

NBER WORKING PAPER SERIES

DIFFERENTIATED ACCOUNTABILITY AND EDUCATION PRODUCTION:  
EVIDENCE FROM NCLB WAIVERS

Steven W. Hemelt  
Brian Jacob

Working Paper 23461  
<http://www.nber.org/papers/w23461>

NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
June 2017

The research reported here was supported by grants from the Spencer, William T. Grant, and Walton Family Foundations. This research used data structured and maintained by the Michigan Consortium for Educational Research (MCER). MCER data are modified for analysis purposes using rules governed by MCER and are not identical to those data collected and maintained by the Michigan Department of Education (MDE) and/or Michigan's Center for Educational Performance and Information (CEPI). Results, information, and opinions solely represent the analysis, information, and opinions of the author(s) and are not endorsed by, nor do they reflect the views or positions of, grantors, MDE and CEPI or any employee thereof. We are grateful for helpful comments and suggestions from seminar participants at Northwestern University, Brigham Young University, Harvard University, University of North Carolina at Charlotte and conference participants at the 2016 meetings of the Association for Education Finance and Policy (AEFP) in Denver, CO. We thank Elizabeth Mann, Max Kapustin, and Pieter De Vlieger for excellent research assistance. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2017 by Steven W. Hemelt and Brian Jacob. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Differentiated Accountability and Education Production: Evidence from NCLB Waivers  
Steven W. Hemelt and Brian Jacob  
NBER Working Paper No. 23461  
June 2017  
JEL No. I20,I21,I28,J01,J08

**ABSTRACT**

In 2011, the U.S. Department of Education granted states the opportunity to apply for waivers from the core requirements of No Child Left Behind (NCLB). In exchange, states implemented systems of differentiated accountability in which they identified and intervened in their lowest-performing schools (“Priority” schools) and schools with the largest achievement gaps between subgroups of students (“Focus” schools). We use administrative data from Michigan in a series of regression-discontinuity analyses to study the effects of these reforms on schools and students. Overall, we find that neither reform had appreciable impacts on various measures of school staffing, student composition, or academic achievement. We find some evidence that the Focus designation led to small, short-run reductions in the within-school math achievement gap – but that these reductions were driven by stagnant performance of lower-achieving students alongside declines in the performance of their higher-achieving peers. These findings serve as a cautionary tale for the capacity of the accountability provisions embedded in the recent reauthorization of NCLB, the Every Student Succeeds Act (ESSA), to meaningfully improve student and school outcomes.

Steven W. Hemelt  
Department of Public Policy  
University of North Carolina at Chapel Hill  
Abernethy Hall, Campus Box 3435  
Chapel Hill, NC 27599  
hemelt@email.unc.edu

Brian Jacob  
Gerald R. Ford School of Public Policy  
University of Michigan  
735 South State Street  
Ann Arbor, MI 48109  
and NBER  
bajacob@umich.edu

A online appendix is available at <http://www.nber.org/data-appendix/w23461>

## **I. Introduction**

One of the central goals of K-12 education reform is to improve the performance of chronically struggling schools and students. Test-based accountability has been the hallmark of such reform efforts for decades. In the early 2000s, the No Child Left Behind (NCLB) Act<sup>1</sup> brought test-based accountability to scale across the United States by requiring states to implement accountability systems that applied to all public schools and students. NCLB mandated annual testing in reading and mathematics of students in grades 3 through 8 and at least once in high school. States set annual goals for the share of students in the aggregate and in specific demographic subgroups that should demonstrate proficiency on state-created, high-stakes tests, with the proficiency target rising to 100 percent in the 2013-2014 academic year. Schools and districts that failed to meet the aggregate or subgroup benchmarks were subject to a series of escalating consequences under the federal law (Mills, 2008).

The effects of NCLB and similar accountability systems have been mixed, with evidence of improvement in student outcomes (Dee & Jacob, 2011; Deming et al., 2016) in tension with unintended negative consequences such as an overemphasis of tested material (Jacob, 2005; Rothstein, Jacobsen, & Wilder, 2008; Koretz, 2008) and cheating by teachers (Fausset & Blinder, 2014; Jacob & Levitt, 2003).<sup>2</sup>

Over the ensuing decade, many states and districts grappled with the increasingly unrealistic expectation that all students demonstrate proficiency by 2013-2014 and scores of schools fell into deeper sanctions as Congress failed to make changes to the legislation. In the fall of 2011, Secretary Duncan and President Obama articulated a process by which states could

---

<sup>1</sup> Prior to the recent passage of the Every Student Succeeds Act (ESSA), NCLB was the operational, governing reauthorization of the Elementary and Secondary Education Act (ESEA), the key piece of federal legislation that applies to K-12 schooling in the United States.

<sup>2</sup> For a more comprehensive assessment of these policies, see Jacob (2017), Ladd (2017), and Figlio and Ladd (2015).

request relief from some of NCLB's provisions.<sup>3</sup> In exchange, states had to implement reforms that aligned with a set of principles detailed by the Obama Administration. The waiver process targeted some of the most widely criticized components of NCLB in an attempt to “[allow states] to move beyond the one-size-fits-all mandates of NCLB, to be more innovative, and to engage in continued improvement in ways that benefit educators and students” (Arne Duncan, U.S. Department of Education, 2014a). As of fall 2015, 43 states had approved waiver plans in action.<sup>4</sup>

In exchange for receiving a waiver, states were required to implement reforms targeted at their lowest performing schools (“Priority schools”) as well as schools with the largest achievement gaps (“Focus schools”). For Priority schools, the required reforms are quite prescriptive and demand multiple changes to school organization and instructional practice consistent with elements of school turnarounds. The motivation behind such interventions is that chronically struggling schools face myriad challenges that require extensive and multi-pronged reforms, rather than marginal or targeted actions. Suggested interventions for Focus schools are less clearly defined at the federal level, reflecting a belief such schools require external incentives to focus attention and coordinate action on achievement gaps, but do not require prescriptive guidance on effective strategies for ameliorating such gaps.

Proponents of these reforms argue that states will be able to craft and implement differentiated accountability initiatives that better address the varied and multifaceted problems facing chronically low-performing schools as well as the specific barriers that depress the

---

<sup>3</sup> Common concerns about NCLB from education leaders in states that applied for waivers include mandating consequences for schools that did not always lead to improved student achievement, setting an unrealistic goal of 100 percent student proficiency by 2014, stifling progress with overly prescriptive sanctions, inaccurately identifying schools as “needing improvement,” overemphasizing standardized tests, and creating disincentives for states to set ambitious standards (CEP, 2013, p. 11).

<sup>4</sup> U.S. Department of Education, “ESEA Flexibility,” Retrieved from <http://www2.ed.gov/policy/elsec/guid/eSEA-flexibility/index.html>.

achievement of certain subgroups of students in other schools.<sup>5</sup> Critics characterize the waiver initiative as overly prescriptive, inflexible, unfunded, and an unwarranted extension of federal authority into local public schooling.

In this paper, we study the effects of Priority and Focus reforms on a range of educational outcomes. Research on the causal effects of these new reforms is scarce. These waiver-based reforms are one of the most direct, ambitious, and wide-ranging efforts in place to improve the performance and prospects of disadvantaged students across the United States. Our analysis centers on Michigan, a state with one of the clearest and most systematically applied procedures for identifying Priority and Focus schools. Michigan identifies Priority and Focus schools using a baseline measure of prior academic performance that is not manipulable at the time of identification. These sharp, discontinuous assignment rules enable us to examine the causal effects of Michigan's Priority and Focus school reforms through the use of a regression-discontinuity (RD) design. This design essentially compares subsequent outcomes of schools that performed just poorly enough to garner a Priority or Focus designation and be required to implement the associated reforms to schools that performed just well enough to be free of such requirements. We use administrative data from Michigan to provide one of the first comprehensive looks at the causal effects of Priority and Focus reforms on measures of achievement as well as candidate mechanisms such as staffing and student composition.

To preview our results, we find little to no effect of Priority status and associated reforms on measures of school staffing, composition, or academic achievement. In fact, we find some evidence of negative achievement effects, especially for students in the upper part of the within-school achievement distribution. For Focus schools, we find some evidence that the designation

---

<sup>5</sup> One implication of findings from some of the early state-specific studies on the effects of NLCB's accountability provisions on academic performance was to differentiate approaches and supports to schools based on the manner in which they failed to meet proficiency benchmarks (e.g., Hemelt, 2011).

led to small reductions in within-school math achievement gaps – but any such reductions were short-lived and likely driven by the stagnant performance of lower-achieving students alongside declines in the performance of their higher-achieving peers.

The paper proceeds as follows. In the next section, we briefly synthesize overarching findings from the literature on whole-school reform and the recent, rapidly developing literature on school turnarounds, and then describe the “reform principles” that guided waiver implementation. Section III describes the Michigan policy context along with details on Michigan’s waiver-based reforms. Section IV details our primary empirical approach. Section V describes the data we employ. Section VI presents the main findings related to Priority interventions. Section VII discusses findings concerning Focus schools. Section VIII explores results from a dynamic regression-discontinuity approach. Section IX concludes and offers policy implications of our main findings.

## **II. Literature Review**

In this section, we briefly summarize earlier work on whole-school reform. We next discuss prior studies on the recent set of school turnaround efforts that started shortly after the Great Recession. Finally, we outline the differentiated approach to accountability instituted by the NCLB waiver process.

### *A. The Challenges of Whole-School Reform*

Whole-school reforms are notoriously difficult to implement well. A prominent example of federal efforts to promote whole-school reform began in 1998 with the introduction of the “Comprehensive School Reform Demonstration” (CSR D) program. This initiative provided three-year grants to schools that could then be used to purchase the services of independent, school-reform developers using research-based designs. This grant program developed into a

“leading strategy” of the U.S. Department of Education between 1998 and 2005 (Gross, Booker, & Goldhaber, 2009) when nearly \$2 billion were distributed to roughly 6,700 schools to support Comprehensive School Reform (CSR) efforts. Many additional schools also implemented CSR reforms using non-federal funding sources. The U.S. Department of Education defined CSR models as consisting of 11 key components with a prominent emphasis on the use of “scientifically based” teaching and management methods and the school-wide integration of instruction, assessment, professional development, and school management (U.S. Department of Education, 2010).

The evaluation evidence on the achievement effects of CSR is somewhat mixed. A federally sponsored evaluation concluded that CSR schools did not demonstrate larger achievement gains than comparison schools up to five years after receiving the award (U.S. Department of Education, 2010). Similarly, a recent study of CSR awards in Texas (Gross, Booker, & Goldhaber, 2009) concludes that CSR awards led to only modest achievement gains among white students (0.04 effect size) and no detectable effects among minority students.

However, these studies acknowledge that they focus on the effects of receiving CSR funding rather than on the effects of the highly diverse set of CSR reform efforts. Other evidence suggests that the quality of CSR implementation was uneven in ways that mattered for sustaining school improvement (Desimone, 2002; Bifulco, Duncombe, & Yinger 2005; U.S. Department of Education, 2010). A meta-analytic review (Borman, Hewes, Overman, & Brown, 2003) that considered the efficacy of specific CSR models characterized three (i.e., Direct Instruction, Success for All, and the School Development Program) as having the “strongest evidence of effectiveness” in terms of the comparative quality and quantity of evidence suggesting

meaningful impacts on student achievement. However, Borman et al. (2003) also suggest that CSR is more likely to have positive impacts when implemented over several years.

### *B. Literature on Recent School Turnaround Efforts*

There is a growing literature on the recent school turnaround efforts that originate as part of accountability schemes. To date, this research has yielded quite mixed results. In several cases, reform efforts have led to improved student performance; in other instances, schools subject to turnaround made no progress or experienced declines in achievement. This is perhaps not surprising given the widely varying nature and context of such reforms. A review of this research can help shed light on potential of the new NCLB-mandated reforms.

Several studies find that state-led turnaround efforts in Massachusetts have been quite successful. Using a difference-in-differences approach, Papay (2015) finds that traditional public schools in Massachusetts that participated in state-led turnaround efforts experienced significant gains in student performance. Schueler, Goodman, and Deming (2017) study the state's efforts to turn around performance in the Lawrence (Massachusetts) school district. Using a differences-in-differences framework, they find that the state-takeover of the Lawrence Public Schools resulted in substantially higher math achievement and moderately higher ELA achievement.

Turnaround efforts in Tennessee have been less successful. Zimmer et al. (2016) study school turnarounds in Tennessee that were implemented as part of the state's Race to the Top grant, which included both state takeovers as well as district-managed reform efforts. They note that all reform models undertaken involved a change in school leadership and a substantial turnover of teachers. Using a difference-in-differences approach, the authors find that schools chosen for district-led turnaround efforts experienced significant achievement gains while the schools taken over by the state or operated by charter management organizations saw no



improvement. They note that the three districts that led school turnarounds provided considerable support to the target schools, which included offering substantial raises to teachers who remained in or transferred to a target school.

In other cases, individual districts have pursued a more tailored approach to school turnaround. These reforms differ from the more recent type of turnarounds attempted under NCLB waivers and ESSA, not only in that schools are more selectively chosen but also because more resources are typically devoted to the reform effort. While there is some evidence that such reforms have been successful (see Player and Katz, 2013 and de la Torre et al., 2013), these cases are not amenable to more rigorous research designs.

Strunk et al. (2016) present a similarly mixed assessment of turnaround efforts in Los Angeles. Using a comparative interrupted time series approach, they find no sign of improvement among the first cohort of schools (n=14), improvement in reading but not math for the second cohort of schools (n=5), and declines in student performance for the third cohort (n=9). They attribute the success of the second cohort to greater assistance on the part of the district and greater reliance on school restarts and reconstitutions, and note that there were substantial implementation problems with the third cohort.

The use of external operators (e.g., CMOs) to manage failing schools has also met with mixed results. Gill et al. (2007) examined reform efforts in Philadelphia in which the state transferred the operation of 45 low-performing schools to education management organizations. They find no impact on student performance despite the fact that these schools were provided with extra resources. On the other hand, Ruble (2015) finds positive effects of a similar reform strategy in New Orleans, where the state transferred operation of failing schools (both traditional and charter) to non-profit school management organizations.

Several prior studies have used a regression-discontinuity approach to study the impact of turnaround efforts, comparing schools just above and below the cutoff used to identify schools for reform. Mathematica Policy Research examined school turnaround reforms adopted in 2011-2012 and 2012-2013 as part of the federal School Improvement Grant (SIG) initiative. The study covered more than 400 schools across roughly 60 districts in 22 states (Dragoset et al., 2017). The report concludes that the implementation of a SIG-funded reform model had no impact on math or reading achievement, high school graduation, or college enrollment. The authors note modest differences in the use of recommended practices in treatment schools, although these differences were not statistically significant in the regression-discontinuity analysis. An earlier study of SIG-funded reforms in California found that they did improve student performance, particularly among schools that engaged in the most substantial reforms such as replacing the principal and at least 50 percent of staff (Dee, 2012). Consistent with these results from California, the national study found suggestive evidence in favor of the more severe “turnaround” model in contrast to the “transformation” model.

Three other regression-discontinuity studies find such reforms resulted in zero or even negative achievement effects. Studying turnaround efforts in North Carolina implemented as part of the state’s Race-to-the-Top application, Heissel and Ladd (2016) find that students in the affected schools actually experienced a decline in math and reading scores. Interestingly, most schools in this study opted for the less severe transformation model of reform. The authors find that teachers in reform schools reported spending more time on professional development collaborative planning, but also reported more time on administrative duties (e.g., meetings and reporting) not relevant for instruction.

Dougherty and Weiner (2017) use a regression-discontinuity design to study Priority school reforms in Rhode Island. In the first two years, they find that schools just above the cutoff do no differently in math or reading achievement than schools just below the cutoff. They find suggestive evidence that the lowest performing schools in the state have even lower subsequent performance than slightly higher-performing but still sanctioned schools. Finally, Saw et al. (2016) find little impact of school turnaround efforts in Michigan that preceded NCLB waivers, though their analysis is limited to high schools in a single year.

### *C. Differentiated Accountability under NCLB Waivers*

Secretary Duncan announced the opportunity for states to apply for waivers from NCLB's main requirements in the fall of 2011. There were three submission windows spanning from September of 2011 through September of 2012. States' waiver plans were assessed according to the degree to which they adhered to four key "reform principles" articulated by the U.S. Department of Education (2012; CEP, 2013). First, states had to adopt "college- and career-ready standards" in at least math and reading,<sup>6</sup> align those standards to "high-quality" assessments, and then use those assessments to measure growth in student achievement over time and hold schools accountable. Second, states were required to craft and adopt "differentiated recognition, accountability, and support systems." Specifically, states had to propose how they would identify the lowest-performing schools in their states ("Priority" schools) and the schools that had the lowest performance for certain subgroups or the largest gaps in performance between subgroups ("Focus" schools).<sup>7</sup> States were then required to field different interventions

---

<sup>6</sup> The aim of the Common Core State Standards Initiative (CCSSI) is to develop a set of standards that defines the knowledge and skills students should acquire during their K-12 educational careers so that they will graduate "college and career ready." These standards are intended to provide parents and educators with a sense of students' academic progress, regardless of state or locality.

<sup>7</sup> In December of 2015, Congress finally passed a reauthorization of the ESEA entitled the "Every Student Succeeds Act" (ESSA; Public Law 114-95). The law will take full effect starting in the 2017-18 academic year. Broadly, the law aims to provide states and districts with more flexibility (relative to NCLB) in crafting and implementing

in these groups of schools. Third, states had to design and implement systems to evaluate and support teachers and principals. Finally, states were required to evaluate their reporting requirements and ditch duplicative, unhelpful ones.

Priority schools are among the lowest-performing schools in their states (i.e., the bottom 5% of all Title I schools based on state-determined measures of prior performance). Federal guidance requires Priority schools in all states to implement multiple initiatives that are consistent with federal “turnaround principles.” Acceptable approaches closely mirror those required of schools that received School Improvement Grants (SIGs) funded by the 2009 stimulus package. That is, Priority schools must implement a specific reform approach: transformation, turnaround, or restart (or otherwise close). These reforms emphasize new leadership, teacher evaluations, professional development, and instructional changes that rely on the continuous, formative use of data, as well as extended learning time.

Focus schools must constitute at least 10% of Title I schools in a waiver state. Focus schools (and their encompassing districts) are required to implement unspecified interventions to reduce achievement gaps, which “may include tutoring and public school choice” (U.S. Department of Education, 2012, p. 2). Thus, we expect ample heterogeneity in the types of Focus interventions adopted, in contrast to the limited set of federally specified Priority-school interventions shared across all waiver states.

---

accountability policies for schools and evaluation systems for teachers. Within these state systems, ESSA requires states to identify the lowest-performing schools (i.e., the bottom 5 percent) and to intervene to improve performance. Thus, some of the key elements of the waiver reforms will continue during the implementation of ESSA – but with greater flexibility for states in determining the interventions to field in their chronically struggling schools.

### III. Policy Implementation Context in Michigan

Michigan's waiver application was approved in July of 2012,<sup>8</sup> and Michigan identified its first set of Priority and Focus schools in August of 2012. For this group of schools, the 2012-13 year served as a planning year during which schools flagged as either Priority or Focus consulted with district- and state-level supporters in order to craft plans of action consistent with federally required intervention components. These plans were then implemented during the 2013-14 academic year.

#### *A. Priority and Focus Assignment in Michigan*

Michigan adopted a cohort model of Priority and Focus school identification: once identified as a Priority or Focus school, a school remains designated as such for four years, with the first year being the planning period noted above. In practice, many schools began implementation efforts immediately.

In Michigan, schools were identified as Priority, Focus, or Reward using a “top-to-bottom” (TTB) ranking, in which each school received a percentile score that located it in a statewide performance distribution of schools. In Appendix B, we provide a detailed description of the steps involved in calculating the TTB index as well as the manner in which the state used this index to identify Priority and Focus schools. In broad terms, the TTB score was a weighted function of subject-specific achievement measured in three ways: level achievement, growth in achievement, and the within-school gap between the top 30% and bottom 30% of students – all based on two to four years of prior data.

Schools were arrayed from lowest (i.e., the worst performing schools) to highest by their overall TTB score (regardless of level). The Michigan Department of Education (MDE) included

---

<sup>8</sup> A complete list of all states with approved or pending waiver applications (as well as the full applications) can be accessed through the U.S. Department of Education: <http://www2.ed.gov/policy/elsec/guid/esea-flexibility/index.html>.

both Title I and non-Title-I schools in its rankings. Priority schools were identified as schools in the bottom 5% of the TTB ranking distribution.<sup>9</sup> The Priority school accountability designation replaced Michigan’s previous label of “persistently low-performing” (PLA) schools,<sup>10</sup> which was solely based on math and reading scores, and divided by school level (e.g., elementary, middle and high) – making it more difficult to array all schools in Michigan along a more comprehensive index of achievement. Reward schools were those that constituted the top 5% of the TTB distribution.<sup>11</sup>

Focus schools were identified using one particular component of the TTB ranking: the achievement-gap score. Within each school and tested subject, MDE calculated the gap in performance as the difference between the average score of the top 30% of students and the bottom 30% of students. Next, MDE averaged these subject-specific gap values to arrive at a “composite gap index” for each school. Based on this composite measure, a stock of schools that included 10% of non-Priority, Title I schools with the largest gaps were labeled Focus schools.<sup>12</sup> Focus schools appear at all points in the TTB ranking list of schools – and are thought of as having particular difficulty in supporting the specific needs of their lower-performing students, even though the school as a whole may be achieving at an adequate (or superior) level.

---

<sup>9</sup> Federal guidance required states to identify the bottom 5% of Title I schools in the state as Priority schools. Michigan wanted to identify all low-performing schools as Priority, regardless of Title I status. In the case of Priority schools, the number of Title I schools that fell below the 5<sup>th</sup> percentile Priority cutoffs in 2012, 2013, and 2104 was equal to or greater than 5% of the total population of Title I schools in Michigan in those years. Thus, Michigan did not have to move farther up the TTB ranking distribution to capture 5% of the population of Title I schools. See Appendix B for additional details.

<sup>10</sup> Any school that (a) was identified in 2010 or 2011 as a PLA school or (b) received SIG funds to implement a turnaround model was automatically identified as a Priority school in August of 2012.

<sup>11</sup> Michigan also identified two other types of Reward schools: (a) the top 5% of schools making the greatest gains in achievement and (b) schools determined to be “beating the odds” (i.e., those outperforming other schools with similar socio-demographic makeups). For more details on Reward schools, please consult <http://www.michigan.gov/rewardschools>.

<sup>12</sup> Once again, federal guidance required states to identify 10% of the population of Title I schools in a state as Focus schools. Michigan wanted to identify all schools with large within-school achievement gaps as Focus, regardless of Title I status. Thus, in each ranking year, Michigan found the standardized gap value below which 10% of the stock of Title I schools would be identified as Focus schools. All non-Title I schools that fell below that value were also labeled as Focus schools. See Appendix B for more details.

Schools cannot be labeled as both a Priority and a Focus school. If a school meets the criteria for both Priority and Focus designations, the Priority label takes precedence. Though Priority and Focus schools remain in a “treated” status for four years, schools that just barely missed being placed in Priority or Focus groups in one year could fall below a subsequent year’s cutoff and become treated.<sup>13</sup> The dynamic nature of a school’s treatment status informs our empirical approach, which we discuss in our methodology section below.

*B. Treatment Components of Priority and Focus Designations in Michigan*

Both Priority and Focus schools receive supports from the state. All Priority schools must develop a reform/redesign plan based on one of the four intervention models established by the U.S. Department of Education: transformation, turnaround, restart, or closure.<sup>14</sup> As of 2015, 60 percent of all Priority schools had chosen the least drastic “transformation” option (J. Doll, personal communication, January 27, 2017). This option requires a school to replace the principal,<sup>15</sup> implement unspecified instructional reform strategies, and extend or repurpose time for additional professional development or teacher planning. The turnaround plan is more severe in that it also requires that schools release and rehire up to 50% of their staff, implement a new governance structure, and implement a new or revised curriculum. The restart option is the most drastic choice (short of permanent closure), as it involves formally closing a school and releasing all teachers and administrators (though some could be hired back by the new school).

---

<sup>13</sup> In addition, since Priority status takes precedence, if a Focus school identified in 2012 (2013) fell below the Priority cutoff in 2013 (2014), it was moved into Priority status and subject to the associated requirements (K. Ruple, Michigan Department of Education, personal communication, June 10, 2016).

<sup>14</sup> By state law (MCL 280.1280c), all schools designated as Priority must submit a redesign plan that addresses one of four federal intervention models identified by the U.S. Department of Education to Michigan’s School Reform Office (SRO) for approval, regardless of Title I status. Priority schools that are Title I receive additional assistance from the state.

<sup>15</sup> Schools that adopted the “transformation” model could retain the current principal if she had been in the role for two or fewer years.

Title I Priority schools must set aside 10% of their building-specific Title I funding to support the implementation of the school’s reform plan.<sup>16</sup> Title I Priority schools also work with their district to assemble a “School Support Team” composed of a school improvement facilitator assigned by the state, a district representative, and an intervention specialist (trained and assigned by the School of Education at Michigan State University).<sup>17</sup> All Priority schools participate in “diagnostic dialogues”<sup>18</sup> with stakeholders in which achievement data are examined to determine relevant changes in the school’s teaching and learning practices.

All Focus schools must undergo district-led, school-specific “data dialogues” to identify district-level system changes needed to help Focus schools make substantial progress toward closing achievement gaps. Focus schools receive a “District Toolkit” which outlines practices, tools, and strategies that have been successful in helping other districts improve struggling schools. The “data dialogues” are intended to generate discussion anchored on local achievement information that will lead to locally appropriate, customized changes in learning practices, curricula, and teaching necessary to improve student outcomes.

Title I Focus schools receive a “District Improvement Facilitator” (DIF) who works with the central office staff (at the district level) about 40 hours per week during the school year to provide implementation assistance.<sup>19</sup> Specifically, the DIF helps each Focus school identify one or two major changes in such practices to implement during the academic year. The DIF also leads changes in any district-level policies or practices necessary to support the school-level

---

<sup>16</sup> In addition, the district in which a Title I Priority school resides must set aside 20% of the district-level Title I funding to support changes in its Priority schools.

<sup>17</sup> The state funded the intervention specialist (at the district level) for 50 days and the school improvement facilitators (at the school level) for 40 days (J. Doll, personal communication, January 10, 2017).

<sup>18</sup> For a detailed treatment of the questions and structure that guide these data dialogues, please consult the document “Data Dialogue Booklet” on the following MDE website: <http://www.michigan.gov/focusschools>.

<sup>19</sup> DIFs were initially funded and trained by the School of Education at Michigan State University. In subsequent years, the state increased funding to Michigan’s Intermediate School Districts (ISDs) by about \$8 million through regional assistance grants. ISDs were then responsible for finding and hiring DIFs for any of their districts that contained Title I Focus schools (V. Keesler, personal communication, March 18, 2016).



changes. Finally, Title I Focus schools in year two of Focus status are required to allocate a share of their Title I building-specific funding toward implementation (with districts required to set aside additional district-specific Title I funds for schools in year three of Focus identification).<sup>20</sup>

Though Priority and Focus schools are assigned to cohorts for four years, Michigan decided that if a school made consistent progress over the first two to three years of its cohort membership, it could be released early. Criteria for early release differed by Priority and Focus reforms. Priority schools had to maintain a TTB ranking of at least 15 for two consecutive years after initial identification and meet all other testing stipulations (e.g., testing participating rate of at least 95 percent). Release criteria for Focus schools were less well specified. The state looked at growth and gaps in math and reading performance to make its early release decisions for Focus schools (J. Doll, personal communication, January 27, 2017).

#### **IV. Empirical Approach**

We use a regression-discontinuity (RD) design to study Priority- and Focus-school interventions in Michigan. This RD approach leverages comparisons of schools that are just above and below the threshold values of the ranking variables used to identify Priority and Focus schools. In this section, we discuss the basic issues underlying a standard RD analysis, which we then use to examine effects of Priority and Focus designations by cohort (2012, 2013, and 2014). We then raise some of the complexities that arise from the dynamic nature of these treatments

---

<sup>20</sup> These funds are most commonly used to support targeted professional development concerning the implementation of a multi-tiered system of support within the school, with a focus on low-achievers; or to provide space and time for weekly (or even daily) teacher collaboration (MDE, Office of Improvement and Innovation, Focus School Technical Assistance, 2013). Whether Title I Priority and Focus schools were required to offer (and pay) for the option of students to attend a different, non-Priority (or non-Focus) school depends on the academic year. In 2012-2013, Title I Priority and Focus schools were required to offer school choice and transportation, and to use Title I dollars to pay for transportation. From 2013-2014 onward, Priority and Focus schools were not required to offer choice and transportation.

across cohorts, laying the groundwork for our consideration of results from a dynamic regression-discontinuity approach later in the paper.

### *A. Regression Discontinuity in a Cross-Section*

As described above, school  $j$  is assigned to intervention if its performance ranking,  $r$ , falls below a specific cutoff:  $b_j = 1(r_j < r^*)$ . We are interested in the relationship between the intervention and some school outcome. Suppressing time-related subscripts, we can write:

$$(1) \quad y_j = \kappa + b_j\theta + u_j$$

Here,  $\theta$  is the causal effect of Priority or Focus status on the outcome. The identification concern is that schools below the cutoff have unobservable characteristics that are correlated with the outcome:  $E[u_j \neq 0]$ . However, as long as there is some random component that translates a school's performance into a ranking, if one focuses on schools sufficiently close to the cutoff, the comparison of treated and untreated units approximates a randomized experiment. The schools that just "miss" passing the cutoff and are thus assigned the intervention become our treatment group, while those that just "pass" the cutoff and avoid intervention serve as the control group.

There are two common approaches to estimating RD models. The nonparametric approach estimates local polynomial regressions on either side of the cutoff, relying on various algorithms to select the optimal bandwidth. The parametric approach includes a flexible polynomial of the running variable in an OLS regression to absorb variation from schools farther from the cutoff, as in the following equation:

$$(2) \quad y_j = \kappa + b_j\theta + P_g(r_j, \gamma) + u_j$$

where  $P_g(r_j, \gamma)$  is a polynomial in  $r$  of order  $g$ , with coefficients  $\gamma$ . The primary advantage of the parametric approach is that by including a wider range of data around the cutoff, it typically yields more precise estimates. In the analysis below, we demonstrate that parametric and nonparametric results yield comparable point estimates in the cross-section.<sup>21</sup> For the dynamic regression-discontinuity approach described below, we use a parametric model. Regardless of the estimation technique, a RD design should be interpreted as identifying the impact of a particular intervention on units close to the cutoff.

The key identifying assumption of any RD analysis is that the outcome is related to the running variable in a continuous fashion. This might be violated for several reasons. In some cases, one might be concerned that the actors themselves can manipulate the running variable such that those with particularly good (or bad) unobservable characteristics fall on one side of the cutoff. In practice, the algorithms used to rank schools for the purpose of Priority and Focus determination were quite complicated and, more importantly, depended on a school's relative position. For these reasons, we believe that it was not possible for schools to intentionally manipulate their performance ranking.

A more realistic concern is that state officials determined the algorithm with the intention of including or excluding specific schools from the intervention. We believe that this is unlikely for several reasons. The 5 and 10 percent thresholds were determined by federal guidelines, as was the mandate to incorporate performance levels and gaps to some extent. The components that the state chose, and the weights attached to those components, appear quite standard. For example, performance level, growth, and gaps received weights of 0.5, 0.25, and 0.25

---

<sup>21</sup> We employ the data-driven nonparametric commands developed by Calonico, Cattaneo, and Titiunik (2014a, 2014b, 2015).

respectively in the overall school performance metric, and all test scores in all available grades were included.

A final concern is that the threshold for one intervention may correspond to other treatments or policies. For example, if school funding formulas were related to school performance such that schools below the 5<sup>th</sup> percentile received discontinuously more (or less) funding, it would be impossible to disentangle the impact of Priority intervention from school funding. In this case, however, we know of no other state or district policy that used these cutoffs.

We test this assumption by estimating equation (2) with outcomes prior to the intervention, where we would expect  $\theta$  to be zero. We also implement a procedure developed by McCrary (2008) to determine whether there is any bunching in the distribution of the running variable around the cutoff, which might suggest manipulation of the running variable itself. As discussed below, both tests suggest that the key assumptions are met in our context.

### *B. Accounting for Treatment Crossover in a Cross-Section*

In any individual year, there is strict adherence to the cutoffs described above. That is, every eligible school scoring below the 5<sup>th</sup> percentile is assigned Priority status and every school scoring in the bottom 10 percent of the gap measure is assigned Focus status.<sup>22</sup> However, each year after the initial set of schools was announced in 2012, the state followed a similar procedure – calculating top-to-bottom rankings based on measures of student achievement level, growth, and gaps, and assigning schools to Priority or Focus status depending on their rank in the relevant distribution. All schools are included in the TTB rankings in each year, regardless of whether they had been labeled Focus or Priority in prior years. Because school performance and

---

<sup>22</sup> As noted above, there are a handful of schools that qualify for both Focus and Priority interventions, in which case they are assigned to the Priority category. For this reason, adherence to the Focus cutoff is not strictly perfect.

thus the TTB rankings change each year, additional schools can and do fall into the intervention categories.

For example, by Summer 2014, 124 (265) additional schools had been placed in the Priority (Focus) category. More importantly, many of the schools that narrowly missed being assigned Priority or Focus interventions in one year will be assigned in subsequent years. For example, roughly 30 percent of schools in the 6<sup>th</sup> to 10<sup>th</sup> percentile on the composite TTB ranking in 2012 had been assigned Priority status by Fall 2014. Roughly 40 percent of schools in the 11<sup>th</sup> to 20<sup>th</sup> percentile of the gap measure in the initial year had been assigned Focus status by Fall 2014.

This creates a situation which is often referred to as treatment crossover in the context of a cross-sectional RD analysis. To address this, we estimate a “fuzzy” RD (Hahn, Todd, & Van der Klaauw, 2001) within an IV framework. For a given cohort, the equation of interest relates the treatment,  $T$ , to the outcome,  $y$ , conditional on the flexible polynomial of the running variable:

$$(3) \quad y_j = \kappa + \hat{T}_j \theta + P_g(r_j, \lambda) + \varepsilon_j$$

And the first-stage models the treatment as a function of the indicator for falling below the cutoff, along with the flexible polynomial.

$$(4) \quad T_j = \rho + b_j \alpha + P_g(r_j, \lambda) + v_j$$

The estimate of  $\theta$  from equation 3 captures the effect of the “treatment-on-the-treated” whereas the estimate of  $\theta$  from equation 2 captures the effect of the “intent-to-treat.” The estimate of  $\alpha$  above is the first-stage effect of missing the cutoff in year  $t$  on the likelihood of experiencing the Priority (Focus) intervention in a subsequent year.

In the context of a cross-section model, if one is willing to assume that the impact of the intervention increases linearly with time, then one can define the treatment,  $T$ , as the cumulative years of intervention (i.e., the number of years since the school was first identified as Priority or Focus). If one believes that the full impact of the intervention occurs immediately, then one would define  $T$  as “ever identified” by year  $t$ . If one believes that it takes two years for the intervention to take effect, then one could define  $T$  as being identified at least two years earlier.

### *C. Other Concerns with Internal Validity*

In addition to the mechanics of estimation, the fact that the “control” schools in one year might become “treatment” schools in a subsequent year raises the issue of whether merely missing the intervention has an independent impact on school performance. For example, one might imagine that the fear of being sanctioned in subsequent years might motivate teachers and administrators to take actions to affect performance in the short-run. These “threat effects” could lead us to understate the impact of the intervention itself. To explore this possibility, we examine the time trends of various outcomes in treatment schools as well as sets of comparison schools. If we find positive trends in student outcomes following a school’s identification as a Priority (or Focus) school, but that this occurs among observationally similar comparison schools in the same cohort, we might be concerned about the threat effect. As discussed below, we do not find any compelling evidence that threat responses among comparison schools were driving the impact estimates.<sup>23</sup>

---

<sup>23</sup> This analysis resembles a comparative interrupted time series design. A key assumption of this analysis is that it requires the existence of a set of comparison schools that were (a) far enough from the cutoff that they did not experience a threat effect but (b) close enough to the cutoff that their trajectory reflects the counterfactual path for treatment schools.

A second potential concern involves school closures. One of the four approved reform models for priority schools involved closing the school, and dispersing students and teachers to other nearby schools. Because our primary analysis follows schools over time, if treatment schools are more likely to close, our sample may suffer from differential attrition that could introduce bias. In Appendix Table A3, we present results from an analysis of school closures by cohort. We see some evidence that schools labeled as Priority were more likely to close and this pattern is most pronounced among schools in the cohort of 2012: about 39 percent of initial Priority schools in our preferred analytic sample for 2012 had closed by the 2015-2016 year, compared to 18 percent of schools that just barely missed Priority designation in 2012. The differentially higher closure rate of Priority schools is much less pronounced in later cohorts. Very few Focus schools close during our period of analysis.

It turns out that the lion's share of 2012 Priority schools that closed did so during the summer of 2012. The Priority schools were announced in August 2012, so it seems as if these closure decisions were determined prior to or concurrently with the priority determination. For this reason, we limit our analysis sample to schools that were open during the 2012-2013 school year. When examining how the baseline (pre-treatment) characteristics of schools change through the cutoff, we limit the analysis to this sample as well. Any differentiation selection induced by these summer 2012 closures should show up in this analysis. We follow a similar approach in our analysis of the 2013 and 2014 cohorts.

A third concern is that the treatment itself might lead to changes in student or teacher composition in the school. For example, some parents might be reluctant to send their children to a Priority school because of the negative signal. On the other hand, if this designation signals future reforms, parents might choose to send their children to the school. If one views these types

of changes as part of the treatment, they should be thought of as mechanisms through which the intervention operates. This seems particularly appropriate for any staff changes generated by the intervention. In the analysis below, we study whether the intervention has any impact on student and teacher mobility or composition.

## **V. Data**

To identify Priority and Focus schools, we use publicly available data from the Michigan Department of Education (MDE). For each cohort, we also obtained the subcomponents that were used to construct the composite ranking variables. We use measures of school- and district-characteristics collected by MDE and the Common Core of Data (CCD), including measures of school expenditures, school type (e.g., charter, magnet, vocational, etc.), Title I status, and geographic location.

To measure student and teacher characteristics (including standardized test scores), we start with individual-level administrative data, which we aggregate to the school-year level. The use of individual data allows us to construct measures that are typically not provided in publicly available school-level data, such as the fraction of students or teachers that are new to a school in any particular year and various quantiles of student achievement in a school and year. Given that the fundamental unit of variation is the school-year, the use of aggregate data is more appropriate and produces virtually identical results as the use of individual-level data.

In recent years, Michigan has administered its standardized tests in the Fall of each academic year. Given that schools first learned about their Priority (Focus) status in August, it seems unlikely that they would have had an opportunity to implement any substantive reforms before the October testing. For this reason, we use achievement scores from the following Fall to measure the year 1 impact of the policy. When controlling for prior school characteristics,



however, we only use information from the pre-designation school year to avoid any possibility of contamination. In the 2014-15 school year, Michigan switched to a new assessment program (M-STEP), which is administered in the late spring (i.e., April and May) of the academic year. For the 2014 treatment cohort, we therefore use spring 2015 scores as year 1 outcomes and 2016 scores as year 2 outcomes. We consider Spring 2016 (Spring 2015) exams as year 4 (3) effects and Fall 2014 as the year 1 effect for the 2012 cohort. For the 2013 cohort, exams in Spring 2016 (Spring 2015) provide year 3 (2) effects.

In calculating student achievement measures, we standardize test scores by subject-grade-year. In order to examine whether the intervention had different impacts across the achievement distribution, we estimate something akin to quantile regressions. It is possible to estimate quantile regression models in an RD framework. However, estimates from a standard quantile RD model using student-level data will reflect two distinct phenomena: the impact of the intervention on the lowest-performing students *within* a school as well as the average effect on the lowest-performing schools. To better distinguish between these channels, we estimate several different specifications.

To examine the impact of the intervention within schools, we calculate the achievement level at different percentiles within the school as well as measures of within-school dispersion such as the standard deviation, inter-quartile range, or the state-defined achievement gap. We then proceed to use these measures as outcomes in a school-level RD specification like those above. We also estimate the impact of the interventions on the distribution of school mean achievement. That is, do the Priority or Focus reforms disproportionately help schools at the bottom versus top of the achievement distribution? To do so, we estimate simple quantile RD

models at the school level. In addition to standard school-level covariates, we also control for school-level averages of various student characteristics.

### *A. Sample*

In this analysis, we focus on schools serving students in grades 3-8.<sup>24</sup> The broad majorities of Priority and Focus schools are elementary or middle schools: 67 percent and 85 percent, respectively. We exclude charter schools from our analysis because, while they were subject to the same interventions, it is less clear that charter management organizations or charter authorizers enforced the requirements.<sup>25</sup> We also exclude schools that exclusively serve a special needs population, as they were typically not subject to the same mandates as other schools. For each cohort, we include schools that were open in the Fall immediately following the release of the Top-to-Bottom list and identification of new Priority and Focus schools. As discussed above, this will exclude schools closed during the summer when the announcement was made. For the 2013 and 2014 cohorts, we exclude schools that had been previously identified as Priority schools because they were not eligible for additional intervention.<sup>26</sup>

---

<sup>24</sup> We exclude schools that only serve students in grades K-2 because student test scores are not available for these schools, so in practice they were not eligible for Priority or Focus designation. We exclude schools that serve some elementary or middle school students that also serve high school students, including schools with grade configurations like 6-12 and K-12, because the Top-to-Bottom ranking for these schools was based in part on high school graduation rates. We do include a small number of schools that extend to grade 9 because the clear majority of students in these schools are in grades K-8 and high school graduation does not factor into the ranking decision for these schools. Finally, we drop 1 or 2 schools in each year with a total enrollment of fewer than 50 students because these were schools for special populations.

<sup>25</sup> Charter schools make up about 7 percent of all public, K-8, non-special education schools and roughly similar proportions of Priority and Focus schools.

<sup>26</sup> We exclude schools that were identified as “persistently low achieving” (PLA) in 2010 or 2011. The PLA label was the precursor to Priority schools. PLA schools represented the bottom 5 percent of public schools in Michigan based on prior math and reading test scores. These schools were publicly identified, put under the management of the School Reform Office (SRO), and required to design and implement a three-year plan to boost student performance. Finally, we exclude schools that were a part of the first two cohorts of SIG schools as these cohorts preceded the NCLB waiver intervention period. Priority schools in Michigan were encouraged to apply for SIG funding during the third and fourth cohorts, and thus SIG funding during these cohorts constitutes a mechanism through which any effect of Priority designation could operate. Indeed, Table 1 shows that 13 percent of Priority schools (and zero Focus schools) in the 2012 cohort won SIG funding.

Table 1 presents descriptive statistics for the sample of schools in the 2012 analysis. (Appendix Tables A2 and A3 present analogous statistics for the 2013 and 2014 cohorts.) Out of the full group of public K-8 schools in our analytic sample for the cohort of 2012, about 3 percent were labeled Priority schools and roughly 14 percent Focus schools. The share of black and Hispanic children in Michigan’s public elementary and middle schools is about 22 percent, and over half of all students in those schools are considered economically disadvantaged.<sup>27</sup> Only 15 percent of Michigan’s K-8 public schools are located in urban settings, while over 45 percent and about 39 percent are located in suburban and rural settings, respectively.

The three panels allow the reader to compare the characteristics of Priority and Focus schools to the full analytic sample of K-8 public schools in Michigan. In terms of demographics, Priority schools are heavily concentrated in urban areas and serve much higher proportions of black and economically disadvantaged students: nearly 78 percent and 91 percent, respectively. In terms of staff characteristics, the share of teachers new to a school in a given year is twice as high in Priority schools compared to all public K-8 schools.

In terms of demographics, staff characteristics, and location, Focus schools look more like the typical K-8 public school in Michigan than do Priority schools. One difference is that the typical Focus school serves a smaller share of economically disadvantaged students relative to the average Michigan public K-8 school: 39 percent versus 52 percent. In addition, the average academic achievement of Focus schools is above the state average.

Broadly these descriptive characteristics confirm some general notions about Priority and Focus schools. Schools that receive a Priority label exhibit very low overall academic achievement and disproportionately serve minority and disadvantaged students in urban settings.

---

<sup>27</sup> This measure combines students who are eligible for free or reduced-price meals (FARM) with those who are migrants or homeless.

Focus schools appear to have acceptable (or even above-average) overall achievement levels and serve a collection of students that resembles the typical K-8 public school. But, the acceptable global performance of these schools obscures sizeable gaps in academic achievement between within-school subgroups of students.

## **VI. Findings for the Priority Intervention**

In the sections that follow, we consider the effects of the Priority designation and associated reforms on measures of staffing, composition, and academic achievement for each of our three cohorts of schools.

### *A. Exploration of RD Assumptions*

The ability to confidently interpret our findings as causal effects of Priority reforms turns on a few key assumptions embedded in our RD setup. We test these assumptions before discussing our main results.

One key assumption is that schools are unable to manipulate the assignment variable. In terms of Priority schools, this variable was an index of achievement levels, growth, and gaps based on several years of historical data – and thus very difficult to manipulate around some unknown threshold value. In Figure 1, we present the distribution of this running variable in 2012, 2013 and 2014, and test for any jump at the cutoff value using the McCrary (2008) test. We find no evidence of bunching near the cutoff suggestive of running variable manipulation for the cohorts of 2012 and 2013. We see noisy evidence of possible bunching to the passing side of the cutoff for the 2014 cohort.

The second key assumption is that in the neighborhood of the cutoff, we can consider schools to either side as equivalent in all observed and unobserved respects except one: the receipt of the Priority label and attendant intervention requirements. If true, we should see few

differences in observable baseline characteristics of Priority and non-Priority schools across the cutoff.

Figure 2 shows scatterplots of various baseline (i.e., pre-intervention) school characteristics by the school performance index (SPI), where the cutoff is re-centered on zero. Hence, all schools with a negative SPI were labeled priority and no schools with SPI greater than zero were labeled in that first year. To highlight changes around the cutoff, we limit the figures to schools with SPI values within  $\pm 1$  of the cutoff. Notes at the bottom of each figure report estimates of  $\theta$  from equation 2, using both the nonparametric approach of Calonico et al. (2014a, 2014b, 2015) as well as a parametric approach. For the parametric models, we limit the sample to  $\pm 0.25$  from the cutoff and include a linear term in the running variable which is allowed to differ on either side of the cutoff. The lines shown in the figure are the estimated regression lines from the parametric specification.<sup>28</sup>

We show results for four baseline characteristics, including prior average math and reading scores along with summary measures of student demographics and school structure. To create these summary measures, we regress baseline math and reading scores on the set of school characteristics shown in Table 1, and calculate predicted baseline math and reading scores. Looking across these graphs, it does not appear that there are any discontinuous changes in observable school characteristics at the cutoff for the first cohort of schools to be considered for the intervention. Appendix Figures A1 and A2 show comparable graphs for the 2013 and 2014 cohorts of schools. As noted above, once a school is identified as Priority, it is essentially out of

---

<sup>28</sup> Graphs, results, and conclusions are similar when we limit the data window to  $\pm 0.50$ , but the  $\pm 0.25$  data window better fits the data – both visually and as evidenced by the average bandwidth selected by the nonparametric calculations across various outcomes.

the risk-set for additional “Priority” treatment.<sup>29</sup> For this reason, these figures exclude any schools that have been previously identified as Priority schools. The number of new Priority schools below the cutoff in these later cohorts is very small, making it difficult to draw strong conclusions from these estimates. To control for any small baseline differences and increase the precision of our estimates, we include a wide range of school characteristics in our preferred specification, all of which are defined in the academic year prior to the Priority designation: shares of students who are female, black, Hispanic, and Asian; share of students in special education, share with limited English proficiency, share economically disadvantaged; average math and reading scores, school type (i.e., elementary, middle, magnet); school enrollment; average teacher experience, share of teachers new to the school, the share of teachers from very competitive colleges; teacher-student and aide-student ratios. In addition, we weight our parametric models by school-level enrollment. Unweighted results are similar in all respects and available upon request from the authors.

### *B. First-Stage Impact on Treatment*

Figure 3 illustrates the extent of dynamic treatment crossover discussed above. As with the earlier figures, schools are plotted with the SPI on the x-axis. Figures 3A and 3B show the probability that a school was identified for Priority intervention by 2013 and 2015, respectively. For the very first cohort of Priority schools identified in 2012, the relationship between the overall school index and Priority status is sharp: that is, as one passes the threshold value from right to left, the probability of being categorized as a Priority school jumps from 0 to 1. In Figure 3A, we see how the movement of some control schools in 2012 to the treatment side of the cutoff in 2013 mutes the relationship between a school’s initial (2012) rating index value and the

---

<sup>29</sup> Recall that a school that was identified as Focus in an earlier cohort remains in the “risk set” for possible Priority designation in subsequent years.

cumulative likelihood it is ever flagged for Priority intervention. Figure 3B illustrates that nearly 35 percent of schools that fell just to the control side of the initial 2012 cutoff had been identified as Priority schools by 2015.

### *C. Impacts on Academic Achievement*

Figure 4 provides graphical evidence regarding the impact of the Priority designation on math scores. Since any effects from large, whole-school reforms tend to take time to manifest, we focus on test score outcomes by cohort that are at least two years from initial designation. As above in our figures that showed baseline characteristics, each figure graphs the binned raw data along with the estimated regression line from our simple, linear, parametric RD specification. Below each figure are the point estimates from this basic parametric specification as well as the nonparametric RD estimator, neither of which includes control variables. In Table 2, we present estimates of intent-to-treat effects of Priority status on math and reading achievement from our preferred RD specification that adds baseline controls. Thus, comparing estimates from Figure 4 and Table 2 allows the reader to assess the degree to which covariates matter for our conclusions. In Table 2, we explore effects on achievement at the mean and at different points in the within-school achievement distribution. We test the sensitivity of our estimates of the effect of Priority status on math and reading scores to additional RD specifications that vary the data window and the inclusion of controls in Appendix Table A4.

Across the 2012, 2013, and 2014 cohorts, we find little to no evidence of improvements in average math or reading performance or subject-specific performance at different quantiles of within-school achievement distributions. If anything, we see some evidence that Priority designation further depressed short-run performance. For example, we find that students in Priority schools in the 2014 cohort were performing about 0.12 to 0.17 standard deviations below

their counterparts in barely non-Priority schools after one year in math and reading, respectively. These negative effects persist into the second year for reading. For math and reading, the negative achievement effects appear especially pronounced among the upper half of the within-school achievement distribution.

We can observe the most distal outcomes (four years out) for the cohort of 2012. Four years after initial designation as a Priority school, students in the lowest parts of the within-school achievement distribution are performing worse than their barely non-Priority counterparts in mathematics, pulling wider the within-school gap between the top and bottom 30 percent of students (i.e., the map gap score). We see little effect on reading performance by four years after initial designation as a Priority school for the cohort of 2012.

One caveat for all of our Priority estimates is that they are relatively imprecise. This imprecision is a natural consequence of the process used to identify Priority schools: by definition, they are a small group of chronically struggling schools. Estimation of causal effects within an RD framework exacerbates this imprecision since one focuses on a subset of observations close to the threshold that determines treatment.

In Appendix Figure A3, we plot performance trends for the 2012 cohort<sup>30</sup> alongside a comparison group that should be less sensitive to any threat effects but still function as a reasonable approximation of what the treatment group would have experienced absent treatment.<sup>31</sup> We see that performance declined over this period among treatment schools. This suggests that the null effects from our RD models are not driven by short-run, comparable increases among treatment and control groups, which is what one would expect in the case of

---

<sup>30</sup> We focus on the 2012 cohort for two reasons. First, the RD assumptions are mostly strongly met for this cohort. Second, we can observe the longest trajectory of outcomes for this cohort (i.e., up to four years after initial identification as a Priority school).

<sup>31</sup> We define this comparison group as schools that have running variable values between 0.25 and 0.50, which correspond roughly to the 14<sup>th</sup> and 20<sup>th</sup> percentiles of the TTB rating distribution respectively.



threat effects. Further, the average math performance of Priority schools falls well below the performance of the comparison group by the third and fourth years after initial designation. This is consistent with some of our RD estimates that suggest negative effects of Priority status on medium-run measures of achievement.

#### *D. Impacts on School Staffing, Enrollment, and Composition*

In Table 3, we examine the effects of Priority status on measures of enrollment and staffing. We detect sizeable effects of Priority designation on total school enrollment for the first cohort of 2012. This bump in enrollment is sustained through 2016 and jump started by a notable increase in non-entry-grade enrollment. Recall that a large share of Priority schools closed immediately prior to the 2012-2013 academic year (see Appendix Table A3). Students from these schools had to be absorbed by other schools. We know that Priority schools in Michigan tend to be clustered in urban areas such as Detroit. Thus, the bump we observe in non-entry-grade enrollment in 2013 is consistent with the notion that Priority schools cluster together and frequently close. In complementary analyses not shown, we created measures of the number of schools that closed within three miles of each school during each year in our data as well as a measure of enrollment at those schools. If we introduce these variables as additional controls and drop the district of Detroit, we can “explain” about 60 percent of the initial and persistent increase in total enrollment. The remaining, unexplained part of the effect of Priority designation on enrollment suggests that some parents chose to move students into Priority schools. Perhaps parents in nearby schools viewed the Priority label as an indication of additional state involvement, help, and oversight with hope for improvement. We see little effect of Priority status on enrollment for schools in the 2013 and 2014 cohorts.

Almost all of the Priority school interventions required staff replacement and rehiring – of teachers, the principal, or both. We examine effects of Priority designation on teacher mobility and principal replacement. Overall, we find little to no effect of Priority status on these staffing measures. In the top panel of Table 3, we see some evidence that schools designated as Priority in the initial cohort of 2012 experienced greater shares of new teachers in the first post-treatment year, relative to barely non-Priority schools. Yet, this increase in new teachers can be explained by the enrollment boost discussed above. The effect of Priority designation on the share of teachers new to the school in 2014, 2015, and 2016 is close to zero and statistically insignificant. We depict this finding graphically in Figure 5. Further, we do not find evidence of such increases in the share of new teachers as a consequence of Priority status for the cohorts of 2013 and 2014. Our estimates of the effect of Priority designation on whether the principal of the school is new are very noisy, but positive in magnitude for two of the three cohorts (2013 and 2014). We find no effects on teacher-student or aide-student ratios for any of the three cohorts.

One interpretation of these results is weak implementation on the part of schools and districts. This interpretation is supported by a descriptive comparison of the share of teachers new to a school by intervention type. We obtained data from the School Reform Office (SRO) in Michigan on the intervention model (transformation, turnaround, restart, closure) chosen by each Priority school in the cohorts of 2012, 2013, and 2014. If Priority schools that chose the slightly more severe “turnaround” model were following the associated guidelines, the average share of teachers new to those schools in the year following Priority designation should be around 50 percent. For such schools in the cohort of 2012 ( $n = 10$ ), this share is 37 percent – a figure that is not much higher than the corresponding share among schools that chose the most flexible option, “transformation,” which carried no specific requirements regarding teacher and staff

replacement: 34 percent. Percentages for Priority schools in our analytic sample that chose the “turnaround” model in the subsequent cohorts of 2013 ( $n = 1$ ) and 2014 ( $n = 3$ ) look similar, though the number of newly identified Priority schools that chose this option is much smaller in those later cohorts.

In Table 4, we present estimates of several key measures of student composition. Despite the substantial increase in enrollment in the 2012 cohort, we do not see any large or consistent changes in the racial or socioeconomic composition of the schools. In the 2013 cohort, there appears to have been an initial, modest decrease in the fraction of economically disadvantaged students due to Priority designation. In the 2014 cohort, we see some evidence that Priority status reduced the share of Hispanic students enrolled in Priority schools, relative to barely non-Priority schools. Overall, however, most of the effects on measures of school composition are modest. We see no evidence that Priority designation increased the share of students in special education at Priority schools, relative to their barely non-Priority counterparts.

## **VII. Findings for the Focus Intervention**

We next turn our attention to Focus schools. As we did with Priority schools, we first test the key RD assumptions that allow us to interpret any subsequent differences between Focus and barely non-Focus schools as causal effects of Focus reforms. We then discuss our main findings concerning the effects of Focus designation on student achievement, school composition, and staffing on a cohort-by-cohort basis.

### *A. Exploration of RD Assumptions*

The first assumption is that schools are unable to manipulate the assignment variable. In terms of Focus schools, this variable was a measure of within-school gaps in performance between the top 30 percent and bottom 30 percent of students. In Figure 6, we present the

distribution of this running variable in 2012, 2013 and 2014, and test for any jump at the cutoff value using the McCrary (2008) test. We find no evidence of bunching near the cutoff suggestive of running variable manipulation for any of the three cohorts.

The second assumption is that in the neighborhood of the cutoff, we can consider schools to either side as equivalent in all observed and unobserved respects except one: the receipt of the Focus designation and associated intervention requirements. If true, we should see few differences in observable baseline characteristics of Focus and non-Focus schools across the cutoff. As we did with Priority schools, we show results for four baseline characteristics in Figure 7, including prior average math and reading scores along with summary measures of student demographics and school structure.

Looking across these graphs, we fail to find discontinuous changes in the observable school characteristics at the cutoff for the first cohort of schools to be considered for Focus interventions. Appendix Figures A4 and A5 show comparable graphs for the 2013 and 2014 cohorts of schools. As noted above, once a school is identified as Focus or Priority, it is essentially out of the risk-set for subsequent “Focus” treatment. For this reason, these figures exclude any schools that have been previously identified as Focus or Priority schools. For the cohort of 2013, we see some evidence that Focus schools performed better in math relative to their barely non-Focus counterparts at baseline.<sup>32</sup> For the cohort of 2014, we see little evidence of differences in math achievement, reading achievement, or measures of school composition at the cutoff.

---

<sup>32</sup> In all subsequent outcome models, our controls include baseline achievement in math and reading. Nevertheless, we also caution readers in the standalone interpretation of this one cohort.

### *B. First-Stage Impact on Treatment*

Figure 8 illustrates the extent of dynamic treatment crossover for the 2012 Focus treatment cohort.<sup>33</sup> Figures 8A and 8B show the probability that a school was identified for Focus intervention by 2013 and 2015, respectively. In Figure 8A, we see how the movement of some control schools in 2012 to the treatment side of the cutoff in 2013 mutes the relationship between a school's initial (2012) performance gap index value and the cumulative likelihood it is ever flagged for Focus intervention. Figure 8B illustrates that about 45 percent of schools that fell just to the control side of the initial 2012 cutoff had been identified as Focus schools by 2015.

### *C. Impacts on Academic Achievement*

Table 5 presents estimates of the effects of Focus designation on measures of academic achievement, with a special focus on within-school math and reading gaps.<sup>34</sup> Since the principal aim of the Focus label was to bring attention to large within-school gaps in achievement, these outcomes are of prime importance in assessing the efficacy of any attendant interventions implemented by Focus schools. We find little to no effect of Focus designation on reading performance – at the mean and at various points along the within-school achievement distribution. In Appendix Table A5, we test the sensitivity of our estimates of the effect of Focus status on math and reading gap scores to additional RD specifications that vary the data window and the inclusion of controls.

---

<sup>33</sup> We include all schools in the initial cohort of 2012 and thus a few to the left of the Focus cutoff met the Priority-school criteria and were designated as such.

<sup>34</sup> These outcome measures are slightly different than the variable that functions as the running variable for Focus assignment. That running variable is calculated over two years of prior test scores (see Appendix B for details). For each subject area, we generate our outcomes based on data from each single year and calculate the gap between the top 30 percent of students and bottom 30 percent of students, following the general steps of the state's calculation rules outlined in Appendix B.

In Figure 9, we explore the effects of Focus designation on the math gap score for the cohorts of 2012 and 2014.<sup>35</sup> The linear parametric RD estimates below each graph come from specifications without control variables. Estimates from our preferred specification that includes baseline controls appear in Table 5 and tell the same story. For these two cohorts, we find evidence that Focus schools shrunk within-school gaps in math performance more than their barely non-Focus counterparts. Unfortunately, these reductions were small (i.e., about 4 percent of the control mean) and short-lived. In both cohorts, the gap reductions disappeared by the second or third year after initial identification. Further, estimates across the achievement distribution suggest that decreases in the math gap were driven by declines in performance among students in the top of the distribution while the performance of students in the lower parts of the distribution remained steady.<sup>36</sup>

One might be worried that one explanation for these findings is that schools that fell close to the Focus threshold but remained untreated also responded and worked to improve their gap scores. If such threat effects exist, our comparison between treatment and control schools within the RD setup might understate the true causal impact of the reform. We explore this concern in Appendix Figure A6, wherein we examine trends in the math gap score for Focus schools in the cohort of 2012 alongside a comparison group of schools that should be less sensitive to any threat effects but still function as a reasonable approximation of what the treatment group would have experienced absent treatment.<sup>37</sup> The decline in the math gap score from 2012 to 2013 for

---

<sup>35</sup> We focus on the cohorts of 2012 and 2014 since these are the two cohorts for which baseline tests of the RD assumptions that undergird our analyses of the causal effects of Focus status are most compellingly met.

<sup>36</sup> In results not reported, we find no effect of Focus designation on the distribution of school mean achievement. That is, we see no evidence that Focus status benefited some subset of very low-performing schools more than others.

<sup>37</sup> We define this comparison group as schools that have running variable values between 0.25 and 0.50, which correspond roughly to the 24<sup>th</sup> and 36<sup>th</sup> percentiles of the gap-score rating distribution respectively. We focus on the 2012 cohort because the RD assumptions are compellingly met and we can observe the longest trajectory of outcomes for this cohort (i.e., up to four years after initial identification as a Focus school).

the treatment group is larger than that of the comparison group, but the net difference in those changes is comparable to our RD estimate in magnitude. Thus, our RD findings are unlikely to be driven by comparable improvements in this measure due to threat effects among the non-treated schools. The graph in Panel A suggests that the treatment group experienced continued declines in the math gap measure that were a bit larger than the comparison schools for another year or so. Panels B and C of Appendix Figure A6 illustrate that even if the modest reduction in within-school math performance gaps persisted a bit long, the character of the gap closure did not change. We see that any small reductions in within-school math achievement gaps were driven by little to no improvement of those in the bottom of the distribution (Panel B) alongside declines in the performance of those at the top (Panel C).

*E. Impacts on School Staffing, Enrollment, and Composition*

In Table 6, we turn to effects of Focus status on measures of school staffing and enrollment. We do not find an effect of Focus designation on total school enrollment in any of the three cohorts. In terms of staffing, we find some suggestive evidence from the cohorts of 2013 and 2014 that schools under Focus status experienced higher rates of teacher departure than their barley non-Focus counterparts in the year immediately following identification. Figures 10B and 10C show the corresponding graphs for the 2013 and 2014 cohorts, respectively. Yet, we see no such effect on teacher mobility for the cohort of 2012 (see Figure 10A). Of any cohort, the evidence of an effect of Focus status on teacher mobility appears most suggestive for the cohort of 2014, where we observe a concomitant decrease in the teacher-student ratio alongside the increase in the share of teachers leaving the school.

In Table 7 we explore the effects of Focus designation on the composition of students in schools. Across all three cohorts, we find no evidence that Focus status influenced the share of

students who are black or Hispanic, or the share of economically disadvantaged students in a school, or the share of students receiving special education services. The vast majority of results are close to zero in magnitude, fairly precise, and statistically insignificant. In results not reported but available upon request, we similarly find no evidence of an effect of Focus status on a range of additional measures of school composition and student mobility including the share of students new to the school who were not grade limited in their prior school and the share of students with limited English proficiency.

## **VIII. Putting it All Together: Findings from a Dynamic Regression-Discontinuity**

### **Approach**

Given the dynamic nature of the school reforms we study, our analysis can be recast in a frame similar to the study of school bond elections by Cellini, Ferreira, and Rothstein (2010). Our context differs from the school bond case in at least two ways: (i) schools cannot choose to be considered for Priority or Focus status in the way that a district can choose to hold a bond election in a given year; and (ii) a school cannot be labeled Priority or Focus more than once – that is, falling below the cutoff two years in a row does not constitute *more* intervention in the way that passing two bonds constitutes more funding. This also means that once a school is labeled a Priority or Focus school in one year, it is essentially out of the “risk set” for consideration in future years (although these schools remain in the pool of schools that are ranked, and thus influence the cutoff score for future cohorts). In this section, we pool our three cohorts of schools and adapt the approach developed by Cellini et al. (2010) to estimate treatment-on-the-treated effects in a way that accommodates the dynamic nature of the intervention.



Following Cellini et al. (2010), we create a panel in which the school by ranking year is considered the fundamental unit of analysis, with time relative to the ranking year denoted by  $\tau$  (as distinct from academic year which is denoted  $t$ ). We include data from academic year 2009-2010 (2010) through 2015-2016 (2016). For the 2012 ranking,  $\tau$  runs from -3 through +3. For the 2013 and 2014 rankings, respectively,  $\tau$  runs from -3 to +2 and -3 to 1. Hence, a school that existed over the entire six-year period would be included  $7 + 6 + 5 = 18$  times in the panel. However, once a school is designated as Focus/Priority, we exclude it from future risk sets. For example, schools that are designated Priority in 2012 will only appear in the data for this first ranking cohort, which will correspond to 7 annual observations. Likewise, a school that is designated Priority in 2013 will appear in two ranking cohorts, for a total of 13 observations.<sup>38</sup>

We estimate the following model to estimate the intent-to-treat (ITT) effects of Focus/Priority designation:

$$(5) \quad y_{jt\tau} = \sum_{h=0}^2 b_{jt\tau=h} \theta_{\tau=h}^{ITT} + P_g(r_{jt}, \gamma_{\tau}) + \alpha_{\tau} + \kappa_t + \lambda_{jt} + e_{jt\tau}$$

in which  $y_{jt\tau}$  represents the outcome for school  $j$  in the  $t$  ranking cohort,  $\tau$  years after (before) the ranking year. The  $b$  terms are indicators for whether school  $j$ 's ranking in cohort  $t$  was below the cutoff for the first time and, if so, whether the year in the model is  $\tau$  years after the ranking. The estimates of  $\theta_{\tau}^{ITT}$  capture the effect of missing the cutoff on school outcomes  $\tau$  years later. The equation includes fixed effects for academic year ( $\kappa$ ) and year relative to ranking ( $\alpha$ ). Fixed effects for school-by-cohort ( $\lambda$ ) absorb across-school variation.  $P_g(r_{jt}, \gamma_{\tau})$  is a polynomial in the ranking variable for cohort (i.e., ranking year)  $t$ . Both the  $\gamma_{\tau}$  and  $\theta_{\tau}^{ITT}$  are allowed to vary freely

---

<sup>38</sup> Because the Priority designation is the more severe sanction, schools that are designated Priority are no longer eligible to be labeled Focus. Hence, in conducting the analysis of the Focus intervention, we exclude schools that have previously been labeled Priority or Focus.

with  $\tau$  for  $\tau \geq 0$ , but are constrained to zero otherwise. The value of the school performance polynomial,  $r_j$ , is set to zero in years prior to the introduction of the policy (i.e., academic years 2010-2012). Standard errors are clustered by school.

We include all schools that scored within -1 and 1.5 of the cutoff in the 2012 ranking in our sample. This captures approximately 95 percent of schools that that were ever subject to any intervention. In theory, we could include all schools. However, we have less confidence in the ability of even a high-order polynomial to account for unobserved school characteristics if we get too far from the cutoff. To estimate the impact of actually experiencing the intervention, we estimate dynamic treatment-on-the-treated (TOT) parameters using the recursive method outlined in Cellini et al. (2010).

One final issue involves the inclusion of covariates. In a standard cross-section RD model, one can include pre-treatment controls in order to increase the precision of the estimates, or create balance that one believes is necessary for the RD assumptions to hold. In the dynamic context described above, the inclusion of school fixed effects should go a long way toward these goals. However, if one is willing to assume that the intervention did not impact certain outcomes (e.g., the demographic composition of students in the school), then one can further include such time-varying school characteristics in the models. We do not include such time-varying school controls in our dynamic RD model since the interventions of interest could have plausibly affected the composition and staffing of treated schools.

Tables 8 and 9 present the TOT results for measures of school staffing and composition (Table 8) and achievement (Table 9). These estimates should be interpreted as the effect of an additional year under Priority or Focus status on an outcome. Overall, the patterns of coefficients largely mirror the conclusions we drew from our cohort-by-cohort analyses. In Table 8, we find

little effect of Priority intervention on teacher or principal mobility. We see a boost in school enrollment after four years under Priority intervention. Again, this is likely a product of the closure of geographically clustered Priority schools over time. We see small to no effects of Focus treatment on measures of staffing and the demographic composition of schools. In Table 9, we find no evidence that Priority or Focus reforms improved average student achievement in math or reading.

## **IX. Conclusions**

In this paper, we study the effects of waiver-based Priority and Focus reforms on a range of educational outcomes. To date, there has been little research on the causal effects of these new reforms. Our analysis focuses on the state of Michigan, a state with one of the clearest procedures for identifying Priority and Focus schools. We exploit sharp, discontinuous assignment rules used to distinguish Priority and Focus schools in conjunction with rich administrative data on students and schools to study these effects.

In terms of the prescriptive reforms that targeted the lowest-achieving schools, we find little to no effect of Priority status and associated reforms on measures of school staffing, composition, or academic achievement. Based on a review of federal and state reports, as well as interviews with state officials, we suspect that these null results were – at least in part – due to weak implementation at the school and district levels overlaid by poor monitoring and limited capacity on the part of state officials. Further, our own descriptive analysis of teacher movement in and out of Priority schools that chose different intervention models is consistent with such notions of implementation infidelity.

Our analysis of Focus schools yields similar findings. Although we find some evidence that Focus designation leads to short-run reductions in the within-school gap in math

performance, these declines are small (i.e., a reduction of between 2 and 4 percent of the control group mean gap score per year of Focus status) and appear to be driven by stagnant performance of students in the lower part of the within-school achievement distribution in tandem with performance declines by students in the upper quantiles of the distribution. We find no effects of Focus status on measures of school composition or reading performance. As with the Priority reforms, some evidence suggests that weak implementation contributed to the null results. Michigan's most recent federal monitoring report cites the state's inability to adequately identify and oversee implementation of specific interventions fielded in Focus schools (U.S. Department of Education, 2014b, p. 5).

This picture is largely consistent with emerging evidence from other states on Focus-school reforms (Dee & Dizon-Ross, 2017). A recent report from the U.S. Government Accountability Office found that many states struggled with multiple implementation challenges related to waiver-based reforms, including incomplete accountability systems and difficulty monitoring districts and schools (GAO, 2016). The one exception is Kentucky, where Focus designation led to improvements in math and reading achievement for targeted students (Bonilla & Dee, 2017). The authors cite strong state oversight and high-fidelity implementation that included comprehensive school planning and professional development as likely mediators of these positive effects.

Like the NCLB waivers, the Every Student Succeeds Act (ESSA) continues to require states to identify their lowest-performing schools (i.e., the bottom 5 percent) and to intervene in those schools; but, states will have much more flexibility in determining the attributes of the interventions they field in those schools (Burnette II, 2016; National Council of State Legislatures, 2016). However, the new accountability provisions in the ESSA tied to a state's

lowest-performing schools differ in at least two additional ways from the waiver era reforms. First, states must include at least one non-academic factor (such as student engagement or school safety) along with academic indicators in their school rating systems. Second, districts with low-performing schools will receive extra funds from the state-level pot of Title I dollars that are specifically allocated to support the turnaround of chronically struggling schools (Burnette II, 2016). The degree to which these changes will translate into meaningful action on the part of states, districts, and schools that leads to improved student outcomes is unclear. The first round of EESA plans suggests that states have closely hewn to the accountability structures they set up under their NCLB waivers (Klein & Ujifusa, 2017). Further research will be needed to determine if the EESA reforms are able to improve student performance.

## REFERENCES

- 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.
- Bonilla, S., & Dee, T. S. (2017). The Effects of School Reform Under NCLB Waivers: Evidence from Focus Schools in Kentucky. NBER Working Paper.
- 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, 125–230.
- Burnette II, D. (2016, January 5). States, districts to call shots on turnarounds under ESSA. *Education Week*, 35(15), 18.
- Calonico, S., Cattaneo, M. D., & Titiunik, R. (2015). Optimal data-driven regression discontinuity plots. *Journal of the American Statistical Association*, 110(512), 1753-1769.
- Calonico, S., Cattaneo, M. D., & Titiunik, R. (2014a). Robust data-drive inference in the regression-discontinuity design. *The Stata Journal*, 14(4), 909-946.
- Calonico, S., Cattaneo, M. D., & Titiunik, R. (2014b). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, 82, 2295-2326.
- Cellini, S., Ferreira, F., and Rothstein, J. (2010). The value of school facilities: Evidence from a dynamic regression discontinuity design. *Quarterly Journal of Economics*, 125(1), 215-261.
- Center on Education Policy. (2013). States' perspectives on waivers: Relief from NCLB, concern about long-term solutions. George Washington University, Retrieved from [http://www.cep-dc.org/cfcontent\\_file.cfm?Attachment=McMurrerYoshioka%5FReport%5FStatesPerspectivesonWaivers%5F030413%2Epdf](http://www.cep-dc.org/cfcontent_file.cfm?Attachment=McMurrerYoshioka%5FReport%5FStatesPerspectivesonWaivers%5F030413%2Epdf)
- de la Torre, M., Allensworth, E., Jagesic, S., Sebastian, J., Salmonowicz, M., Meyers, C., & Gerdeman, R.D. (2013). Turning around low-performing schools in Chicago. Research Report. Chicago, IL: The University of Chicago Consortium on Chicago School Research.
- Dee, T. S., & Dizon-Ross, E. (2017). School Performance, Accountability and Waiver Reforms: Evidence from Louisiana. NBER Working Paper.
- Dee, T. S. (2012). School turnarounds: Evidence from the 2009 Stimulus. NBER Working Paper Series, No. 17990. Retrieved from <http://www.nber.org/papers/w17990>

- Dee, T. S., & Jacob, B. (2011). The impact of No Child Left Behind on Student Achievement. *Journal of Policy Analysis and Management*, 30(3), 418-446.
- Deming, D. J., Cohodes, S., Jennings, J., & Jencks, C. (2016). School accountability, postsecondary attainment and earnings. *Review of Economics and Statistics*, 98(5), 848-862.
- Desimone, L. (2002). How can comprehensive school reform models be successfully implemented? *Review of Educational Research*, 72(3), 433-479.
- Doll, J. (2017, January 10, 27). Personal communication.
- Dougherty, S.M., & Weiner, J. (2016). "The Rhode to Turnaround? The Impact of Waivers to No Child Left Behind School Performance." Working Paper, University of Connecticut.
- Dragoset, L., Thomas, J., Herrmann, M., Deke, J., James-Burdumy, S., Graczewski, C., Boyle, A., Upton, R., Tanenbaum, C., & Giffin, J. (2017). *School Improvement Grants: Implementation and Effectiveness* (NCEE 2017-4013). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.
- Every Student Succeeds Act (ESSA; 2015). Public Law No. 114-95. Retrieved from <https://www.congress.gov/bill/114th-congress/senate-bill/1177/text>
- Fausset, R., & Blinder, A. (2014, April 14). Atlanta school workers sentenced in test score cheating case, *New York Times*, p. A1.
- Figlio, D. N., & Ladd, H. F. (2015). School accountability and student achievement. In H. F. Ladd & M. E. Goertz (Eds.), *Handbook of Research in Education Finance and Policy* (pp. 194-210).
- Gill, B., Zimmer, R. Christman, J.B., & Blanc, S. (2007). State Takeover, Restructuring, Private Management, and Student Achievement in Philadelphia Santa Monica, CA: RAND Corporation and Research for Action.
- Gross, B., Booker, K. T., & Goldhaber, D. (2009). Boosting student achievement: the effect of comprehensive school reform on student achievement. *Educational Evaluation and Policy Analysis*, 31(2), 111-126.
- Hahn, J., Todd, P., & Van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, 69(1), 201-209.
- Heissel, J. A., & Ladd, H. F. (2016). School turnaround in North Carolina: A regression discontinuity analysis. CALDER Discussion Paper No. 156.

- Hemelt, S. W. (2011). Performance effects of failure to make Adequate Yearly Progress (AYP): Evidence from a regression discontinuity framework. *Economics of Education Review*, 30(4), 702-723.
- Jacob, B. A. (2005). Accountability, incentives and behavior: the impact of high-stakes testing in the Chicago Public Schools. *Journal of Public Economics*, 89(5-6), 761-796.
- Jacob, B. A. (2017). The changing federal role in school accountability. *Journal of Policy Analysis and Management*, 36(2), 469-477.
- Jacob, B. A., & Levitt, S. D. (2003). Rotten apples: An investigation of the prevalence and predictors of teaching cheating. *Quarterly Journal of Economics*, 118(3), 843-877.
- Klein, A. & Ujifusa, A. (2017, April 19). First wave of ESSA plans gives early look at state priorities. *Education Week*, 36(28), 16-19.
- Keesler, V. (2016, March 18). Personal communication.
- Koretz, D. (2008). *Measuring up: What educational testing really tells us*. Cambridge, MA: Harvard University Press.
- Ladd, H. F. (2017). No Child Left Behind: A deeply flawed federal policy. *Journal of Policy Analysis and Management*, 36(2), 461-469.
- McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, 142(2), 698-714.
- Michigan Department of Education (MDE, 2013). Office of Improvement and Innovation, Focus School Technical Assistance.
- Mills, J. I. (2008). A legislative overview of No Child Left Behind. *New Directions for Evaluation*, 117, 9-20.
- National Council of State Legislatures. (2016). "Summary of the Every Student Succeeds Act, Legislation Reauthorizing the Elementary and Secondary Education Act." Retrieved from [http://www.ncsl.org/documents/educ/ESSA\\_summary\\_NCSL.pdf](http://www.ncsl.org/documents/educ/ESSA_summary_NCSL.pdf)
- Player, D., & Katz, V. (2013). School improvement in Ohio: An evaluation of the School Turnaround Specialist Program. CEPWC Working Paper Series No. 10.
- Papay, J. P., & Hannon, M. (2016). "The Effect of School Turnaround Strategies in Massachusetts." Working paper, Brown University.
- Rothstein, R., Jacobsen, R., & Wilder, T. (2008). *Grading education: Getting accountability right*. New York: Teachers College Press.



- Ruble, W. (2015). The Effect of Contracting Out Low Performing Schools on Student Performance (Working Paper No. 1521). Tulane University, Department of Economics. Retrieved from <https://ideas.repec.org/p/tul/wpaper/1521.html>
- Ruple, K. (2016, June 10). Personal communication.
- Saw, G. K., Schneider, B., Frank, K., I-Chien, C., Keesler, V., & Martineau, J. (2016). The impact of being labeled as a persistently lowest achieving schools: Regression discontinuity evidence on school labeling. Working paper, Michigan State University.
- Schueler, B. E., Goodman, J., & Deming, D. J. (2017). Can states take over and turn around school districts? Evidence from Lawrence, Massachusetts. *Educational Evaluation and Policy Analysis*, forthcoming.
- U.S. Department of Education. (2014a, July 3) press release, "U.S. Department of Education Approves Extensions for States Granted Flexibility from No Child Left Behind;" Retrieved from <http://www.ed.gov/news/press-releases/us-department-education-approves-extensions-states-granted-flexibility-no-child-left-behind>.
- U.S. Department of Education (2014b). *ESEA Flexibility Part B Monitoring Report: Michigan Department of Education*. Retrieved from <https://www2.ed.gov/admins/lead/account/monitoring/reports13/mipartbrpt2014.pdf>
- U.S. Department of Education. (2012, June 7). ESEA Flexibility. Retrieved from <http://www2.ed.gov/policy/eseaflex/approved-requests/flexrequest.doc>
- U.S. Department of Education (2010). *Evaluation of the comprehensive school reform program implementation and outcomes: fifth-year report*. Washington DC: Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service.
- U.S. Government Accountability Office (2016). *K-12 Education: Education's experiences with flexibility waivers could inform efforts to assist states with new requirements*. Retrieved from <http://www.gao.gov/assets/680/678258.pdf>
- Zimmer, R., Henry, G. T., & Kho, A. (2016). The Role of Governance and Management in School Turnaround Policies: The Case of Tennessee's Achievement School District and iZones. Presented at the Association of Education Finance and Policy, Denver, Colorado: [https://aefpweb.org/sites/default/files/webform/41/ASD\\_iZone%20Impact%20Paper%20conference%20online%20version.pdf](https://aefpweb.org/sites/default/files/webform/41/ASD_iZone%20Impact%20Paper%20conference%20online%20version.pdf)

**Table 1. Descriptive Statistics: Cohort of 2012**

Variable	All K-8 Schools	Priority K-8 Schools	Focus K-8 Schools
	Mean (1)	Mean (2)	Mean (3)
Priority	0.030	1.000	--
Focus	0.144	--	1.000
Average Math Score (std)	0.031 (0.441)	-0.712 (0.319)	0.381 (0.452)
Average Reading Score (std)	0.030 (0.356)	-0.690 (0.299)	0.232 (0.307)
Priority Running Variable	1.520 (0.840)	-0.330 (0.334)	1.706 (0.819)
Focus Running Variable	0.641 (0.707)	1.311 (1.029)	-0.489 (0.461)
<i><u>Student Characteristics</u></i>			
Share black	0.153	0.775	0.130
Share white	0.730	0.143	0.712
Share Hispanic	0.070	0.067	0.055
Share Asian	0.033	0.012	0.090
Share economically disadvantaged	0.521	0.906	0.388
Share LEP	0.051	0.077	0.066
Share special education	0.149	0.168	0.138
<i><u>Staff Characteristics</u></i>			
Teacher-student ratio	0.081 (0.018)	0.076 (0.014)	0.082 (0.016)
Aide-student ratio	0.023 (0.015)	0.021 (0.011)	0.022 (0.013)
Share of teachers in first year at school	0.118	0.230	0.121
Average teacher experience (years)	14.7 (2.6)	17.3 (3.0)	14.6 (2.4)
<i><u>School Characteristics</u></i>			
Total enrollment	463 (191)	498 (220)	511 (203)
Elementary	0.684	0.661	0.638
Middle	0.245	0.054	0.309
Magnet	0.137	0.107	0.109
Urban	0.150	0.786	0.174
Suburban	0.455	0.196	0.570
Rural or Town	0.394	0.018	0.257
SIG Cohort III or IV	0.009	0.125	0.000
N(schools)	1841	56	265

Notes: Standard deviations of select continuous variables appear in parentheses below the means. Sample is limited to K-8, non-special-education, non-charter schools open as of the fall of 2012 with total enrollment in the baseline year of at least 50 students. Schools identified as persistently low achieving (PLA) prior to 2012 and schools that received SIG grants during the first two competitions (prior to 2012) are also excluded from the sample. The running variable for Priority schools is an index of prior level achievement, growth in achievement, and performance gaps; the running variable for Focus schools is a measure of the gap in performance between the top 30 percent and bottom 30 percent of students within a school. See text for additional details about running variables. LEP = limited English proficient; SIG = school improvement grant. Cohorts I and II of the SIG competition were identified before the first cohort of Priority and Focus schools. Priority and Focus schools were eligible to compete for SIG funding in Cohorts III and IV.

**Table 2. ITT Effects of Priority Designation on Achievement**

Cohort Outcome	1 year after		2 years after		3 years after		4 years after
	2012 (1)	2014 (2)	2013 (3)	2014 (4)	2012 (5)	2013 (6)	2012 (7)
<i>A. Mathematics</i>							
Average math score (std)	0.003 (0.101) [-0.566]	-0.122 (0.075) [-0.536]	-0.109 (0.123) [-0.656]	-0.079 (0.085) [-0.541]	0.013 (0.117) [-0.642]	-0.160 (0.129) [-0.680]	-0.134 (0.172) [-0.643]
Math gap score	0.040 (0.121) [1.75]	-0.038 (0.080) [1.89]	-0.037 (0.082) [1.89]	-0.086 (0.083) [1.89]	0.139 (0.089) [1.89]	0.010 (0.087) [1.89]	0.249** (0.105) [1.93]
Math score, 10th percentile	-0.060 (0.075)	-0.082 (0.081)	-0.105 (0.150)	-0.047 (0.084)	-0.143 (0.132)	-0.178 (0.132)	-0.330* (0.179)
Math score, 25th percentile	0.037 (0.083)	-0.057 (0.078)	-0.124 (0.116)	-0.049 (0.075)	-0.001 (0.119)	-0.125 (0.120)	-0.243 (0.159)
Math score, 50th percentile	-0.007 (0.107)	-0.111 (0.069)	-0.166 (0.120)	-0.083 (0.088)	0.003 (0.110)	-0.161 (0.125)	-0.088 (0.172)
Math score, 75th percentile	-0.034 (0.130)	-0.127 (0.093)	-0.111 (0.138)	-0.147 (0.112)	0.053 (0.144)	-0.170 (0.139)	-0.003 (0.192)
Math score, 90th percentile	0.103 (0.176)	-0.219** (0.107)	-0.047 (0.143)	-0.082 (0.122)	0.129 (0.166)	-0.166 (0.167)	-0.020 (0.204)
<i>B. Reading</i>							
Average reading score (std)	-0.086 (0.100) [-0.513]	-0.173** (0.077) [-0.523]	0.160 (0.109) [-0.603]	-0.151** (0.073) [-0.526]	0.232 (0.145) [-0.588]	-0.035 (0.135) [-0.612]	0.066 (0.147) [-0.569]
Reading gap score	0.027 (0.074) [2.00]	-0.046 (0.101) [2.04]	0.016 (0.089) [1.99]	-0.169** (0.077) [2.01]	0.038 (0.149) [1.98]	-0.077 (0.098) [1.97]	0.114 (0.115) [1.94]
Reading score, 10th percentile	-0.127 (0.085)	-0.091 (0.080)	0.214* (0.109)	-0.049 (0.064)	0.212 (0.150)	-0.032 (0.116)	-0.019 (0.096)
Reading score, 25th percentile	-0.114 (0.108)	-0.110 (0.082)	0.144 (0.111)	-0.080 (0.072)	0.208 (0.158)	-0.001 (0.123)	0.032 (0.130)
Reading score, 50th percentile	-0.109 (0.115)	-0.207** (0.090)	0.142 (0.131)	-0.160** (0.076)	0.280* (0.154)	-0.041 (0.145)	0.051 (0.172)
Reading score, 75th percentile	-0.037 (0.110)	-0.190* (0.098)	0.141 (0.108)	-0.227** (0.099)	0.221 (0.153)	-0.024 (0.169)	0.096 (0.191)
Reading score, 90th percentile	-0.079 (0.135)	-0.152 (0.103)	0.161 (0.110)	-0.197* (0.114)	0.227 (0.179)	-0.111 (0.167)	0.097 (0.189)
N(schools)	79	75	72	70	76	70	74

Notes: The analytic sample for the 2012 cohort includes all students enrolled in K-8, non-special-education, non-charter schools open as of the fall of 2012 with total enrollment in the baseline year of least 50 students. For subsequent cohorts, we exclude schools identified as Priority in prior years since they leave the risk set for Priority treatment after initial designation for intervention.

Coefficients presented in the table come from our preferred parametric specification, which is estimated on a data window of +/- 0.25 from the cutoff, includes covariates, and is weighted by total enrollment. For a full list of covariates, please consult the text. Robust standard errors appear in parentheses. Control group means appear in brackets. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1. ITT = intent to treat

**Table 3. ITT Effects of Priority Designation on School Staffing and Enrollment**

Outcomes	Outcome Year			
	2013 (1)	2014 (2)	2015 (3)	2016 (4)
<i>A. 2012 Cohort</i>				
Log(enrollment), total	0.225*** (0.077) [5.97]	0.268** (0.132) [5.97]	0.323*** (0.122) [5.95]	0.301** (0.136) [5.94]
Log(enrollment), entry grades	0.106 (0.160) [4.38]	0.202 (0.189) [4.36]	0.287 (0.212) [4.31]	0.375* (0.216) [4.28]
Log(enrollment), new in non-entry grades	0.672*** (0.236) [4.30]	0.460 (0.348) [4.36]	0.230 (0.231) [4.21]	0.347 (0.228) [4.23]
Share of teachers new to school	0.197* (0.120) [0.215]	0.037 (0.055) [0.228]	0.005 (0.061) [0.204]	-0.009 (0.059) [0.172]
Share of teachers who left school within past year	0.109 (0.108) [0.227]	-0.034 (0.067) [0.219]	0.071 (0.054) [0.186]	0.013 (0.055) [0.192]
New principal as of outcome year	-0.285 (0.315) [0.293]	-0.274 (0.312) [0.288]	-0.153 (0.334) [0.275]	-0.366 (0.389) [0.275]
Teacher-student ratio	-0.004 (0.006) [0.082]	0.003 (0.010) [0.085]	-0.012 (0.010) [0.086]	-0.011 (0.009) [0.076]
Aide-student ratio	-0.002 (0.004) [0.024]	0.005 (0.005) [0.024]	-0.002 (0.006) [0.023]	-0.000 (0.006) [0.021]
N(schools)	88	79	76	74
<i>B. 2013 Cohort</i>				
Log(enrollment), total		-0.007 (0.081) [5.99]	-0.037 (0.076) [5.96]	-0.080 (0.100) [5.97]
Log(enrollment), entry grades		0.108 (0.188) [4.32]	0.120 (0.193) [4.26]	0.216 (0.184) [4.27]
Log(enrollment), new in non-entry grades		-0.229 (0.297) [4.42]	-0.048 (0.237) [4.17]	0.081 (0.187) [4.25]
Share of teachers new to school		0.059 (0.080) [0.209]	0.070 (0.076) [0.183]	0.079 (0.048) [0.174]
Share of teachers who left school within past year		0.029 (0.071) [0.192]	0.077 (0.051) [0.180]	0.107 (0.067) [0.174]
New principal as of outcome year		0.102 (0.208) [0.245]	0.111 (0.198) [0.260]	0.152 (0.195) [0.229]
Teacher-student ratio		0.005 (0.006) [0.088]	0.010 (0.009) [0.089]	0.008 (0.011) [0.079]

Aide-student ratio	0.001 (0.004) [0.021]	0.004 (0.008) [0.021]	0.007 (0.008) [0.019]
N(schools)	74	72	70
<i>C. 2014 Cohort</i>			
Log(enrollment), total		0.048 (0.041) [5.98]	-0.031 (0.081) [5.97]
Log(enrollment), entry grades		-0.193 (0.191) [4.28]	-0.173 (0.229) [4.29]
Log(enrollment), new in non-entry grades		0.314* (0.180) [4.11]	0.047 (0.163) [4.10]
Share of teachers new to school		-0.043 (0.039) [0.194]	-0.049 (0.050) [0.173]
Share of teachers who left school within past year		-0.095*** (0.032) [0.184]	-0.015 (0.052) [0.181]
New principal as of outcome year		0.222 (0.230) [0.196]	0.473* (0.241) [0.143]
Teacher-student ratio		0.003 (0.006) [0.089]	0.000 (0.008) [0.083]
Aide-student ratio		-0.004 (0.003) [0.025]	-0.004 (0.005) [0.022]
N(schools)		75	70

Notes: The analytic sample for the 2012 cohort includes all students enrolled in K-8, non-special-education, non-charter schools open as of the fall of 2012 with total enrollment in the baseline year of at least 50 students. For subsequent cohorts, we exclude schools identified as Priority in prior years since they leave the risk set for Priority treatment after initial designation for intervention. Coefficients presented in the table come from our preferred parametric specification, which is estimated on a data window of +/- 0.25 from the cutoff, includes covariates, and is weighted by total enrollment. For a full list of covariates, please consult the text. Robust standard errors appear in parentheses. Control group means appear in brackets. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1. ITT = intent to treat.

**Table 4. ITT Effects of Priority Designation on School Composition**

Outcomes	Outcome Year			
	2013 (1)	2014 (2)	2015 (3)	2016 (4)
<i>A. 2012 Cohort</i>				
Share black	0.006 (0.013) [0.505]	-0.004 (0.019) [0.513]	-0.017 (0.030) [0.510]	-0.018 (0.033) [0.517]
Share Hispanic	0.000 (0.007) [0.113]	-0.008 (0.011) [0.106]	-0.013 (0.016) [0.111]	-0.001 (0.016) [0.111]
Share economically disadvantaged	-0.023 (0.021) [0.837]	-0.010 (0.027) [0.837]	-0.038 (0.029) [0.842]	-0.026 (0.043) [0.800]
Share special education	-0.018 (0.012) [0.174]	-0.004 (0.017) [0.178]	-0.019 (0.022) [0.174]	-0.026 (0.025) [0.178]
N(schools)	88	79	76	74
<i>B. 2013 Cohort</i>				
Share black		-0.029 (0.021) [0.558]	-0.048 (0.033) [0.557]	-0.038 (0.031) [0.567]
Share Hispanic		0.020 (0.012) [0.089]	0.010 (0.016) [0.092]	0.016 (0.016) [0.090]
Share economically disadvantaged		-0.069*** (0.025) [0.863]	-0.002 (0.024) [0.860]	-0.009 (0.035) [0.823]
Share special education		0.008 (0.013) [0.185]	-0.001 (0.022) [0.185]	0.021 (0.028) [0.182]
N(schools)		74	72	70
<i>C. 2014 Cohort</i>				
Share black			0.006 (0.010) [0.472]	0.011 (0.021) [0.472]
Share Hispanic			-0.015** (0.007) [0.115]	-0.018* (0.011) [0.121]
Share economically disadvantaged			-0.039 (0.026) [0.823]	-0.033 (0.033) [0.794]
Share special education			0.004 (0.014) [0.177]	-0.005 (0.018) [0.180]
N(schools)			75	70

Notes: The analytic sample for the 2012 cohort includes all students enrolled in K-8, non-special-education, non-charter schools open as of the fall of 2012 with total enrollment in the baseline year of least 50 students. For subsequent cohorts, we exclude schools identified as Priority in prior years since they leave the risk set for Priority treatment after initial designation for intervention. Coefficients presented in the table come from our preferred parametric specification, which is estimated on a data window of +/- 0.25 from the cutoff, includes covariates, and is weighted by total enrollment. For a full list of covariates, please consult the text. Robust standard errors appear in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1. Control group means appear in brackets. ITT = intent to treat.

**Table 5. ITT Effects of Focus Designation on Achievement**

Cohort Outcome	1 year after		2 years after		3 years after		4 years after
	2012 (1)	2014 (2)	2013 (3)	2014 (4)	2012 (5)	2013 (6)	2012 (7)
<i>A. Mathematics</i>							
Average math score (std)	0.001 (0.030) [0.128]	-0.023 (0.039) [0.039]	-0.035 (0.043) [0.001]	0.039 (0.042) [0.038]	-0.055* (0.032) [0.159]	-0.053 (0.040) [0.007]	-0.016 (0.035) [0.152]
Math gap score	-0.083*** (0.029) [2.06]	-0.051 (0.040) [1.99]	0.012 (0.035) [1.99]	0.030 (0.041) [2.00]	-0.031 (0.028) [2.00]	0.019 (0.039) [2..00]	-0.033 (0.028) [2.00]
Math score, 10th percentile	0.026 (0.026)	0.005 (0.050)	-0.044 (0.048)	0.030 (0.046)	-0.049 (0.035)	-0.056 (0.046)	-0.017 (0.039)
Math score, 25th percentile	0.032 (0.029)	-0.008 (0.040)	-0.026 (0.045)	0.020 (0.043)	-0.045 (0.033)	-0.064 (0.041)	-0.002 (0.038)
Math score, 50th percentile	0.018 (0.031)	-0.016 (0.039)	-0.049 (0.047)	0.059 (0.045)	-0.042 (0.033)	-0.067 (0.041)	-0.015 (0.037)
Math score, 75th percentile	-0.014 (0.037)	-0.058 (0.043)	-0.031 (0.047)	0.047 (0.049)	-0.065* (0.036)	-0.047 (0.046)	-0.023 (0.038)
Math score, 90th percentile	-0.051 (0.044)	-0.046 (0.047)	-0.033 (0.047)	0.044 (0.054)	-0.093** (0.040)	-0.035 (0.048)	-0.042 (0.039)
<i>B. Reading</i>							
Average reading score (std)	0.010 (0.021) [0.111]	0.011 (0.042) [0.044]	-0.049 (0.045) [0.017]	0.013 (0.038) [0.036]	-0.027 (0.036) [0.142]	-0.040 (0.037) [0.015]	0.016 (0.030) [0.134]
Reading gap score	-0.022 (0.025) [2.07]	-0.026 (0.036) [2.10]	0.014 (0.034) [2.09]	-0.005 (0.038) [2.11]	-0.010 (0.026) [2.07]	0.033 (0.032) [2.11]	-0.015 (0.025) [2.10]
Reading score, 10th percentile	0.016 (0.027)	0.025 (0.054)	-0.055 (0.052)	0.004 (0.046)	-0.027 (0.042)	-0.037 (0.042)	0.018 (0.034)
Reading score, 25th percentile	0.021 (0.026)	0.027 (0.048)	-0.067 (0.052)	0.017 (0.045)	-0.032 (0.040)	-0.058 (0.041)	0.022 (0.035)
Reading score, 50th percentile	0.013 (0.022)	0.013 (0.045)	-0.034 (0.049)	0.015 (0.044)	-0.016 (0.039)	-0.043 (0.041)	0.034 (0.034)
Reading score, 75th percentile	0.006 (0.025)	0.002 (0.044)	-0.048 (0.046)	0.012 (0.041)	-0.024 (0.038)	-0.027 (0.043)	0.003 (0.032)
Reading score, 90th percentile	-0.011 (0.034)	-0.010 (0.046)	-0.045 (0.048)	-0.030 (0.039)	-0.035 (0.039)	-0.010 (0.042)	0.014 (0.032)
N(schools)	625	366	398	361	616	393	607

Notes: The analytic sample for the 2012 cohort includes all students enrolled in K-8, non-special-education, non-charter schools open as of the fall of 2012 with total enrollment in the baseline year of at least 50 students. For subsequent cohorts, we exclude schools identified as Focus or Priority in prior years since they leave the risk set for Focus treatment after initial designation for intervention. Coefficients presented in the table come from our preferred parametric specification, which is estimated on a data window of +/- 0.50 from the cutoff, includes covariates, and is weighted by total enrollment. For a full list of covariates, please consult the text. Robust standard errors appear in parentheses. Control group means appear in brackets. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1. ITT = intent to treat

**Table 6. ITT Effects of Focus Designation on School Staffing and Enrollment**

Outcomes	Outcome Year			
	2013 (1)	2014 (2)	2015 (3)	2016 (4)
<i>A. 2012 Cohort</i>				
Log(enrollment), total	0.000 (0.020) [6.11]	0.024 (0.025) [6.10]	0.025 (0.028) [6.10]	0.012 (0.030) [6.10]
Share of teachers new to school	0.019 (0.019) [0.130]	0.007 (0.017) [0.109]	-0.035** (0.016) [0.119]	-0.002 (0.012) [0.104]
Share of teachers who left school within past year	0.005 (0.014) [0.112]	0.001 (0.013) [0.101]	-0.018 (0.014) [0.103]	-0.017 (0.012) [0.102]
New principal as of outcome year	0.007 (0.057) [0.148]	0.030 (0.068) [0.173]	0.027 (0.071) [0.182]	0.008 (0.071) [0.179]
Teacher-student ratio	0.002 (0.002) [0.082]	0.003 (0.002) [0.083]	-0.000 (0.002) [0.084]	0.002 (0.002) [0.078]
Aide-student ratio	0.001 (0.001) [0.024]	0.001 (0.002) [0.024]	0.001 (0.002) [0.024]	0.001 (0.002) [0.022]
N(schools)	640	625	616	607
<i>B. 2013 Cohort</i>				
Log(enrollment), total		-0.004 (0.025) [6.11]	-0.025 (0.022) [6.11]	-0.038 (0.027) [6.10]
Share of teachers new to school		0.025 (0.026) [0.118]	0.047*** (0.017) [0.116]	0.004 (0.017) [0.105]
Share of teachers who left school within past year		0.032* (0.020) [0.104]	0.020 (0.015) [0.103]	0.026* (0.015) [0.103]
New principal as of outcome year		-0.056 (0.056) [0.099]	-0.101* (0.056) [0.109]	-0.103* (0.057) [0.107]
Teacher-student ratio		-0.001 (0.002) [0.082]	0.003 (0.003) [0.081]	0.001 (0.003) [0.076]
Aide-student ratio		-0.000 (0.002) [0.024]	0.001 (0.003) [0.024]	0.002 (0.002) [0.022]
N(schools)		405	398	393
<i>C. 2014 Cohort</i>				
Log(enrollment), total			-0.019 (0.017) [6.10]	-0.007 (0.023) [6.09]
Share of teachers new to school			0.009 (0.021) [0.124]	0.009 (0.016) [0.108]
Share of teachers who left school within past year			0.043** (0.018) [0.107]	-0.011 (0.015) [0.108]
New principal as of outcome year			0.073 (0.064) [0.102]	.121 (0.088) [0.116]
Teacher-student ratio			-0.005* (0.003) [0.083]	-0.002 (0.003) [0.077]
Aide-student ratio			-0.003** (0.002) [0.024]	-0.005** (0.002) [0.021]
N(schools)			366	361

Notes: The analytic sample for the 2012 cohort includes all students enrolled in K-8, non-special-education, non-charter schools open as of the fall of 2012 with total enrollment in the baseline year of at least 50 students. For subsequent cohorts, we exclude schools identified as Focus or Priority in prior years since they leave the risk set for Focus treatment after initial designation for intervention. Coefficients presented in the table come from our preferred parametric specification, which is estimated on a data window of +/- 0.50 from the cutoff, includes covariates, and is weighted by total enrollment. For a full list of covariates, please consult the text. Robust standard errors appear in parentheses. Control group means appear in brackets. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1. ITT = intent to treat.



**Table 7. ITT Effects of Focus Designation on School Composition**

Outcomes	Outcome Year			
	2013 (1)	2014 (2)	2015 (3)	2016 (4)
<i>A. 2012 Cohort</i>				
Share black	-0.002 (0.003) [0.101]	0.001 (0.004) [0.099]	0.003 (0.005) [0.102]	0.003 (0.006) [0.107]
Share Hispanic	-0.001 (0.002) [0.056]	-0.001 (0.003) [0.057]	0.001 (0.004) [0.059]	0.000 (0.004) [0.060]
Share economically disadvantaged	0.010 (0.009) [0.468]	0.005 (0.009) [0.467]	0.005 (0.011) [0.454]	-0.005 (0.011) [0.449]
Share special education	0.000 (0.003) [0.145]	0.005 (0.005) [0.141]	0.009 (0.006) [0.142]	-0.001 (0.006) [0.145]
N(schools)	640	625	616	607
<i>B. 2013 Cohort</i>				
Share black		0.001 (0.004) [0.111]	-0.002 (0.006) [0.115]	0.003 (0.007) [0.118]
Share Hispanic		0.000 (0.003) [0.062]	0.005 (0.005) [0.063]	0.005 (0.005) [0.065]
Share economically disadvantaged		0.005 (0.009) [0.539]	0.000 (0.013) [0.532]	0.018 (0.012) [0.525]
Share special education		-0.000 (0.005) [0.150]	0.002 (0.007) [0.149]	0.003 (0.008) [0.153]
N(schools)		405	398	393
<i>C. 2014 Cohort</i>				
Share black			-0.002 (0.004) [0.110]	-0.004 (0.006) [0.117]
Share Hispanic			-0.000 (0.002) [0.068]	0.002 (0.003) [0.071]
Share economically disadvantaged			0.006 (0.010) [0.506]	0.019 (0.013) [0.500]
Share special education			-0.008* (0.004) [0.149]	-0.003 (0.006) [0.152]
N(schools)			366	361

Notes: The analytic sample for the 2012 cohort includes all students enrolled in K-8, non-special-education, non-charter schools open as of the fall of 2012 with total enrollment in the baseline year of at least 50 students. For subsequent cohorts, we exclude schools identified as Focus or Priority in prior years since they leave the risk set for Focus treatment after initial designation for intervention. Coefficients presented in the table come from our preferred parametric specification, which is estimated on a data window of +/- 0.50 from the cutoff, includes covariates, and is weighted by total enrollment. For a full list of covariates, please consult the text. Robust standard errors appear in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1. Control group means appear in brackets. ITT = intent to treat.

**Table 8. Effects of Priority and Focus Status on School Staffing and Composition: Evidence from a Dynamic RD Approach**

Outcome	Priority				Focus			
	1 year after (1)	2 years after (2)	3 years after (3)	4 years after (4)	1 year after (5)	2 years after (6)	3 years after (7)	4 years after (8)
<i>A. Staffing</i>								
Share of teachers new to school	0.025 (0.054)	0.037 (0.053)	0.059 (0.083)	0.045 (0.133)	0.021 (0.022)	0.010 (0.019)	-0.027 (0.024)	-0.016 (0.028)
Share of teachers who left school within past year	-0.015 (0.044)	-0.023 (0.046)	0.023 (0.062)	0.009 (0.098)	0.021 (0.014)	-0.004 (0.016)	-0.008 (0.017)	-0.026 (0.022)
New principal as of outcome year	0.027 (0.169)	0.053 (0.113)	0.116 (0.225)	-0.085 (0.302)	0.020 (0.039)	0.023 (0.050)	0.028 (0.055)	0.054 (0.060)
<i>B. Composition</i>								
Log of total enrollment	0.026 (0.049)	0.052 (0.084)	0.134 (0.117)	0.274** (0.118)	-0.011 (0.020)	-0.001 (0.030)	0.009 (0.038)	0.013 (0.047)
Share black	-0.004 (0.011)	-0.010 (0.021)	-0.003 (0.032)	0.024 (0.036)	0.005** (0.003)	0.008 (0.005)	0.011 (0.007)	0.008 (0.010)
Share economically disadvantaged	0.004 (0.015)	0.019 (0.022)	0.021 (0.031)	0.027 (0.073)	0.008 (0.007)	0.002 (0.010)	0.004 (0.014)	-0.023 (0.018)
N(schools)	917	917	917	917	1,654	1,654	1,654	1,654

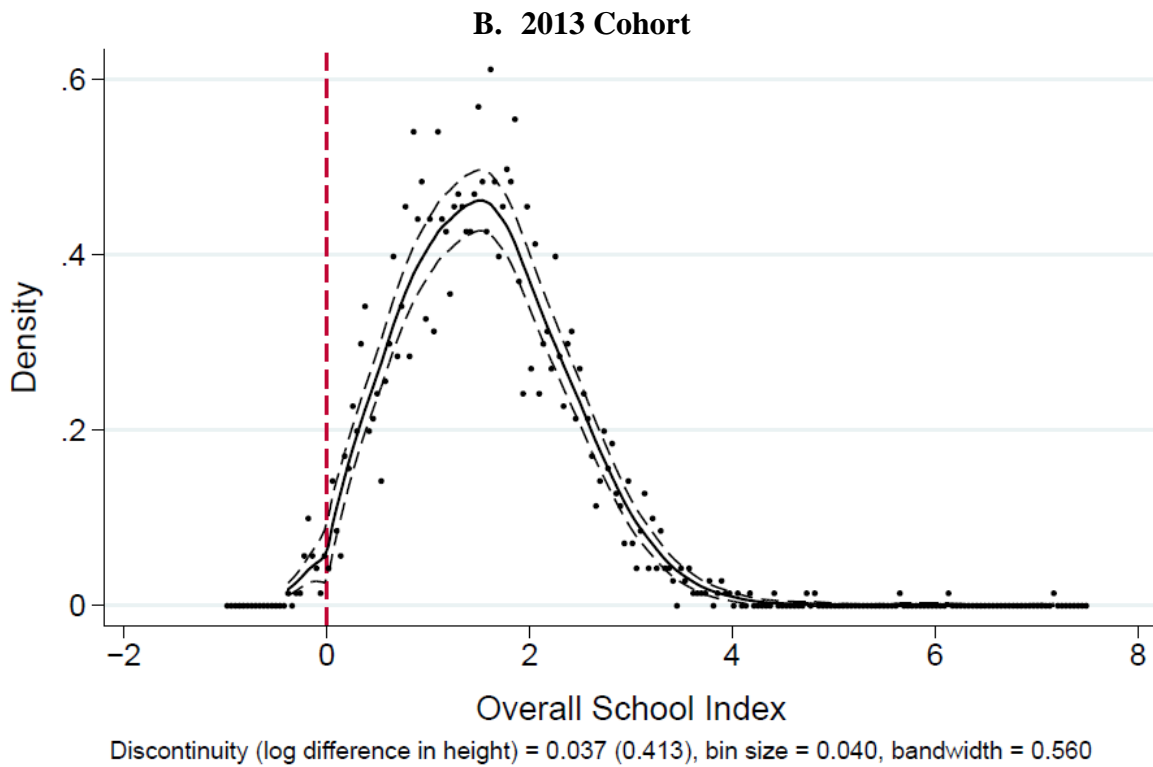
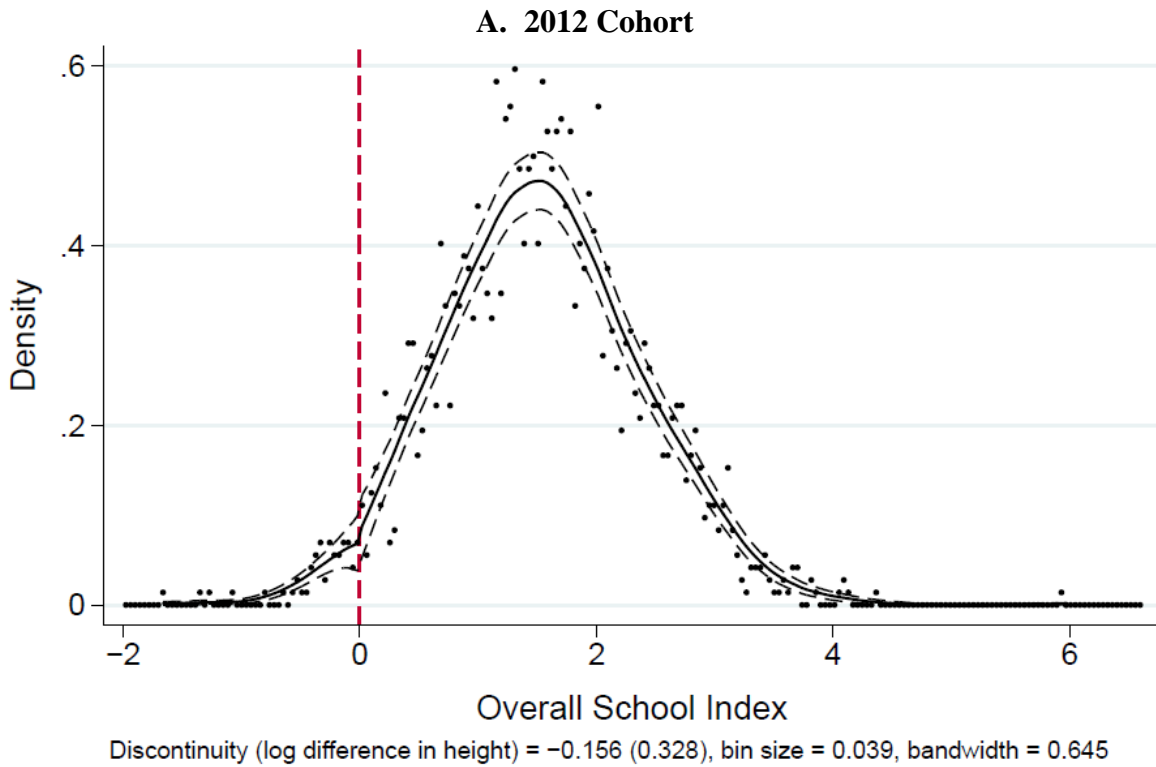
Notes: Each row within each panel (Priority, Focus) represents a separate specification, and reports effects of Priority or Focus designation on outcomes one, two, three, and four years later. Coefficients represent treatment-on-the-treated (TOT) effects estimated using the recursive approach outlined by Cellini et al. (2010). Standard errors clustered by school appear in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

**Table 9. Effects of Priority and Focus Status on Achievement: Evidence from a Dynamic RD Approach**

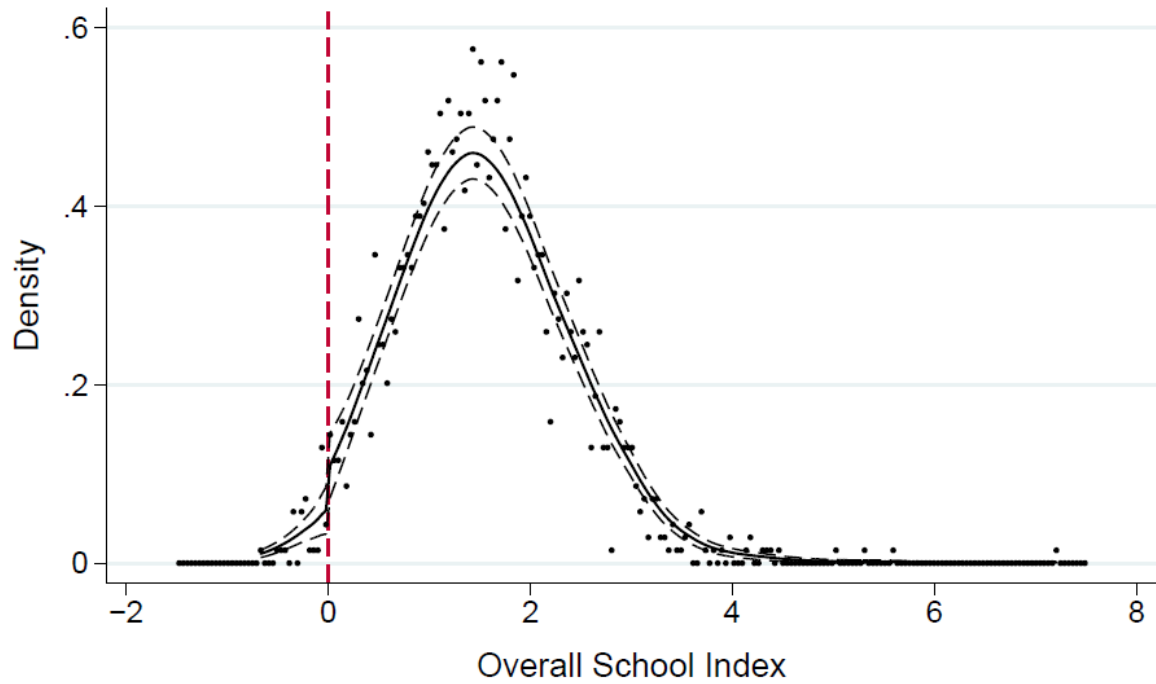
Outcome	Priority				Focus			
	1 year after (1)	2 years after (2)	3 years after (3)	4 years after (4)	1 year after (5)	2 years after (6)	3 years after (7)	4 years after (8)
<i>A. Mathematics</i>								
Average math score (std)	-0.088 (0.084)	-0.036 (0.135)	-0.034 (0.132)	-0.025 (0.225)	-0.005 (0.032)	0.001 (0.055)	-0.040 (0.056)	-0.014 (0.083)
Math gap score	-0.001 (0.081)	-0.147 (0.092)	-0.070 (0.115)	0.113 (0.192)	-0.043 (0.032)	-0.045 (0.059)	0.001 (0.064)	0.061 (0.096)
<i>B. Reading</i>								
Average reading score (std)	-0.053 (0.108)	-0.021 (0.154)	0.028 (0.160)	0.125 (0.236)	-0.004 (0.025)	-0.043 (0.046)	-0.107** (0.048)	-0.068 (0.066)
Reading gap score	0.050 (0.069)	-0.177* (0.095)	-0.063 (0.125)	-0.004 (0.143)	-0.000 (0.028)	0.005 (0.048)	0.042 (0.045)	0.071 (0.053)
N(schools)	917	917	917	917	1,654	1,654	1,654	1,654

Notes: Each row within each panel (Priority, Focus) represents a separate specification, and reports effects of Priority or Focus designation on outcomes one, two, three, and four years later. Coefficients represent treatment-on-the-treated (TOT) effects estimated using the recursive approach outlined by Cellini et al. (2010). Standard errors clustered by school appear in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

**Figure 1. Distributions of Running Variables: Priority Schools**



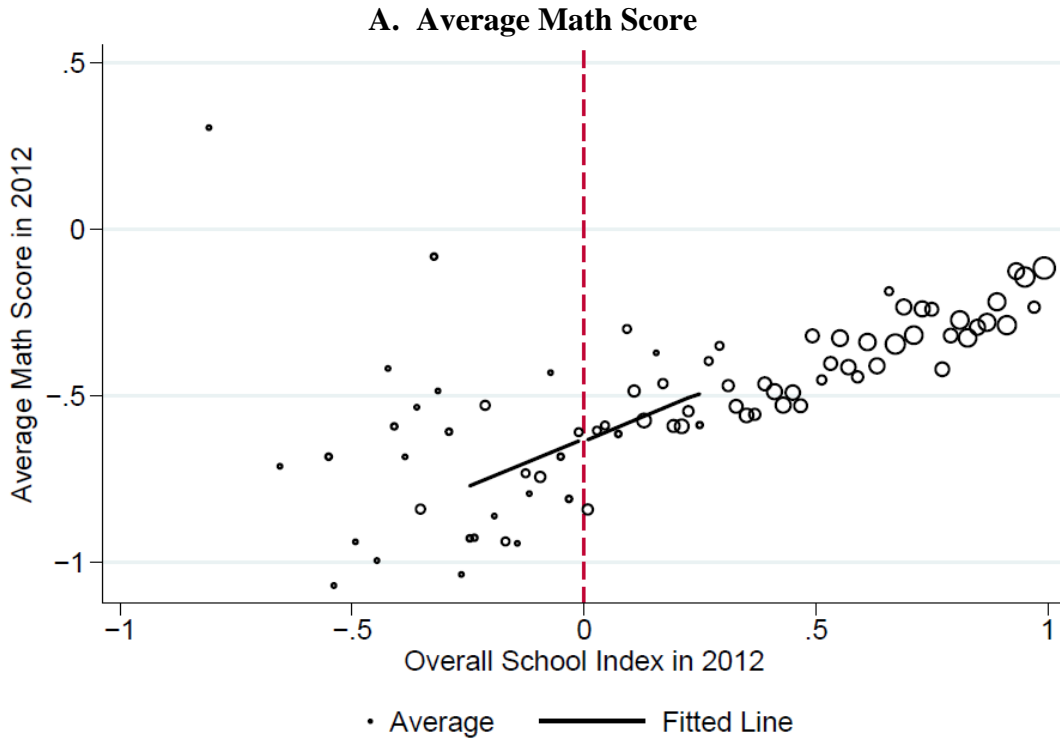
### C. 2014 Cohort



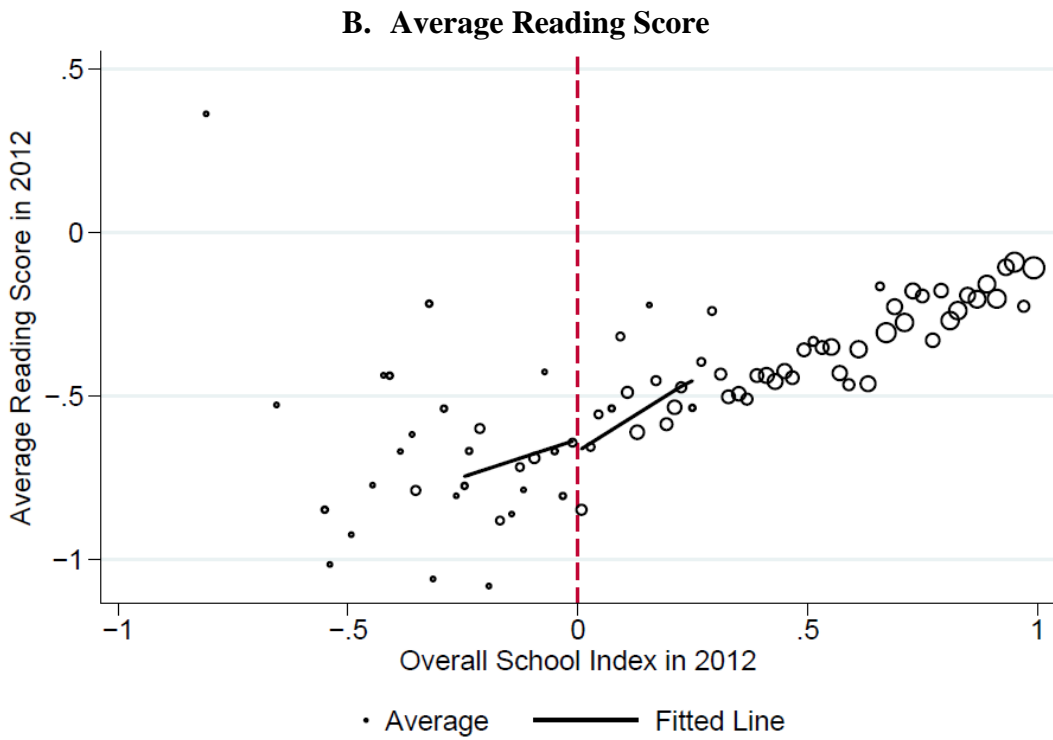
Discontinuity (log difference in height) =  $-0.535$  (0.300), bin size = 0.040, bandwidth = 0.807

Notes: Analytic samples for the 2013 and 2014 cohorts exclude schools identified as Priority in prior years. The running variable for Priority assignment is a school's zero-centered Top-to-Bottom (TTB) index value. Graphs depict McCrary (2008) test for discontinuities in density of running variable at the cutoff for Priority designation.

**Figure 2. Baseline Equivalence: Priority Schools, 2012 Cohort**

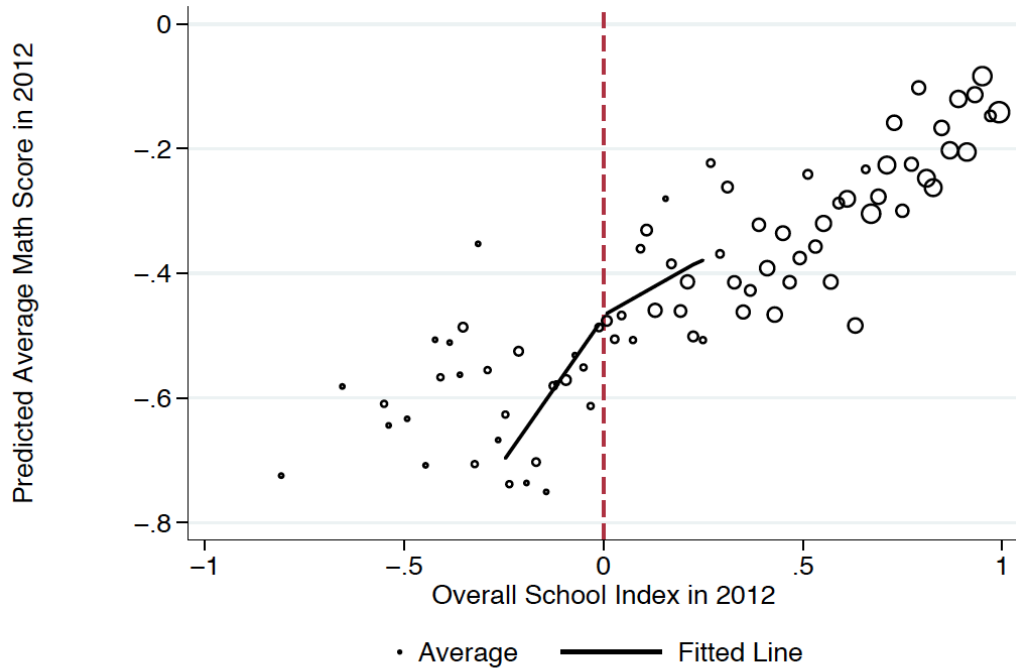


Linear: Estimate = 0.007 (0.151), N(schools) = 88, CM = -0.556, SD = 0.250  
Nonparametric (mserd): Estimate = -0.101 [-0.337, 0.159], p = 0.482, bw = 0.477, N(schools) = 188, CM = -0.627



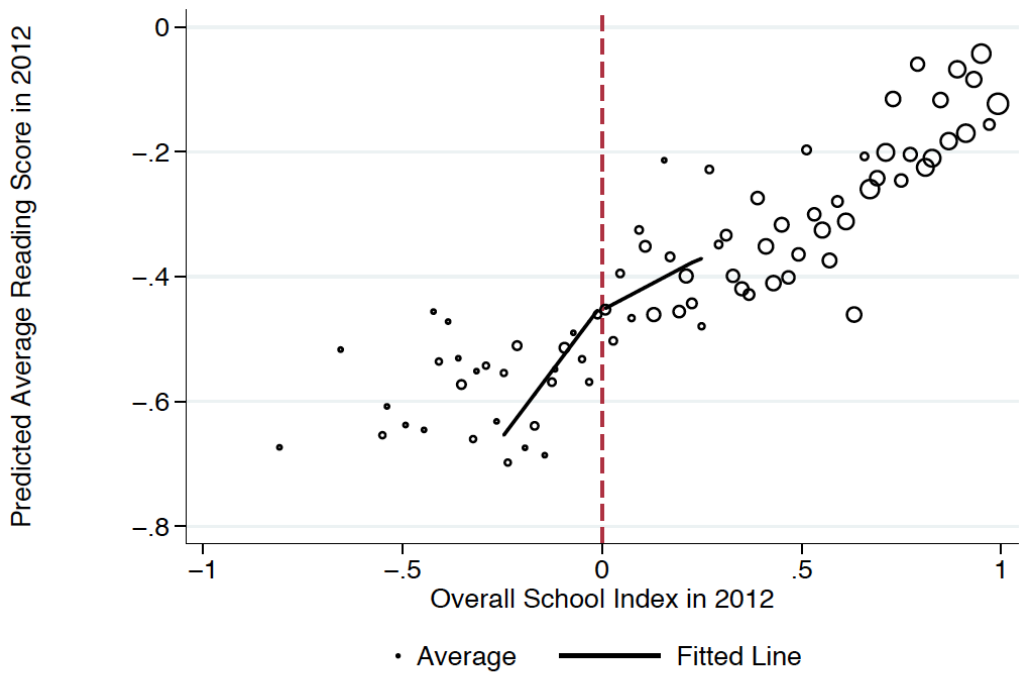
Linear: Estimate = 0.036 (0.144), N(schools) = 88, CM = -0.548, SD = 0.241  
Nonparametric (mserd): Estimate = -0.026 [-0.278, 0.266], p = 0.965, bw = 0.384, N(schools) = 136, CM = -0.677

### C. Predicted Average Math Score



Linear: Estimate = -0.004 (0.104), N(schools) = 88, CM = -0.418, SD = 0.226  
 Nonparametric (mserd): Estimate = -0.064 [-0.251, 0.181], p = 0.751, bw = 0.290, N(schools) = 95, CM = -0.463

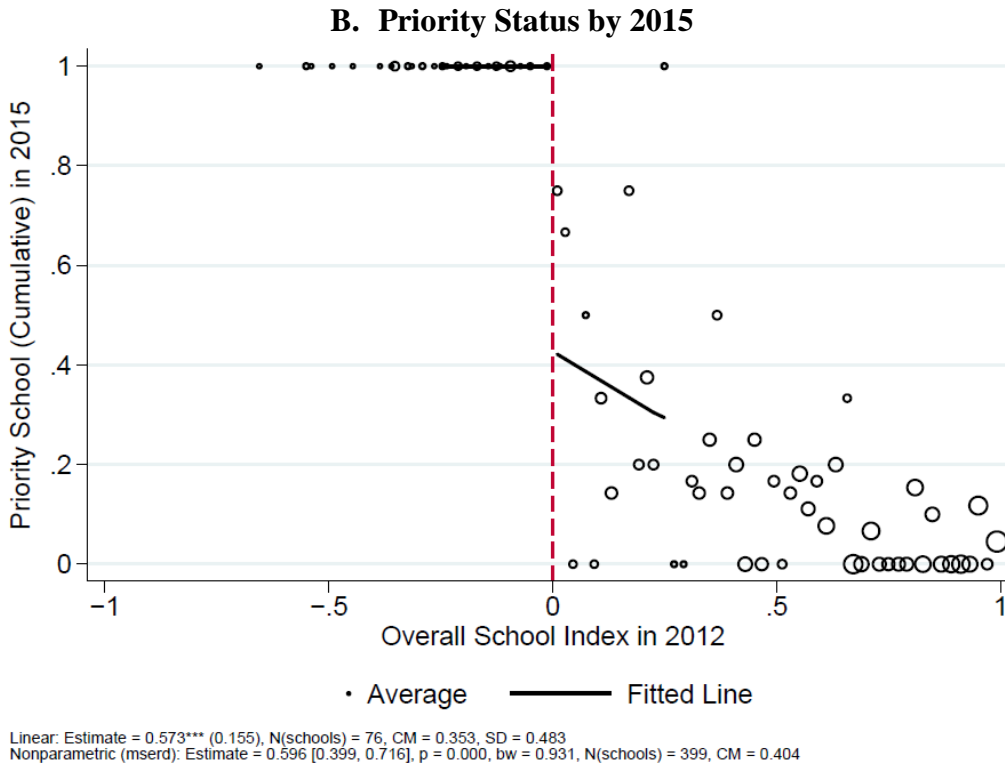
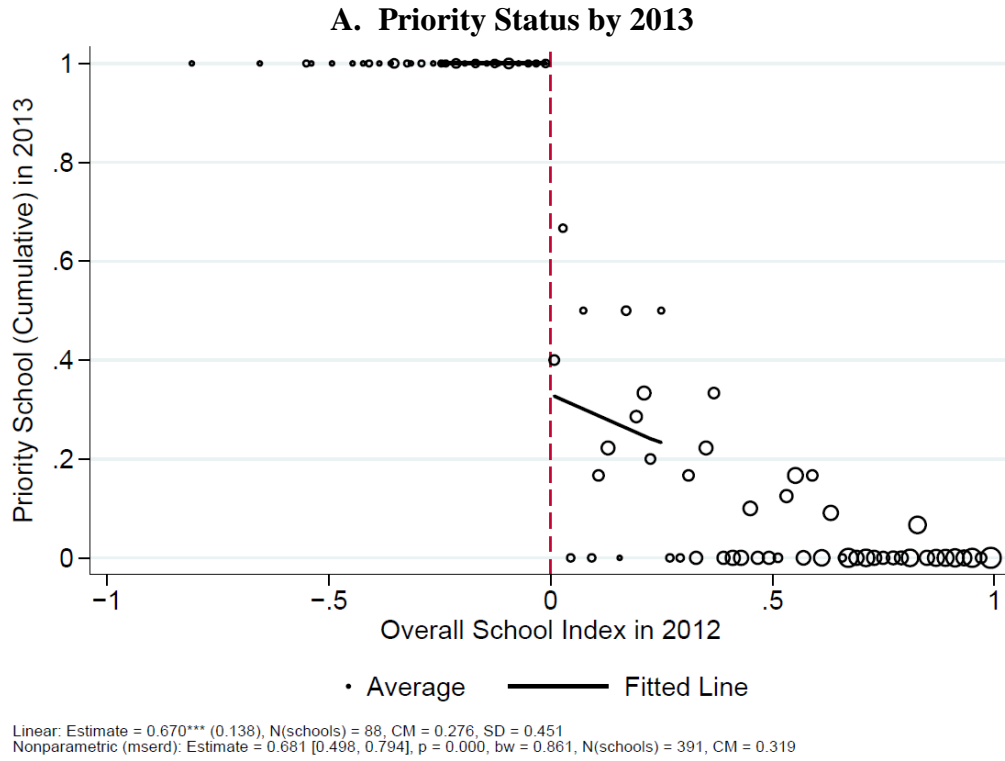
### D. Predicted Average Reading Score



Linear: Estimate = 0.009 (0.089), N(schools) = 88, CM = -0.407, SD = 0.176  
 Nonparametric (mserd): Estimate = -0.055 [-0.229, 0.153], p = 0.698, bw = 0.335, N(schools) = 114, CM = -0.445

Notes: Parametric, linear specification is weighted by school enrollment. Nonparametric estimates are based on approach of Calonico et al. (2014a, 2014b, 2015). Predicted average scores come from regressions of test scores on the set of school characteristics reported in Table 1. CM = control group mean; SD = standard deviation for control group; bw = bandwidth.

**Figure 3. Dynamic Treatment Crossover: Priority Schools**

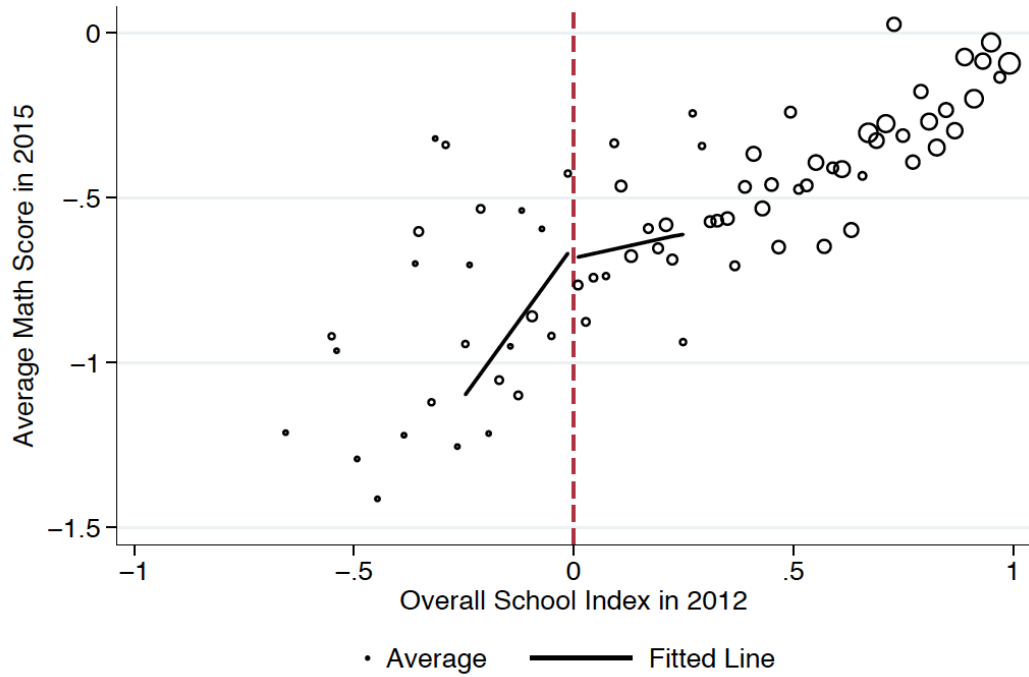


Notes: Parametric, linear specification is unweighted. Nonparametric estimates are based on approach of Calonico et al. (2014a, 2014b, 2015). CM = control group mean; SD = standard deviation for control group; bw = bandwidth.



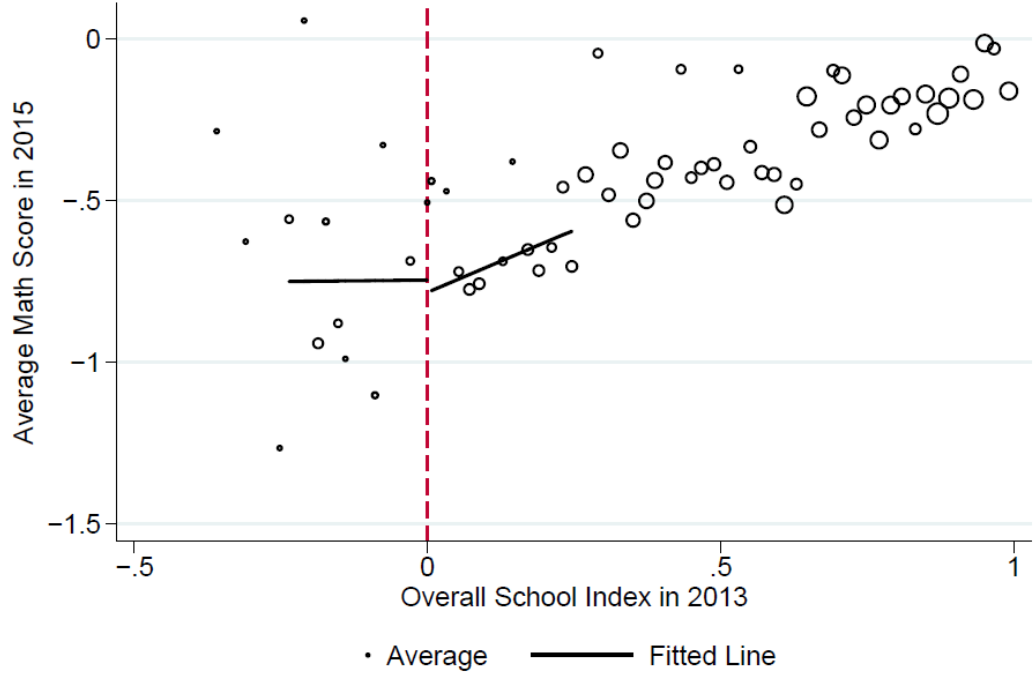
**Figure 4. Effect of Priority Status on Math Achievement**

**A. Average Math Scores, 2012 Cohort, 3 Years After Designation**



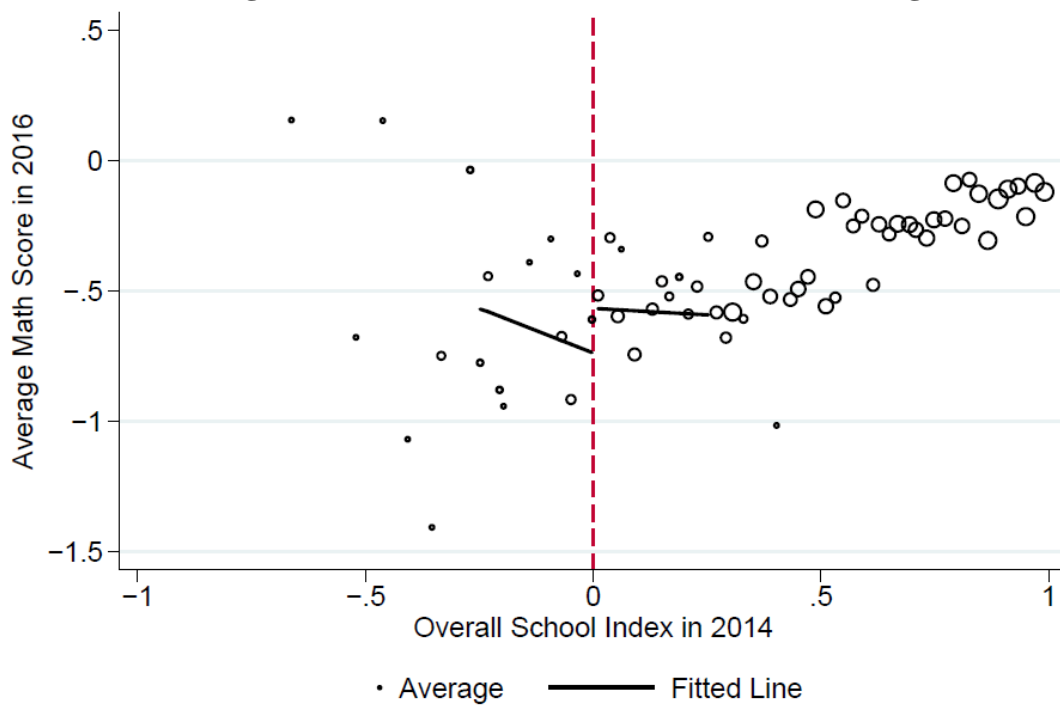
Linear: Estimate = 0.038 (0.204), N(schools) = 76, CM = -0.642, SD = 0.333  
Nonparametric (mserd): Estimate = 0.180 [-0.188, 0.734], p = 0.245, bw = 0.238, N(schools) = 73, CM = -0.758

**B. Average Math Scores, 2013 Cohort, 2 Years After Designation**



Linear: Estimate = 0.038 (0.178), N(schools) = 72, CM = -0.670, SD = 0.357  
Nonparametric (mserd): Estimate = -0.067 [-0.556, 0.423], p = 0.790, bw = 0.159, N(schools) = 34, CM = -0.499

### C. Average Math Scores, 2014 Cohort, 2 Years After Designation

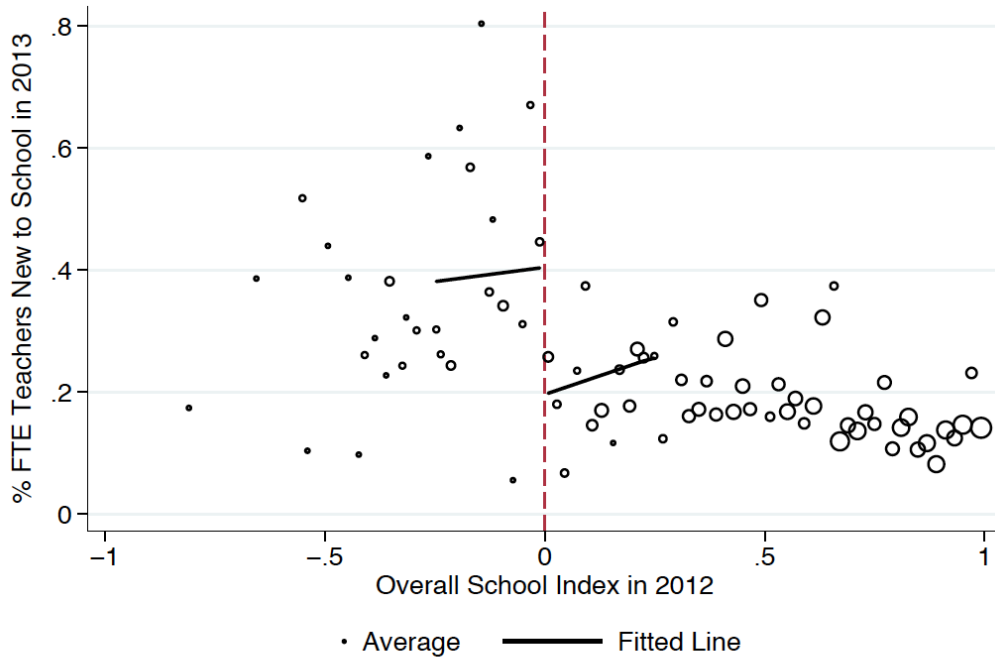


Linear: Estimate = -0.171 (0.178), N(schools) = 70, CM = -0.577, SD = 0.356  
Nonparametric (mserd): Estimate = -0.236 [-0.818, 0.291], p = 0.352, bw = 0.200, N(schools) = 54, CM = -0.477

Notes: Parametric, linear specification is weighted by school enrollment. Nonparametric estimates are based on approach of Calonico et al. (2014a, 2014b, 2015). CM = control group mean; SD = standard deviation for control group; bw = bandwidth.

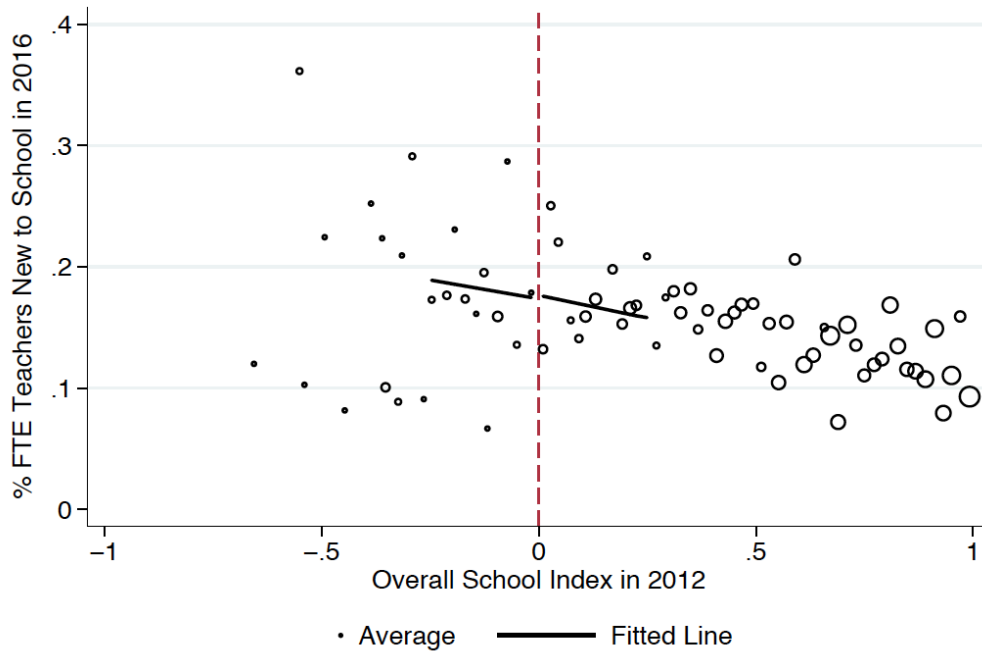
**Figure 5. Effect of Priority Status on School Staffing**

**A. Share of Teachers New to School, Cohort = 2012, Outcome year = 2013**



Linear: Estimate = 0.208\* (0.123), N(schools) = 88, CM = 0.230, SD = 0.203  
 Nonparametric (mserd): Estimate = 0.253 [0.051, 0.493], p = 0.016, bw = 0.422, N(schools) = 160, CM = 0.200

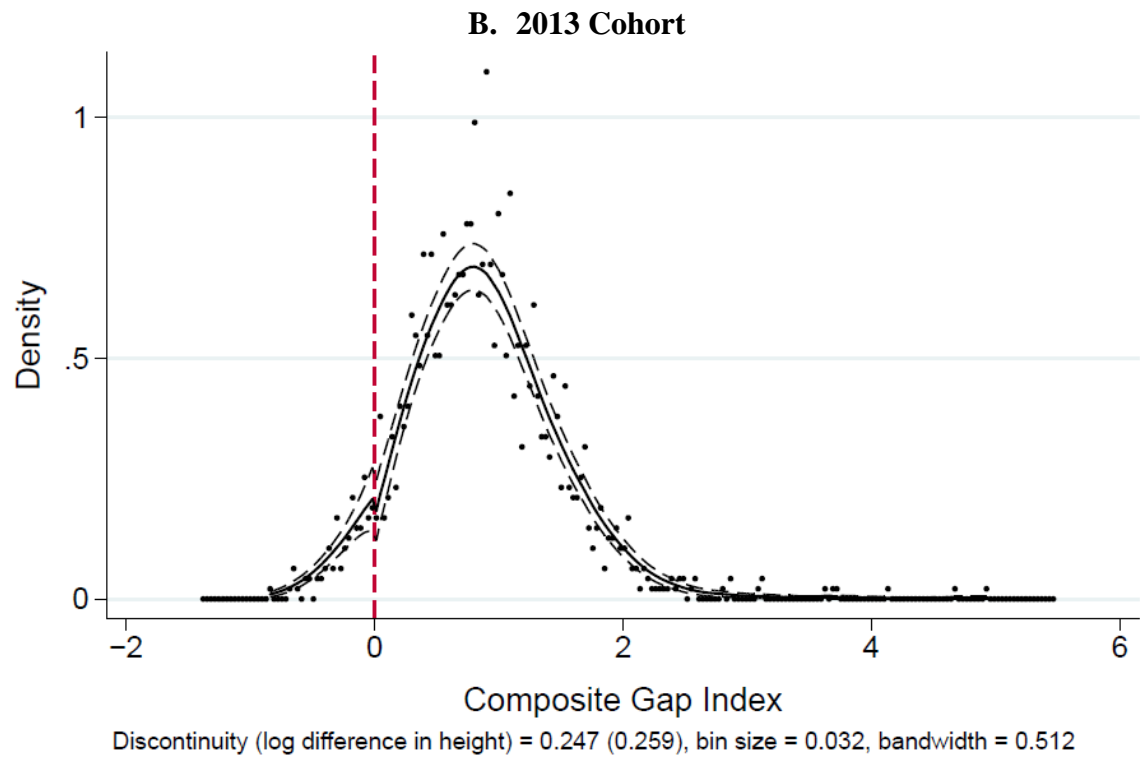
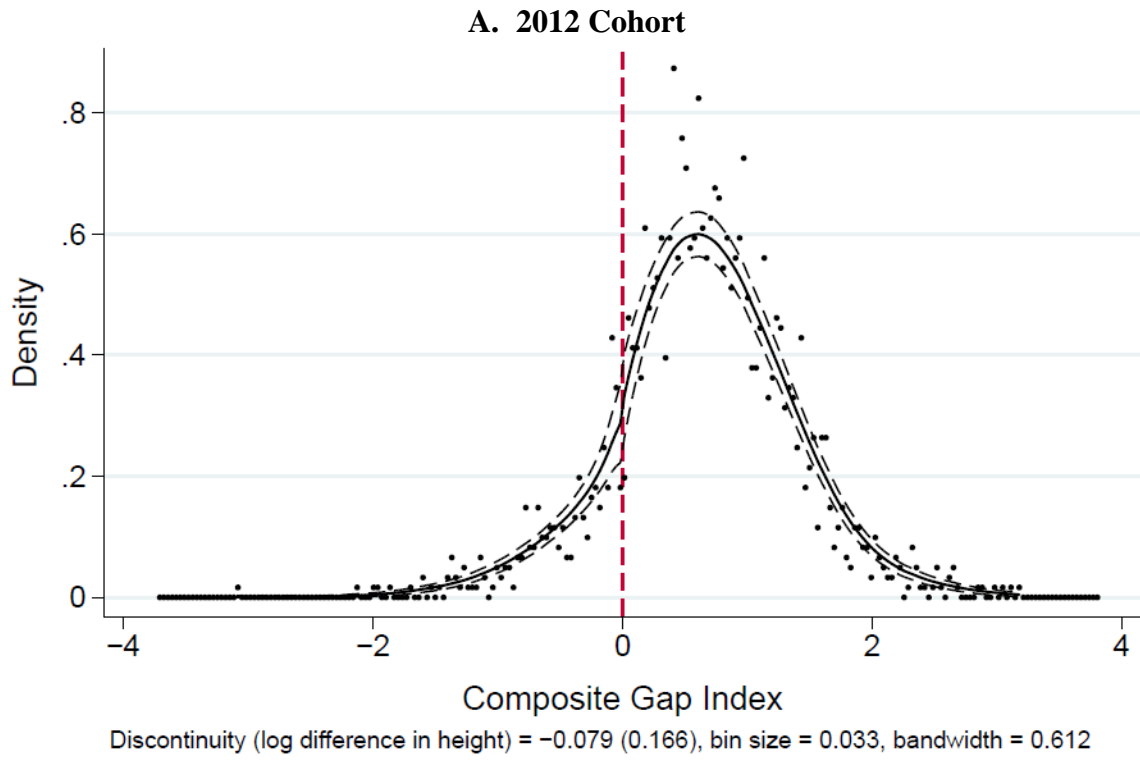
**B. Share of Teachers New to School, Cohort = 2012, Outcome year = 2016**



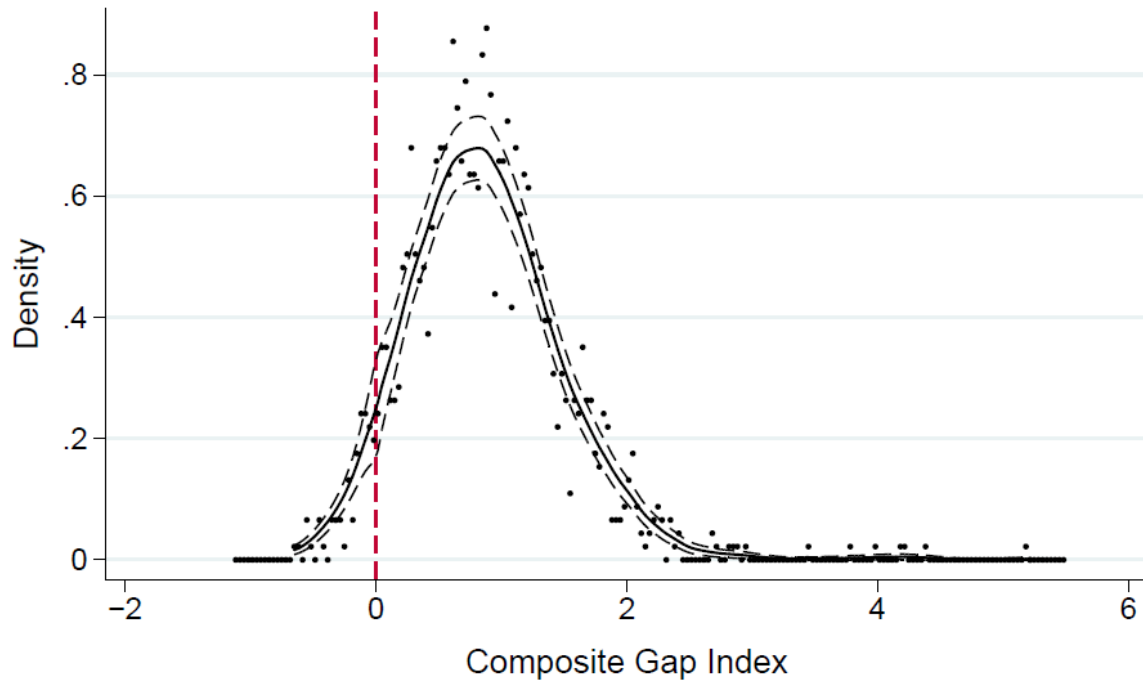
Linear: Estimate = -0.003 (0.041), N(schools) = 74, CM = 0.166, SD = 0.093  
 Nonparametric (mserd): Estimate = -0.021 [-0.149, 0.092], p = 0.643, bw = 0.256, N(schools) = 75, CM = 0.181

Notes: Parametric, linear specification is weighted by school enrollment. Nonparametric estimates are based on approach of Calonico et al. (2014a, 2014b, 2015). CM = control group mean; SD = standard deviation for control group; bw = bandwidth.

**Figure 6. Distributions of Running Variables: Focus Schools**

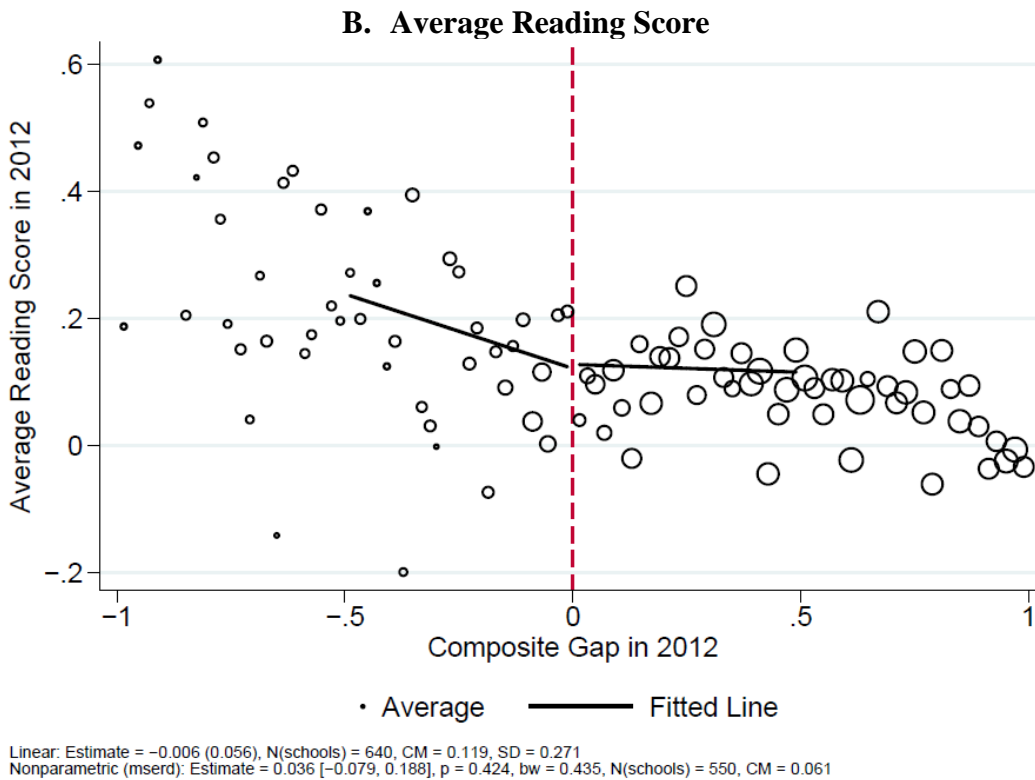
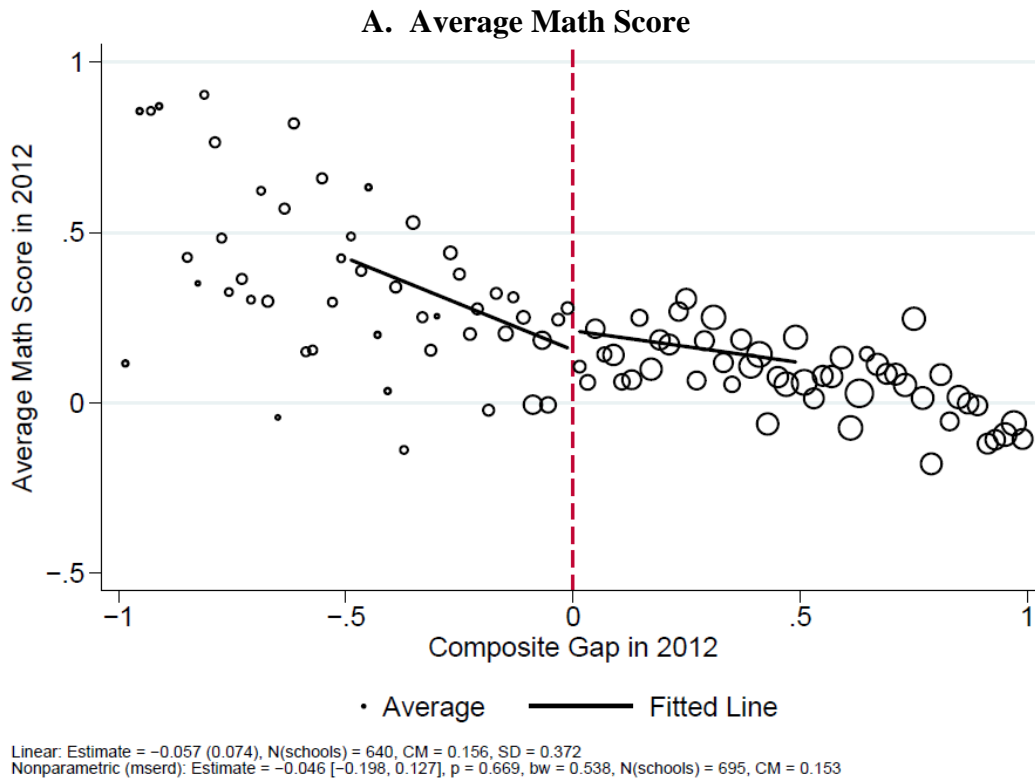


### C. 2014 Cohort

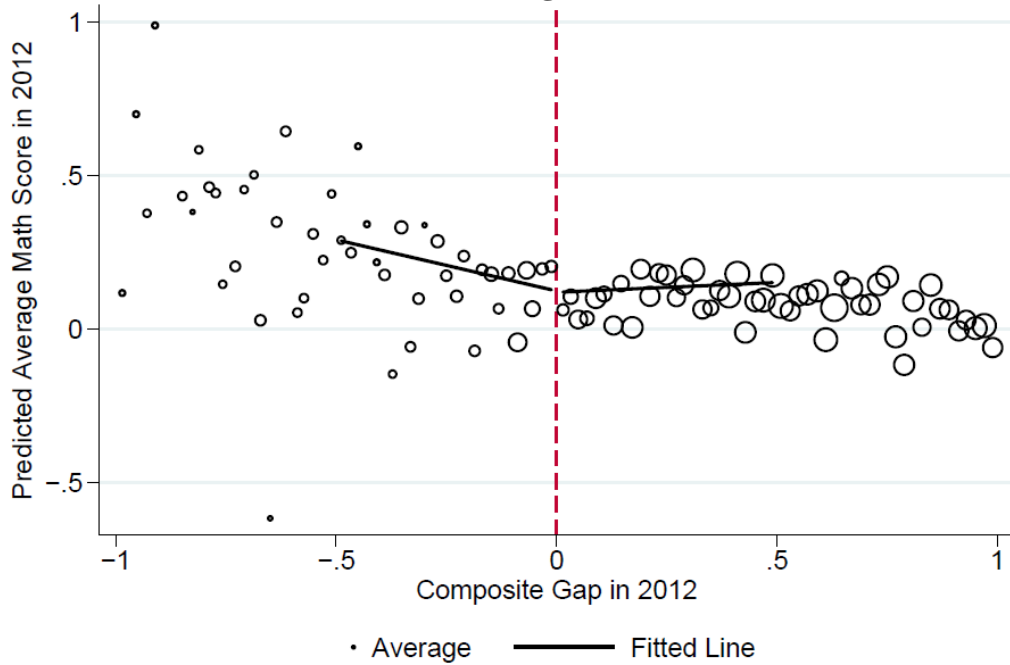


Notes: Analytic samples for the 2013 and 2014 cohorts exclude schools identified as Priority or Focus in prior years. The running variable for Focus assignment is a school's zero-centered index value that measures within-school achievement gaps between the top 30 percent and bottom 30 percent of students. Graphs depict McCrary (2008) test for discontinuities in density of running variable at the cutoff for Focus designation.

**Figure 7. Baseline Equivalence: Focus Schools, 2012 Cohort**

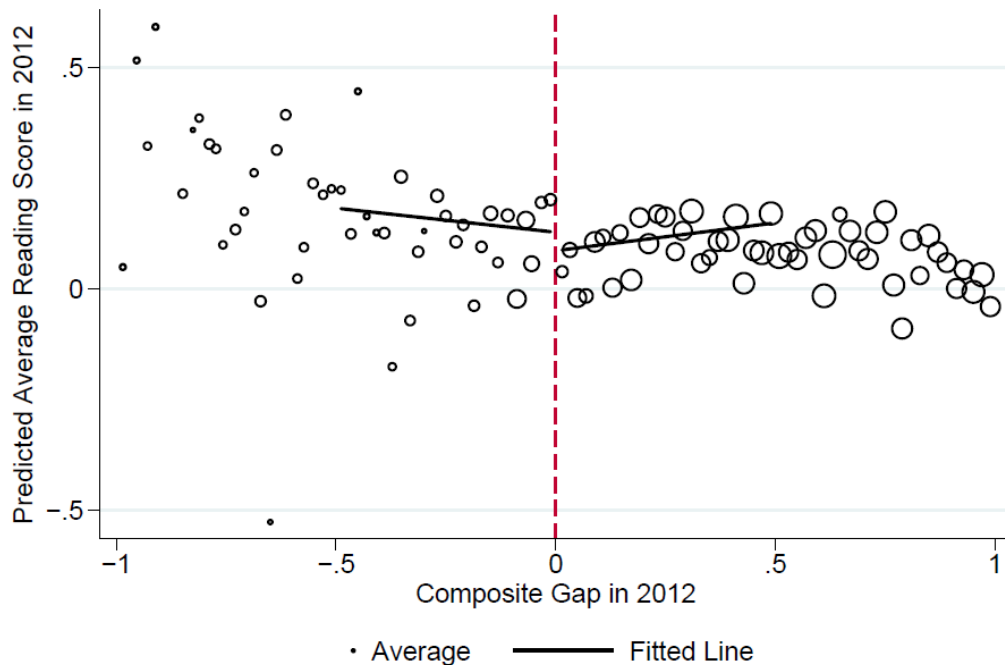


### C. Predicted Average Math Score



Linear: Estimate = 0.004 (0.055), N(schools) = 640, CM = 0.135, SD = 0.290  
 Nonparametric (mserd): Estimate = 0.067 [-0.048, 0.224], p = 0.204, bw = 0.413, N(schools) = 518, CM = 0.045

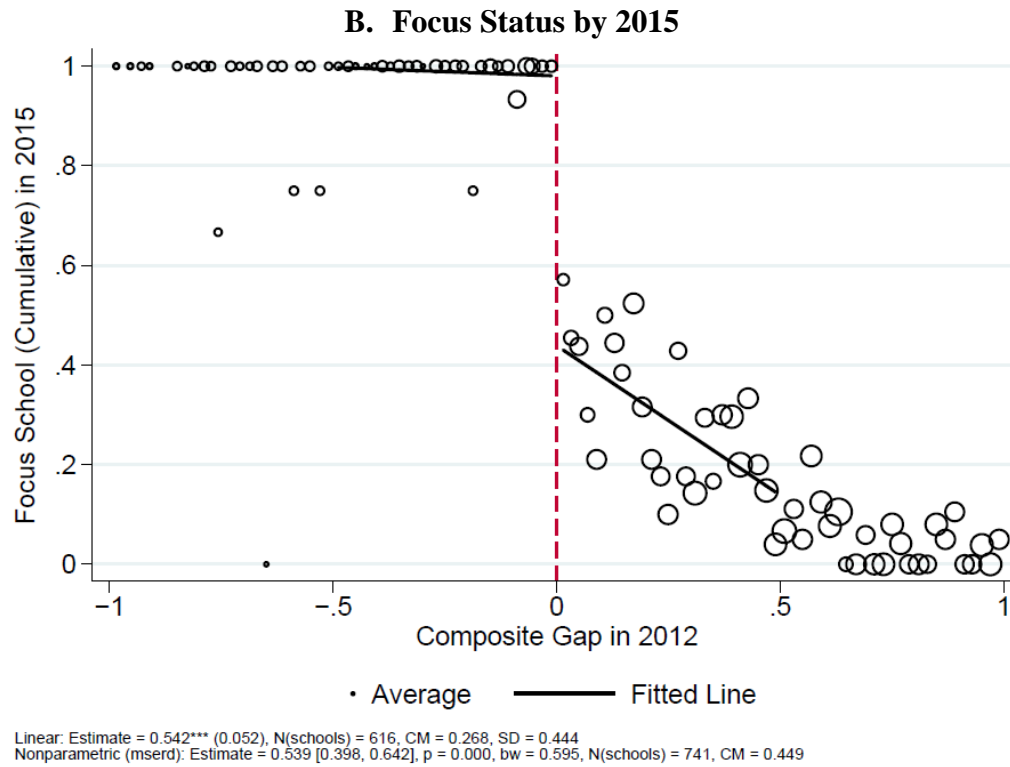
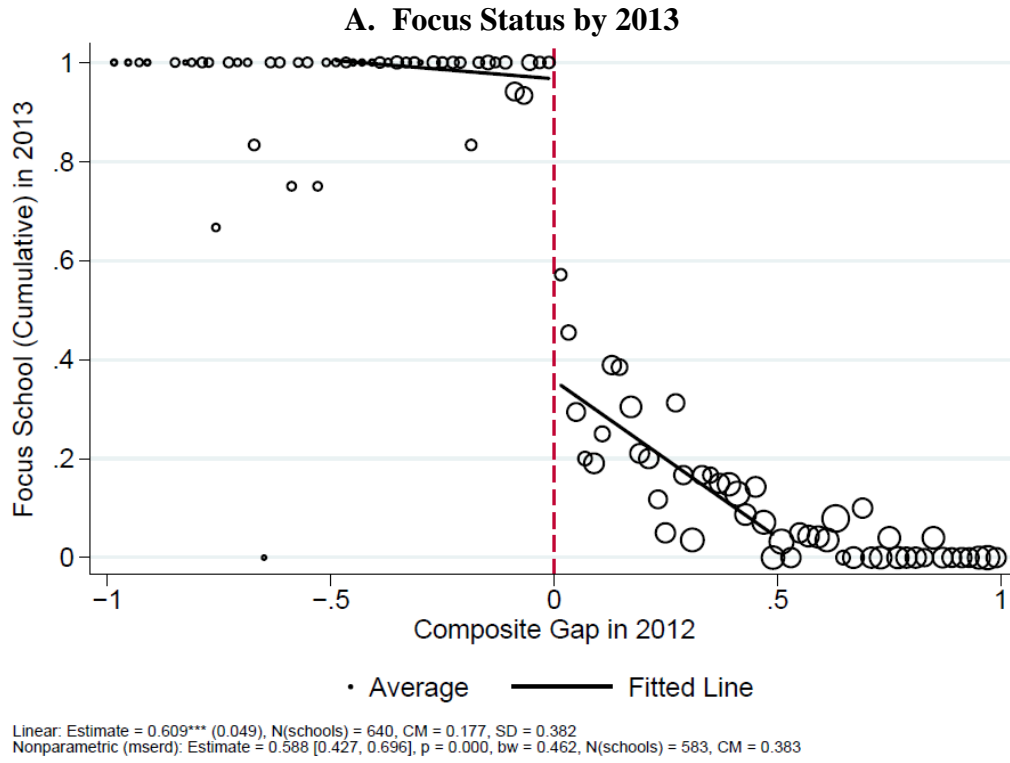
### D. Predicted Average Reading Score



Linear: Estimate = 0.042 (0.049), N(schools) = 640, CM = 0.121, SD = 0.261  
 Nonparametric (mserd): Estimate = 0.098 [-0.001, 0.238], p = 0.052, bw = 0.394, N(schools) = 484, CM = 0.018

Notes: Parametric, linear specification is weighted by school enrollment. Nonparametric estimates are based on approach of Calonico et al. (2014a, 2014b, 2015). Predicted average scores come from regressions of test scores on the set of school characteristics reported in Table 1. CM = control group mean; SD = standard deviation for control group; bw = bandwidth.

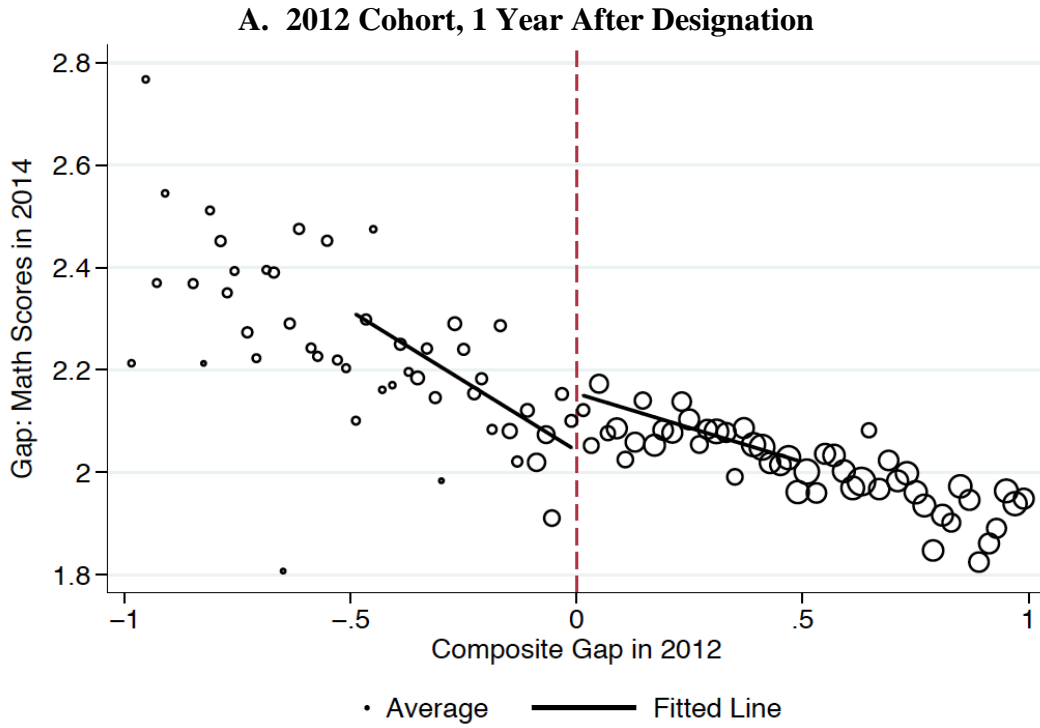
**Figure 8. Dynamic Treatment Crossover: Focus Schools**



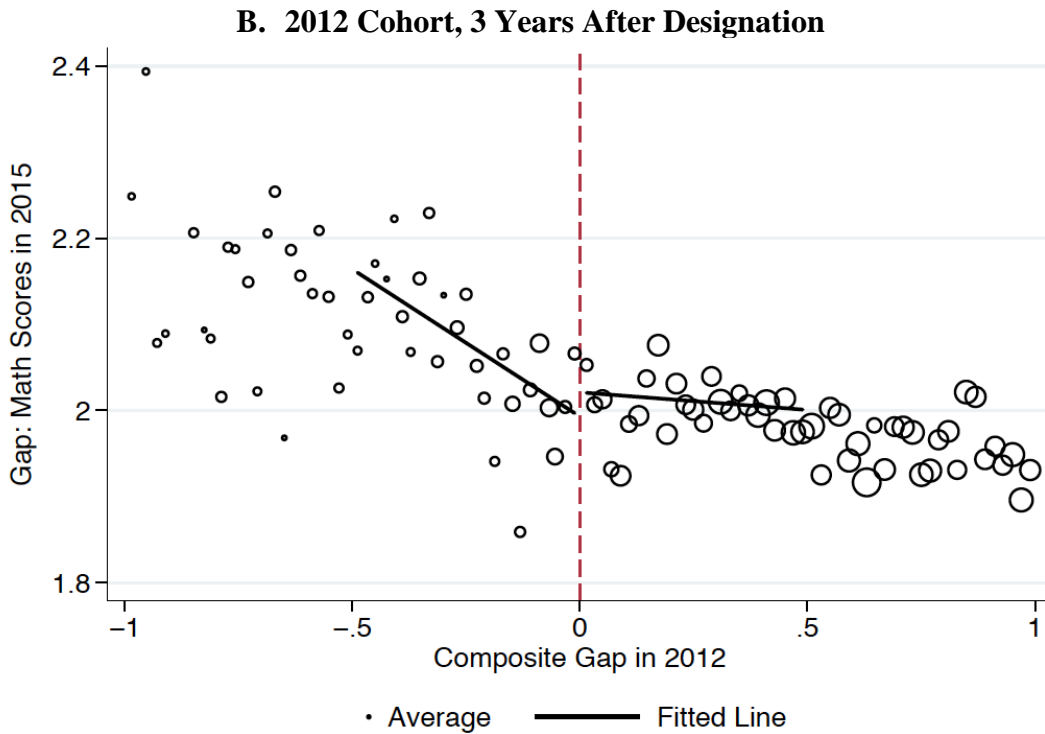
Notes: Parametric, linear specification is unweighted. Nonparametric estimates are based on approach of Calonico et al. (2014a, 2014b, 2015). CM = control group mean; SD = standard deviation for control group; bw = bandwidth.



**Figure 9. Effect of Focus Status on Math Achievement Gap**

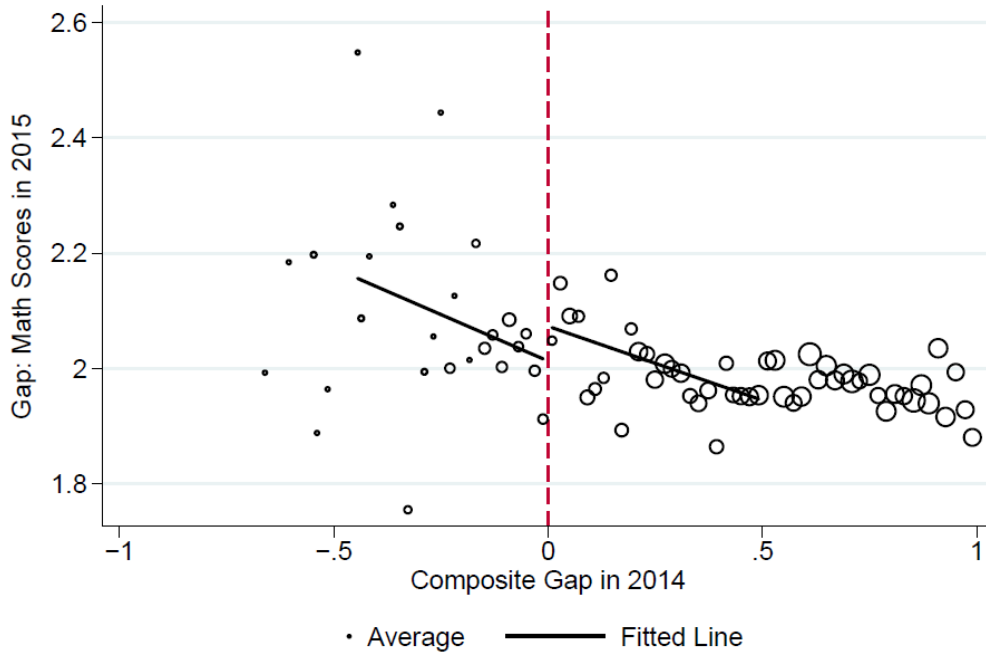


Linear: Estimate =  $-0.112^{***}$  (0.039), N(schools) = 625, CM = 2.077, SD = 0.195  
Nonparametric (mserd): Estimate =  $-0.089$  [-0.187, 0.001],  $p = 0.053$ , bw = 0.480, N(schools) = 597, CM = 2.106



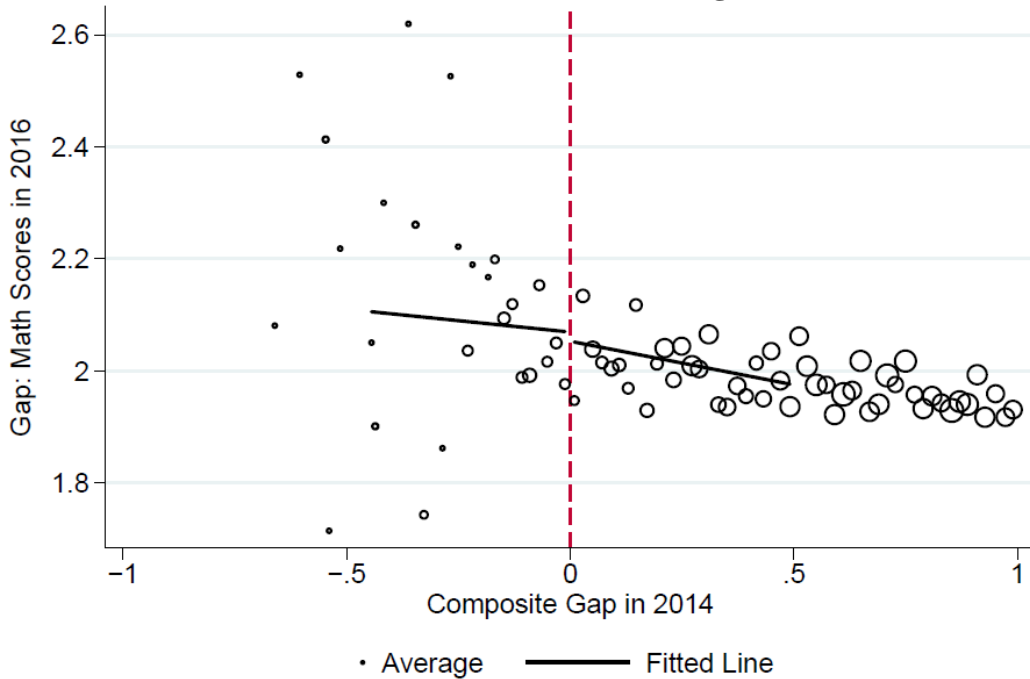
Linear: Estimate =  $-0.028$  (0.028), N(schools) = 616, CM = 2.009, SD = 0.159  
Nonparametric (mserd): Estimate =  $-0.010$  [-0.077, 0.069],  $p = 0.911$ , bw = 0.487, N(schools) = 597, CM = 1.995

### C. 2014 Cohort, 1 Year After Designation



Linear: Estimate = -0.060 (0.040), N(schools) = 366, CM = 2.004, SD = 0.152  
 Nonparametric (mserd): Estimate = -0.208 [-0.432, -0.045], p = 0.016, bw = 0.139, N(schools) = 98, CM = 2.145

### D. 2014 Cohort, 2 Years After Designation

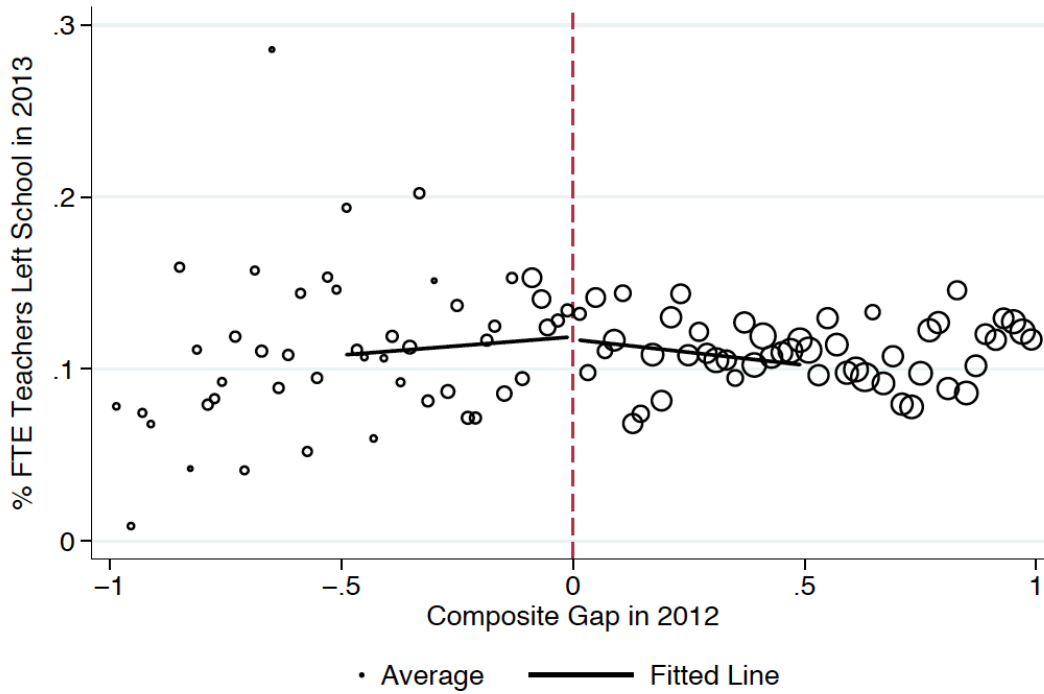


Linear: Estimate = 0.016 (0.040), N(schools) = 361, CM = 2.011, SD = 0.155  
 Nonparametric (mserd): Estimate = -0.050 [-0.210, 0.101], p = 0.492, bw = 0.185, N(schools) = 124, CM = 2.047

Notes: Analytic samples for the 2013 and 2014 cohorts exclude schools identified as Priority or Focus in prior years. Parametric, linear specification is weighted by school enrollment. Nonparametric estimates are based on approach of Calonico et al. (2014a, 2014b, 2015). CM = control group mean; SD = standard deviation for control group; bw = bandwidth.

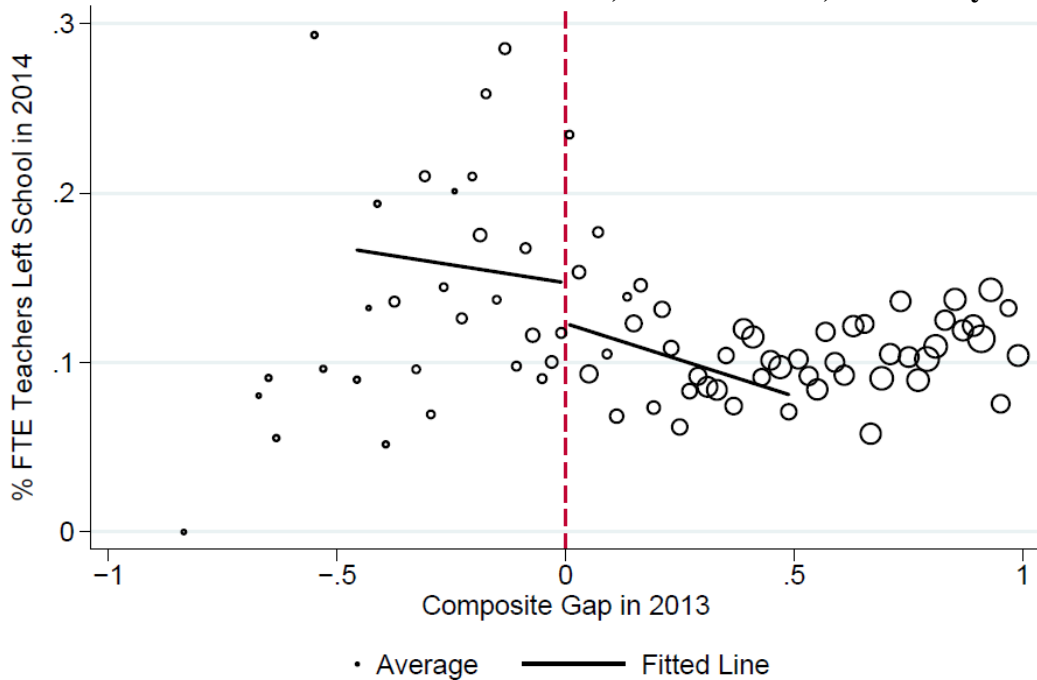
**Figure 10. Effect of Focus Status on School Staffing**

**A. Share of Teachers Left School within Past Year, Cohort = 2012, Outcome year = 2013**



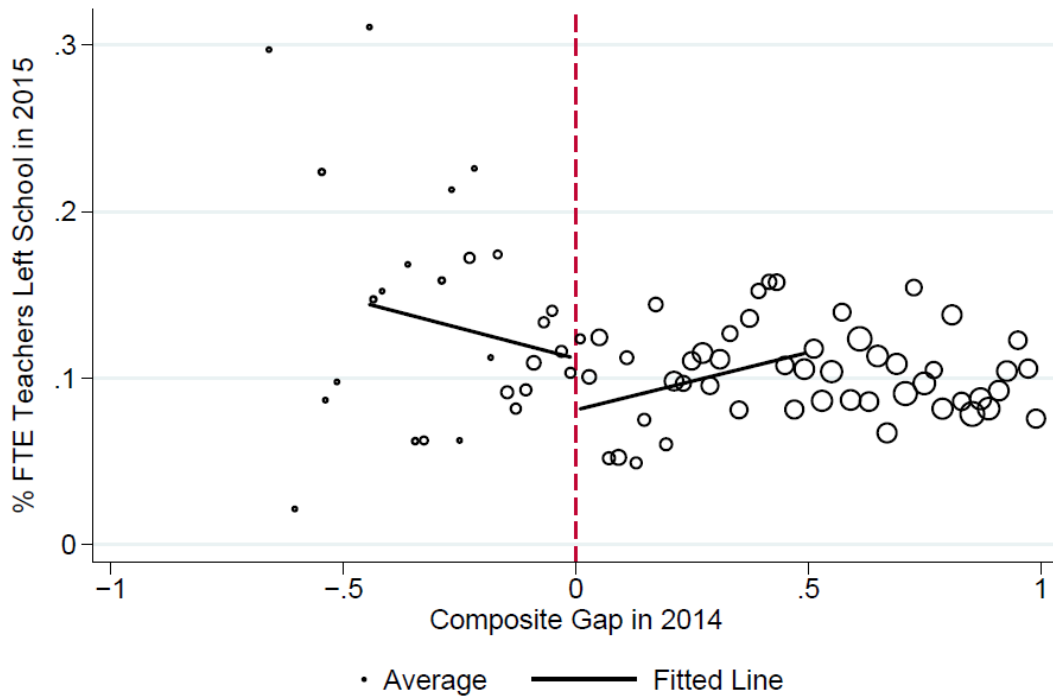
Linear: Estimate = 0.001 (0.016), N(schools) = 640, CM = 0.109, SD = 0.084  
Nonparametric (mserd): Estimate = 0.024 [-0.023, 0.080], p = 0.271, bw = 0.412, N(schools) = 517, CM = 0.116

**B. Share of Teachers Left School within Past Year, Cohort = 2013, Outcome year = 2014**



Linear: Estimate = 0.024 (0.024), N(schools) = 405, CM = 0.101, SD = 0.085  
Nonparametric (mserd): Estimate = -0.073 [-0.175, 0.007], p = 0.070, bw = 0.164, N(schools) = 108, CM = 0.161

**C. Share of Teachers Left School within Past year, Cohort = 2014, Outcome year = 2015**



Linear: Estimate = 0.031\* (0.017), N(schools) = 366, CM = 0.101, SD = 0.082  
Nonparametric (mserd): Estimate = 0.013 [-0.066, 0.082], p = 0.832, bw = 0.235, N(schools) = 162, CM = 0.105

Notes: Analytic samples for the 2013 and 2014 cohorts exclude schools identified as Priority or Focus in prior years. Parametric, linear specification is weighted by school enrollment. Nonparametric estimates are based on approach of Calonico et al. (2014a, 2014b, 2015). CM = control group mean; SD = standard deviation for control group; bw = bandwidth.