.513	576	2.239	435
	(1.730)	(175)	(2.152)	(48)	(1.270)	(175)	(1.617)	(48)
$|SGG\ Score_i| \leq 10$	4.512**	387	3.477	281	2.443	388	3.545*	281
	(1.984)	(142)	(2.534)	(43)	(1.513)	(142)	(1.893)	(43)
$|SGG\ Score_i| \leq 7$	4.501*	252	2.380	170	4.084**	253	4.962**	170
	(2.343)	(113)	(2.854)	(36)	(1.775)	(113)	(2.176)	(36)
Kernel regression	4.727**	368	3.183	264	3.917**	369	5.028***	264
	(2.222)	(141)	(2.669)	(40)	(1.585)	(138)	(1.823)	(40)

NOTES: *** $p < .01$ ** $p < .05$ * $p < .1$. Robust standard errors in parentheses. Each cell contains a regression of 2013-14 reading or math proficiency rates for schools within the specified bandwidth on I(SGG Score$_i$ < 0), a linear spline of the assignment variable and baseline covariates. Columns 2, 4, 6, and 8 contain the number of schools in the specified bandwidth and the number of treatment schools below the bandwidth total in parentheses. See Table 1 for a description of controls.

Table 6. Auxiliary RD Estimates of Post-treatment Covariate Balance

	Fuzzy RD	Frontier RD
Math achievement index	0.269	0.0328
	(0.456)	(0.624)
Reading achievement index	0.344	0.138
	(0.414)	(0.478)

NOTES: *** p < .001 ** p < .01 * p < .05. Robust standard errors in parentheses. Each cell contains the results of a two-stage regression: (1) in the first stage the 2013-14 test score for the gap subgroup is regressed on all post-treatment school covariates (i.e. percent racial/ethnic enrollment, total enrollment, attendance, student-teacher ratio, percent free/reduced lunch, percent ELL, percent special education and number of tested students in a subgroup) and a predicted composite is generated. (2) In the second stage the predicted post-treatment composite is regressed on I(SGG Score < 0) and a linear spline for the assignment variable and baseline covariates (See Table 1 for a description of covariates).

Table 7. The Effect of Focus School Eligibilty on Gap Group Proficiency Levels

Performance Level		Math		
		Fuzzy RD	Frontier RD	
		Estimate	Estimate	Mean
Apprentice and above		2.828*	2.392	78.5
		(1.591)	(1.841)	
Proficient and above		**5.166***	**4.754**	38.8
		(1.641)	**(1.976)**	
Distinguished and above		0.993	0.430	8.6
		(0.678)	(0.938)	
	N	916	765	
		Reading		
		Fuzzy RD	Frontier RD	
		Estimate	Estimate	Mean
Apprentice and above		2.393*	3.484**	74.3
		(1.246)	(1.522)	
Proficient and above		**2.349**	**3.555***	46.2
		(1.088)	**(1.332)**	
Distinguished and above		1.780**	2.172**	12.5
		(0.719)	(0.981)	
	N	918	765	

NOTES: *** p < .01 ** p < .05 * p < .1. Robust standard errors in parentheses.
Each cell contains a separate regression of the effect of I(SGG Score$_i$ < 0) on 2013-14
proficiency levels for gap group students. All models contain a linear spline for the assignment
variable and controls from the preferred specification in Table 4. The lowest proficiency level
is novice followed by apprentice, proficienct and distinguished.

Table 8. The Effect of Focus School Eligibility on Subgroup Proficiency Rates

Sample	Math Proficiency Rate				Reading Proficiency Rate			
	Fuzzy RD	n	Frontier RD	n	Fuzzy RD	n	Frontier RD	n
Subgroup								
Gap Group	**5.166*****	916	**4.754****	765	**2.349****	918	**3.555*****	765
	(1.641)		**(1.976)**		**(1.088)**		**(1.332)**	
Free/Reduced-Price Lunch	4.077**	905	3.128	754	2.003*	904	3.033*	752
	(1.614)		(2.081)		(1.186)		(1.441)	
Special education	6.378**	854	5.294	707	2.613	852	4.293	706
	(2.578)		(3.783)		(2.422)		(3.309)	
Non-gap	3.365	914	8.531**	764	1.528	916	7.272*	764
	(3.167)		(4.183)		(2.941)		(3.272)	

NOTES: *** $p < .01$ ** $p < .05$ * $p < .1$. Robust standard errors in parentheses. Each cell contains the result of a separate regression of $I(SGG \ Score_i < 0)$ on proficiency rates for different subgroups. All models condition on a linear spline of the assignment variable and the full set of controls from the preferred specification in Table 4.

Table A1. The Effect of Focus School Status on Gap Group Proficiency Rates, by Alternative Samples

Sample:	Math Proficiency Rate, Gap Group					
	Actual Sample (Excluding districts with closures)		All Schools with Results		Last Observation Carry Forward (LOCF)	
	Estimate	n	Estimate	n	Estimate	n
	(1)	(2)	(3)	(4)	(5)	(6)
Fuzzy RD 2013-14	5.166***	916	5.286***	1012	5.141***	1054
	(1.641)		(1.570)		(1.547)	
Frontier RD 2013-14	4.754**	765	5.001***	853	4.701**	890
	(1.976)		(1.864)		(1.827)	
Fuzzy RD 2014-15	1.960	849	3.338**	994	3.713**	1054
	(1.539)		(1.522)		(1.539)	
Frontier RD 2014-15	0.460	701	3.115*	834	3.493*	890
	(1.599)		(1.731)		(1.795)	
	Reading Proficiency Rate, Gap Group					
	Estimate	n	Estimate	n	Estimate	n
Fuzzy RD 2013-14	2.349**	918	3.227***	1013	3.121***	1054
	(1.088)		(1.047)		(1.047)	
Frontier RD 2013-14	3.555***	765	4.573***	852	4.283***	890
	(1.332)		(1.258)		(1.275)	
Fuzzy RD 2014-15	1.154	852	2.591**	997	2.462*	1054
	(1.284)		(1.269)		(1.268)	
Frontier RD 2014-15	0.495	703	2.915*	836	2.559	890
	(1.663)		(1.658)		(1.656)	

NOTES: *** $p < .01$ ** $p < .05$ * $p < .1$. Robust standard errors in parentheses.
Each cell contains a separate regression of the effect of $I(\text{SGG Score}_i < 0)$ on proficiency rates for gap group students. All models contain a linear spline for the assignment variable and controls from the preferred specification in Table 4. See Table 1 for a description of controls. Column (1) contains the preferred estimate from the main results in Table 4 based on a the sample that excludes all districts with any school closures. Column 3 contains estimates that include any school with outcomes in a particular year. Column 5 contains estimates for all schools that received an assignment score at baseline and the last observed proficiency rate is imputed for schools with a missing outcome.

Table A2. Auxiliary RD Estimates of Focus School Eligibility on Teacher Satisfaction Survey

	Fuzzy RD		
	Full Sample	|SGG Score$_i$| ≤ 20	|SGG Score$_i$| ≤ 10
PD Quality	**0.0811**	**0.0793**	**0.102****
	(0.0497)	**(0.0544)**	**(0.0491)**
Sufficient Resources for PD	0.0825	0.0987	0.117**
	(0.0545)	(0.0603)	(0.0544)
Appropriate Time for PD	0.0759	0.0728	0.0893*
	(0.0515)	(0.0563)	(0.0515)
PD is Data Driven	0.0224	0.0232	0.0601
	(0.0560)	(0.0611)	(0.0550)
PD Aligned with School Improvement Plan	0.0778	0.0792	0.0954**
	(0.0474)	(0.0521)	(0.0462)
PD Deepens Content Knowledge	0.0612	0.0456	0.0713
	(0.0531)	(0.0580)	(0.0525)
PD Follow up Provided	0.124**	0.132**	0.149**
	(0.0600)	(0.0671)	(0.0606)
PD Provides Collaboration Opportunities	0.104*	0.0893	0.114**
	(0.0562)	(0.0612)	(0.0557)
PD is Evaluated and Results Communicated	0.0758	0.0776	0.113*
	(0.0637)	(0.0684)	(0.0637)
PD Teaches Strategies to Meet Student Needs	0.104**	0.0964*	0.116**
	(0.0477)	(0.0526)	(0.0479)
PD Helps Teachers Improve Student Learning	0.0903**	0.0768	0.0998**
	(0.0445)	(0.0495)	(0.0444)
Instructional Supports	0.0600	0.0474	0.0694
	(0.0455)	(0.0533)	(0.0492)
Had Common Core PD	-0.00223	-0.0198	-0.0306
	(0.0296)	(0.0330)	(0.0308)
Had Closing Achievement Gap PD	0.0347	0.0196	0.0364
	(0.0310)	(0.0345)	(0.0315)
Had Special Education PD	0.0349	0.0345	0.0310
	(0.0233)	(0.0272)	(0.0251)
n	909	710	382
Linear spline	yes	yes	yes
Quadratic spline	yes	no	no
Baseline controls	yes	yes	yes

NOTES: *** p < .01 ** p < .05 * p < .1. Robust standard errors in parentheses. Each cell contains a separate auxiliary RD regression of I(SGG Score$_i$ <0) on post-treatment covariates from the 2015 TELL Kentucky Teacher Survey. Professional Development (PD) Quality consists of a composite of 10 four-point likert items related to PD quality (e.g. resources, feedback, data usage, follow-up, collaboration etc.).

Table A3. First Stage and Reduced-Form Estimates by Subgroup and Proficiency Level, TSD Threshold

Dependent Variable:		Focus School 2013-14		
	Fuzzy RD		Frontier RD	
Independent Variable:	0.673***	0.662***	0.751***	0.744***
I(TSD Score$_i$ < 0)	(0.054)	(0.054)	(0.061)	(0.061)
R^2	0.638	0.675	0.841	0.846
n	899	899	819	819
Basline controls	no	yes	no	yes

	Dependent Variable: 2014 Math Proficiency Rate			
	Gap Group		Special Education	
Performance Level	Fuzzy RD	Frontier RD	Fuzzy RD	Frontier RD
Apprentice and above	1.516	1.680	-0.256	-0.193
	(1.075)	(1.134)	(2.172)	(2.393)
Proficient and above	1.864	1.912	1.444	0.808
	(1.232)	(1.329)	(1.818)	(1.987)
Distinguished and above	0.635	0.765	1.306	1.440
	(0.529)	(0.551)	(0.862)	(0.962)
n	895	816	843	769

	Dependent Variable: 2014 Reading Proficiency Rate			
	Gap Group		Special Education	
	Fuzzy RD	Frontier RD	Fuzzy RD	Frontier RD
Apprentice and above	1.648*	2.234**	1.252	1.887
	(0.988)	(1.013)	(1.999)	(2.212)
Proficient and above	2.271**	2.922***	-0.614	-0.467
	(0.989)	(1.008)	(1.662)	(1.817)
Distinguished and above	1.057*	1.388**	1.869**	2.108**
	(0.544)	(0.559)	(0.933)	(1.026)
n	897	818	841	768

NOTES: *** p < .01 ** p < .05 * p < .1. Robust standard errors in parentheses. All models contain a linear spline for the assignment variable and the full set of controls from the preferred model in Table 4 (See Table 1 for a description of the controls).

Table A4. Placebo RD Estimates

	Math Proficiency Rate		Reading Proficiency Rate	
	Fuzzy RD	Frontier RD	Fuzzy RD	Frontier RD
Placebo RD: $I(\text{SGG Score}_i \leq -8)$	2.983	0.332	2.322	-0.223
	(2.621)	(3.694)	(2.609)	(3.306)
Placebo RD: $I(\text{SGG Score}_i \leq -4)$	2.561	3.173	-0.825	0.046
	(2.347)	(2.837)	(2.145)	(3.380)
Actual RD: $I(\text{SGG Score}_i < 0)$	**5.166*****	**4.754****	**2.349****	**3.555*****
	(1.641)	**(1.976)**	**(1.088)**	**(1.332)**
Placebo RD: $I(\text{SGG Score}_i \leq +4)$	1.514	2.272	0.114	0.674
	(1.177)	(1.433)	-0.898	(1.112)
Placebo RD: $I(\text{SGG Score}_i \leq +8)$	0.655	1.467	1.488	1.991*
	(0.983)	(1.217)	(0.809)	(0.968)
Placebo RD: $I(\text{SGG Score}_i \leq +12)$	-0.585	-0.506	1.091	1.070
	(0.961)	(1.114)	(0.813)	(0.928)
Placebo RD: $I(\text{SGG Score}_i \leq +16)$	0.276	0.490	0.703	0.243
	(1.059)	(1.144)	(0.883)	(0.953)
N	916	765	918	765

NOTES: *** p < .01 ** p < .05 * p < .1. Robust standard errors in parentheses. Each cell contains an estimate of the focus school classification discontinuity for different values of the assignment variable. All models contain a linear spline of the assignment variable and baseline covariates. See Table 1 for a description of controls.
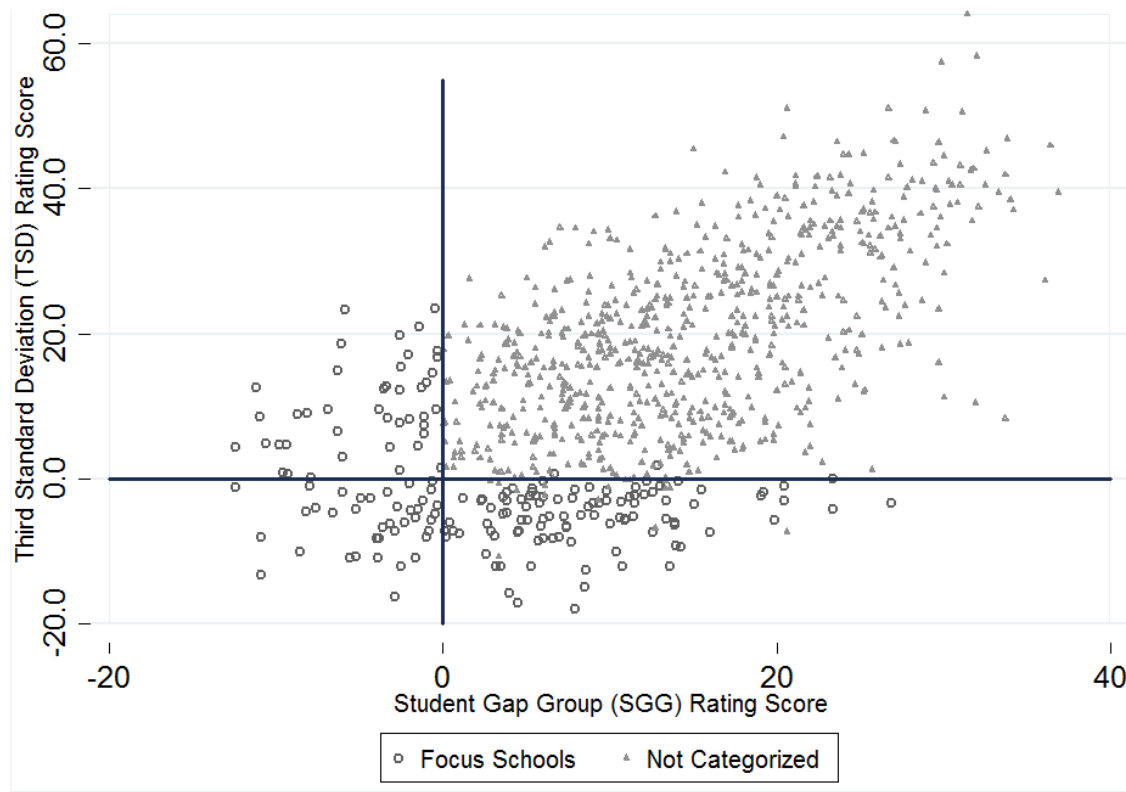
Figure 1. Focus School assignment by forcing variables

NOTES: Graph illustrates Focus School assignment along both forcing variable dimensions, Student Gap Group (SGG) and Third Standard Deviation (TSD) rating scores, on the x and y-axis respectively. Schools with rating scores less than zero on either dimension meet the Focus School classification criteria.
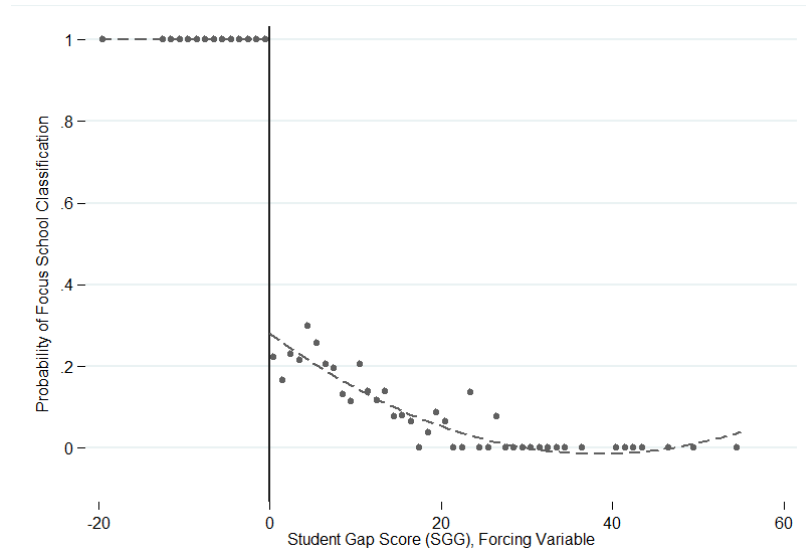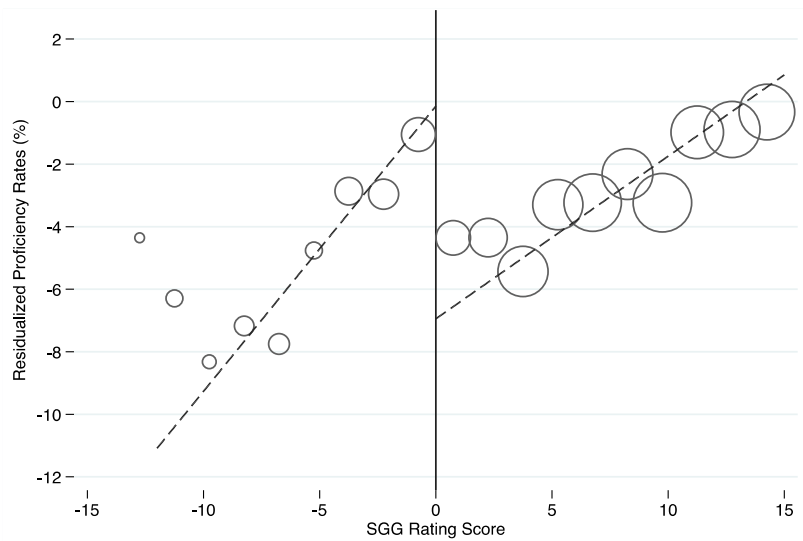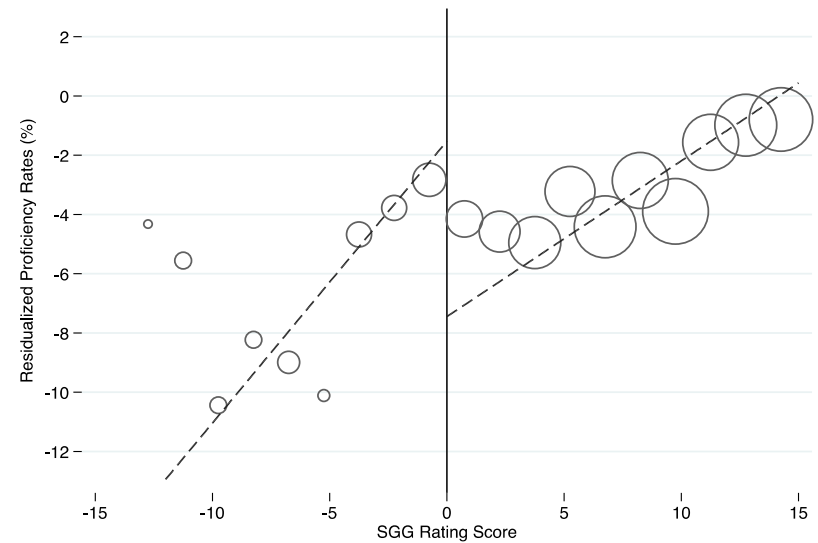
(a) Fuzzy RD
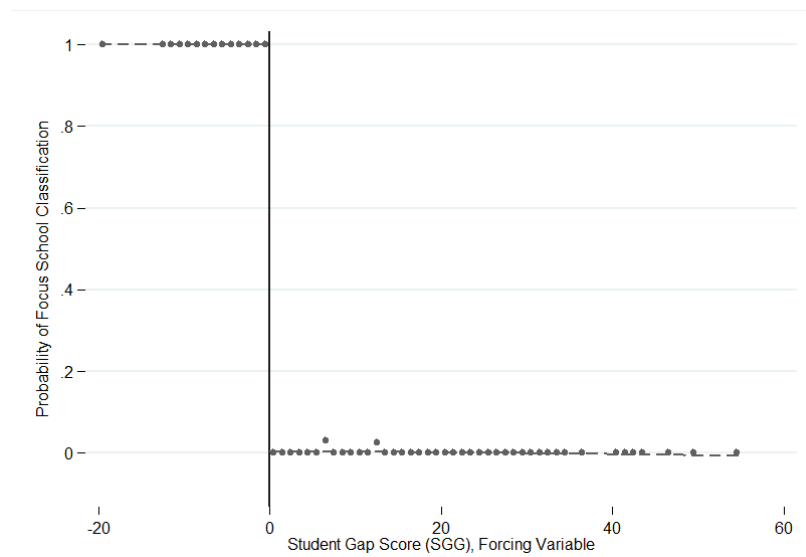
(b) Frontier RD

Figure 2. Focus School Status by Student Gap Group Score, 2013-14

NOTES: Graphs of focus school treatment status by the SGG assignment variables for Fuzzy and Frontier RD samples. Bin width 0.5, full sample.

(a) 2013-14 Math Proficiency Residualized, Gap Group, Fuzzy RD

(b) 2013-14 Math Proficiency Residualized, Gap Group, Frontier RD
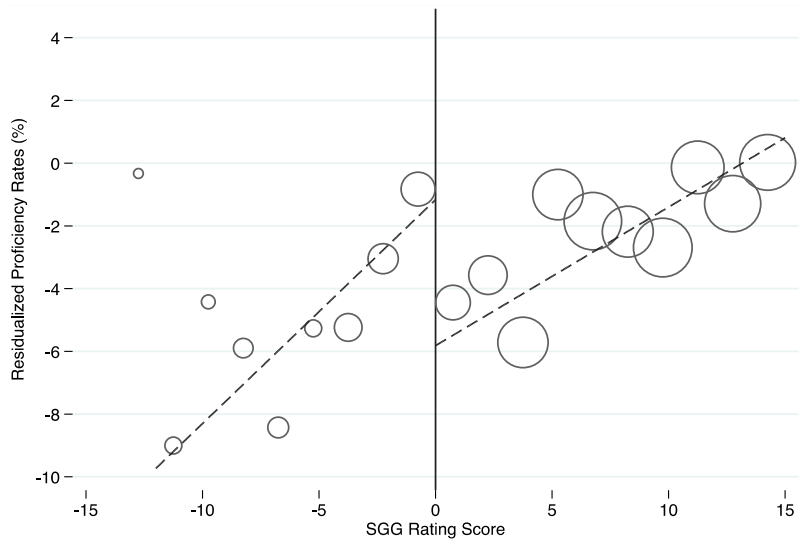
(c) 2013-14 Reading Proficiency Residualized, Gap Group, Fuzzy RD
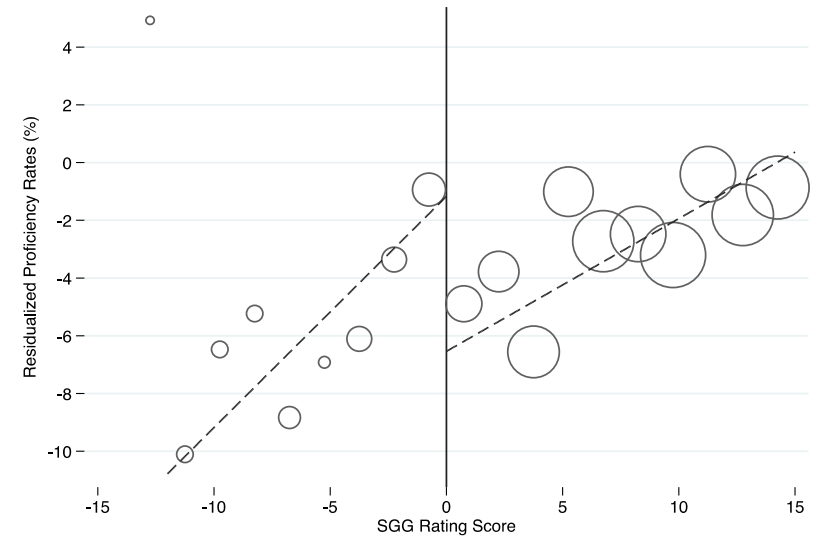
(d) 2014 Reading Proficiency Residualized, Gap Group, Frontier RD

Figure 3. Gap Group Proficiency Rates, by RD Sample, 2013-14

NOTES: Graphs of 2014 residualized proficiency rate for math and reading by the SGG assignment variable for Fuzzy and Frontier RD samples. Bin width 1.5, bandwidth +/-15, Markers weighted by number of schools in bin width.
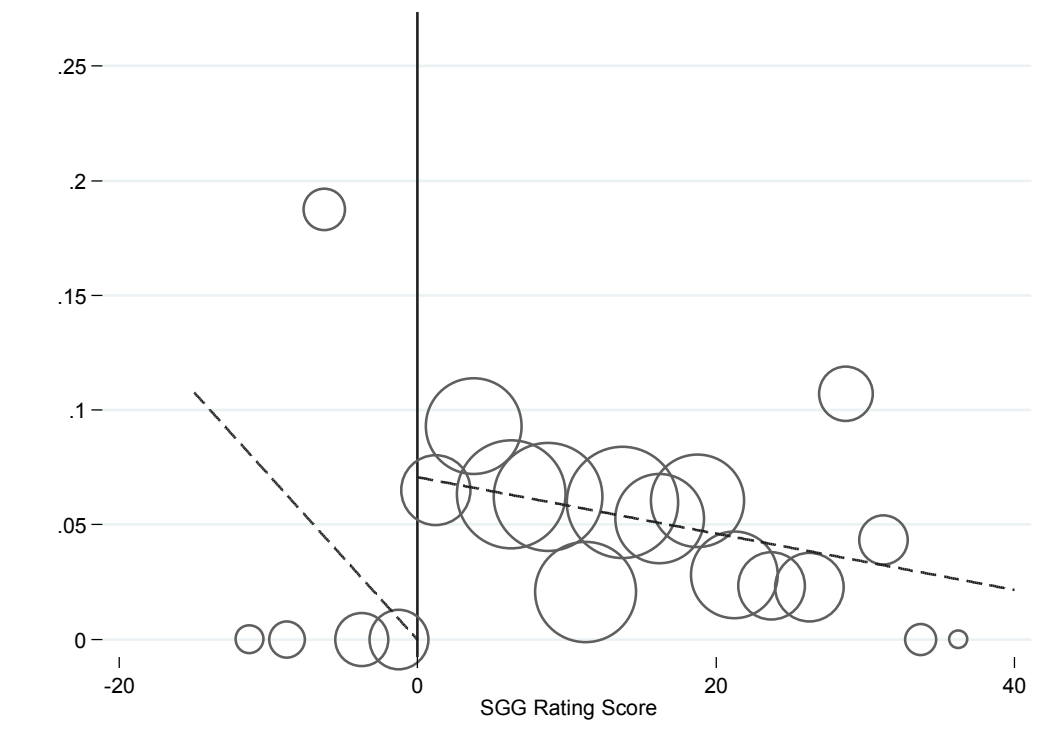
Appendix Figure 1. School closure or grade reconstitution 2013-15

NOTES: Graphs of school closure or grade reconstitution during the treatment period (2013-15) by the SGG assignment variable (n=1054). Bin width 2.5, full sample. Markers weighted by number of schools in bin width. Parametric point estimate -0.0608 (0.0173).
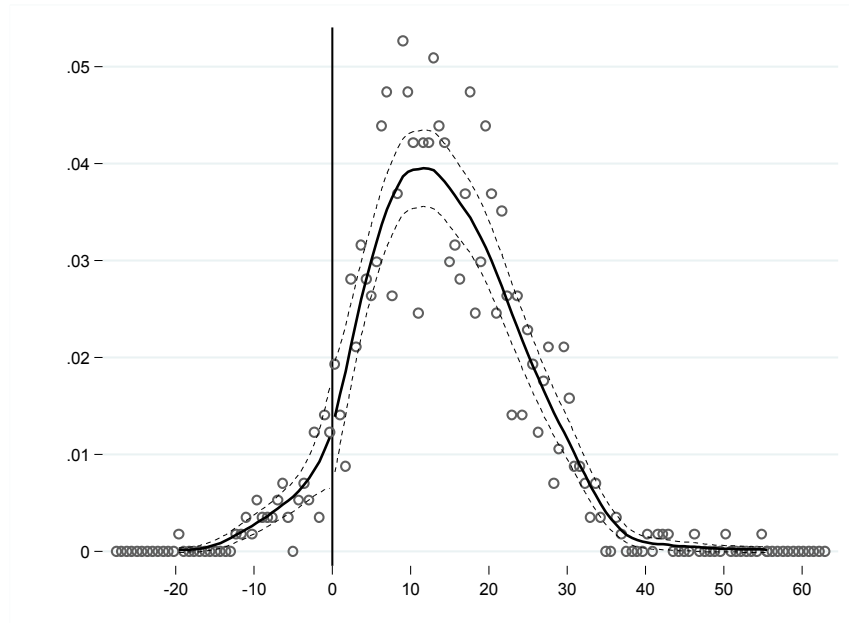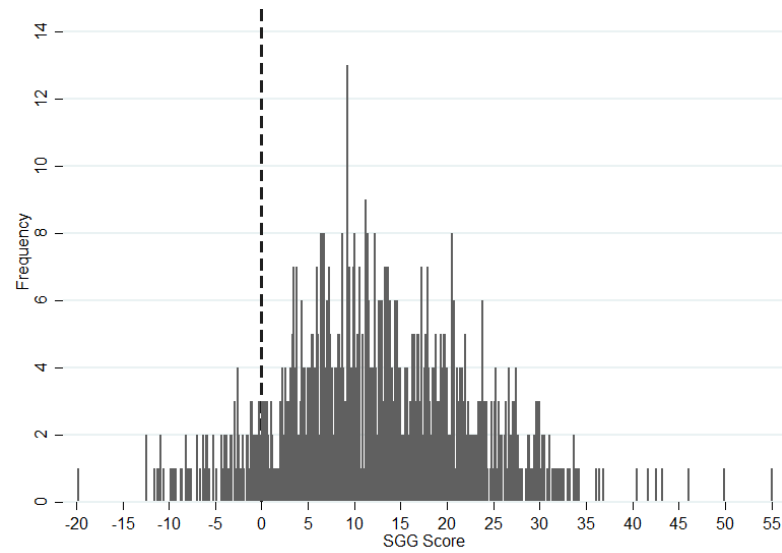
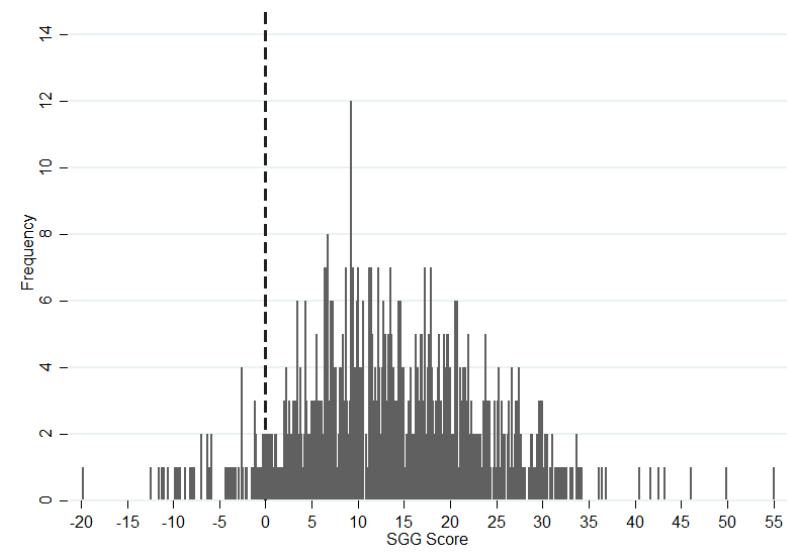(a) Density of SGG Assignment Variable, Fuzzy RD
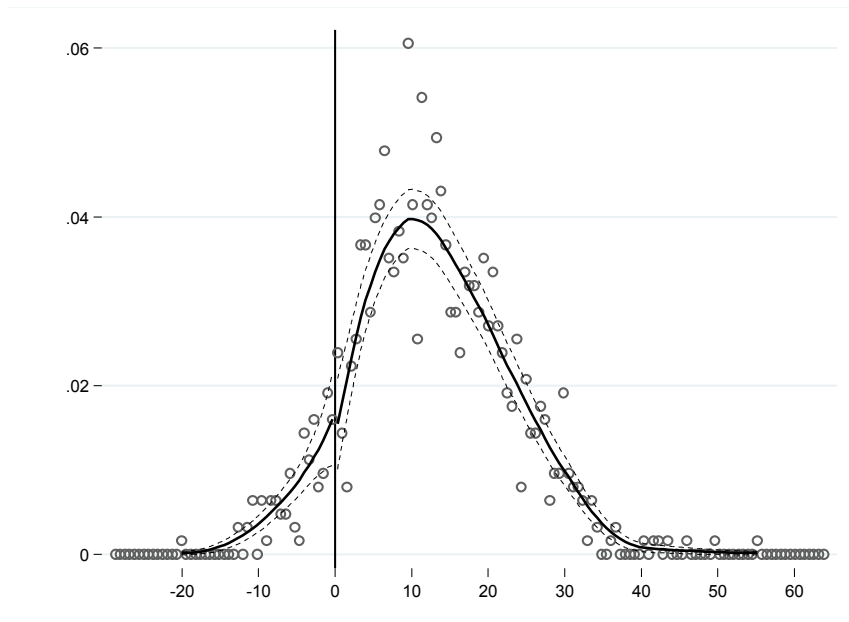
(b) Density of SGG Assignment Variable, Frontier RD

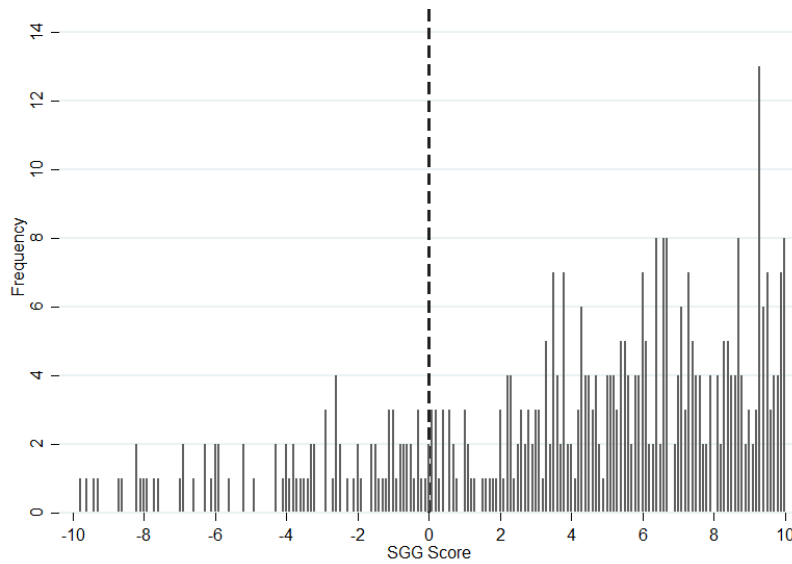Appendix Figure 2. Density of Assignment Variable

NOTES: Assignment variable centered at zero. McCrary density test discontinuity estimates: A(1) -0.181 (.276); and A(2) 0.019 (.344). Standard error in parentheses.
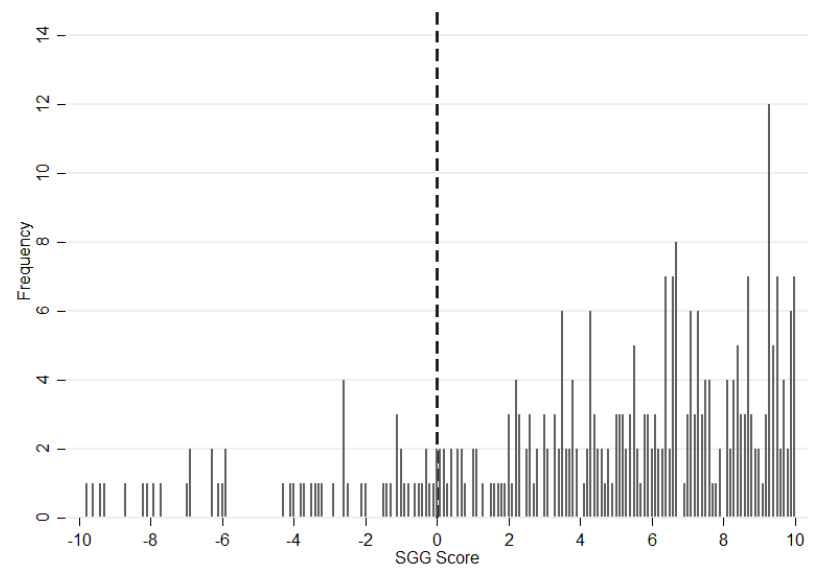
(a) SGG Histogram, Fuzzy RD

(b) SGG Histogram, Frontier RD

(c) SGG Histogram +/- 10, Fuzzy RD

(d) SGG Histogram +/- 10, Frontier RD

Appendix Figure 3. Histogram of Forcing Variable, by RD Sample

NOTES: Graphs contain histograms of the assignment variables for Focus School status, with the assignment variable centered at zero. Bin width 0.1, figures (1) and (2) includes the full sample, figures (3) and (4) includes bandwidth of approximately +/- 1 SD. See Table 1 for exact values.