**Differentiated Accountability and Education Production: Evidence from NCLB Waivers***

Steven W. Hemelt
University of North Carolina at Chapel Hill

Brian A. Jacob
University of Michigan

August 2016

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 a noticeable impact on various measures of school staffing, student composition, or academic achievement. These disappointing 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.

* 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, and conference participants at the 2016 meetings of the Association for Education Finance and Policy (AEFP) in Denver, CO. We thank Elizabeth Mann and Max Kapustin for excellent research assistance. Hemelt can be reached at hemelt@email.unc.edu; Jacob can be contacted at bajacob@umich.edu.

## 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[1] 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 the mathematics achievement of elementary school students (Dee & Jacob, 2011) in tension with observations of curricular narrowing (Au, 2007), overemphasis of tested material (Jacob 2005; Rothstein, Jacobsen, & Wilder, 2008; Koretz, 2008), increased attention to marginally performing students at the expense of high achievers (Neal & Schanzenbach, 2010), and cheating by teachers (Fausset & Blinder, 2014; Jacob & Levitt, 2003).

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 reauthorize ESEA (as amended by NCLB). In the fall of 2011, Secretary Duncan and President Obama articulated a process by

[1] 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.

which states could request relief from some of NCLB's provisions.[2] In exchange, states had to implement reform initiatives 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 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.[3]

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 latent 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. Thus, the more open-ended nature of Focus-school reforms reflects a weaker version of this same implicit theory of change: 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

[2] 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% 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 high standards (CEP, 2013, p. 11).
[3] U.S. Department of Education, "ESEA Flexibility," Access: http://www2.ed.gov/policy/elsec/guid/esea-flexibility/index.html.

facing chronically low-performing schools as well as the specific barriers that depress the achievement of certain subgroups of students in other schools. 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. 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 examine the 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, and academic achievement. For the first cohort of schools, we find that Priority status increased the share of teachers new to a school – but that this increase is largely attributable to an uptick in enrollment. For the next two cohorts, we fail to find much of an effect on measures of teacher mobility. We do not find effects of Priority designation on math or reading achievement – neither at the mean nor for various quantiles of the within-school achievement distribution. For Focus schools, we find no effects on measures of school composition and inconclusive evidence of effects on teacher mobility. We find some evidence that Focus designation 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 NCLB, describe the waiver application process and the "reform principles" that guided waiver implementation, and situate our contributions with the literatures on K-12 accountability and whole-school reform. Section 3 describes the Michigan policy context along with details on Michigan's waiver-based reforms. Section 4 details our primary empirical approach. Section 5 describes the data we employ. Section 6 presents the main findings related to Priority interventions. Section 7 discusses findings concerning Focus schools. Section 8 explores results from a dynamic regression-discontinuity approach. Section 9 concludes and offers policy implications of our main findings.

## II. Background and Literature Review

In this section we briefly summarize work that explores the effects of NCLB on students, teachers, and schools. We next discuss the waiver process, its components, and findings from the limited literature on whole-school reform that examines effects of certain waiver elements.

### *A. The Era of NCLB*

NCLB created a framework within which data on student performance was regularly collected and reported – at the school level and for demographic subgroups of students within schools. NCLB's dual goal to improve overall student achievement and highlight gaps in achievement between advantaged and disadvantaged subgroups of students was laudable.

Implementation of any large policy initiative is warty, and NLCB's rollout was no exception. States differed in how they defined "proficiency" (Peterson & Hess, 2008; Davidson et al., 2015) and parents of students in underperforming schools often lacked information about school choice and supplemental educational services (like tutoring) to which they were entitled under NCLB (Manna, 2007).

To date, the best evidence on the effects of NCLB suggests that it improved the mathematics performance of elementary school students, but had little effect on reading achievement (Dee & Jacob, 2011). Studies that focus on particular states have arrived at varying conclusions about NCLB's effectiveness. In Maryland, Hemelt (2011) found broad, school-wide failure under NCLB unlikely to lead to short-run performance improvements; but, schools that initially failed due to a specific subgroup of students were able to improve subsequent achievement of that subgroup. West and Peterson (2006) concluded that the early sanctions in NCLB had no meaningful effects on performance in Florida. Yet, Springer (2008) found evidence (in an unidentified state) that the threat of NCLB sanctions was positively correlated with test score gains by below-proficient students in failing schools. In more recent work using North Carolina data, Ahn and Vigdor (2014) argue that the performance gains from (most of the) NCLB sanctions were concentrated among low-performing students and unaccompanied by declines among high-performers. Variation in any achievement effects generated by NCLB across states is at least partially a function of state-specific heterogeneity in implementation details (e.g., Davidson et al., 2015).

Congress was due to reauthorize ESEA/NCLB in 2007. For years, policymakers made little headway on a revised bill. Amidst entrenched Congressional inaction, the waiver announcement sought to improve on the aspects of NCLB that appeared most promising while freeing states from its most ineffective elements.

*B. NCLB Waivers and Reform Principles*

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 (2015; CEP, 2013). First, states had to adopt "college- and career-ready standards" in at least math and reading,[4] 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).[5] States were then required to field different interventions 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.

---

[4] 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 a sense of students' academic progress, regardless of state or locality.

[5] 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 2017-2018 academic year. Broadly, the law aims to provide states and districts with more flexibility (relative to NCLB) in crafting and implementing 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.

Our research focuses on the second – and arguably most salient – reform principle: differentiated accountability and associated reform interventions. 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 the recently redesigned School Improvement Grants (SIGs) funded by the 2009 stimulus package (Dee, 2012). That is, Priority schools have to 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.

State and local authorities are required to support reform efforts in Priority and Focus schools out of extant federal funding. This is a notable contrast to the recent SIG initiative (Dee, 2012), in which additional funds were injected into states and districts to support turnaround efforts in the lowest-performing schools. Yet, survey-based evidence from over 38 states with approved waivers (as of 2013) indicates that a majority of these states (i.e., 68 percent) expected the costs of designing and implementing the differentiated accountability systems to be about the same as costs under NCLB (CEP, 2013, p. 8). However, state education officials were more apprehensive about the costs of the attendant supports required by those differentiated accountability systems (CEP, 2013, p. 8). Thus, developing a functional system to identify groups of schools in need of different supports is not synonymous with providing well-developed supports to those schools.

*C. 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" (CSRD) 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 (e.g., 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.

The key features of the school reforms being implemented under NCLB waivers (i.e., their scale, their targeting, and their prescriptive design) mark this new federal effort as an important and unique addition to the existing literature on whole-school reforms. However, the components of these reform models appear to track at least some defining features of specific CSR models with comparatively strong evidence of efficacy. These practices include the use of formative assessment and data-driven instruction (e.g., Success for All), school-wide planning and community engagement (e.g., the School Development Program), and differentiated instruction (e.g., Direct Instruction).

There is some early regression-discontinuity evidence suggesting that the smaller-scale SIG-funded efforts to implement Priority-like reform models have been effective in boosting performance in one state (Dee, 2012). These results appear to be driven by schools that adopted the “turnaround” model of school reform (Dee, 2012, p. 27). Whether the much more widely implemented Priority School reforms will have similarly positive effects is unclear for at least two reasons. First, they have to be implemented in all eligible schools as opposed to only those schools that were able to craft a winning SIG application. Second, Priority-school interventions are not coupled with the same infusion of new federal resources as SIG reforms. The absence of the influx of additional resources that partially defined SIG reforms could have meaningful implications for the efficacy of waiver-based reforms.

One recent study explores the achievement effects of the accountability label that preceded the Priority-school designation in one state: persistently low-achieving (PLA) 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. Saw et al. (2016) use school-level data from Michigan in the context of a regression-discontinuity approach to study effects of this label on high school performance. The authors focus on non-SIG high schools assigned to the PLA list in August of 2010 and examine outcomes in the following academic year (i.e., 2010-2011). They compare outcomes for these high schools to those that barely missed being identified as persistently low performing. The authors find positive effects of the PLA label on subsequent writing performance, but no consistent effects on reading, math, social studies, or science test scores. Thus, this suggests that non-SIG differentiated accountability mandates may have the capacity to raise the performance of schools and students in some areas. NCLB waivers formalized and expanded the notion of enforcing different sanctions and supplying different supports based on the performance profile of a school (e.g., consistently low aggregate performance and low subgroup performance versus large gaps across subgroups in performance but typical to above-average aggregate performance).

### *D. Contributions*

We examine how Priority- and Focus-school reforms in Michigan affect the operational features of schools as well as measures of academic performance. Our research provides the first comprehensive look at the causal effects of NCLB waivers on a range of educational outcomes. These waivers 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. Beyond the direct contribution of our work on the effectiveness of this reform agenda, our research makes two other contributions. First, our work constitutes a novel addition to the literature on whole-school reforms. Second, it comments on the theories that undergird waiver and accountability-based reforms more broadly, such as collective action failures and imperfect information.

## III. Policy Implementation Context in Michigan

Michigan's waiver application was approved in July of 2012,[6] and Michigan identified its first set of Priority and Focus schools in August of 2012. For this group of schools, the 2012-2013 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 2013-2014 academic year.

[6] 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.

*A. Priority and Focus Assignment in Michigan*

Michigan adopted a cohort model of Priority and Focus school identification: once identified as a Priory 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 performance, 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 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.[7] The Priority school accountability designation replaced Michigan’s previous label of “persistently low-performing” (PLA) schools,[8] 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

[7] 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 5th 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.

[8] 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.

comprehensive index of achievement. Reward schools were those that constituted the top 5% of the TTB distribution.[9]

Focus schools were identified using one particular component of the TTB raking: 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.[10] 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.

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.[11] Thus, the dynamic nature of a school's treatment status informs our empirical approach. We tackle this issue in our methodology section below.

---

[9] 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.

[10] 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.

[11] In addition, since Priority status take 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).

*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.[12] Schools' plans must address all requirements of the chosen intervention model. For example, the turnaround and transformation models require evidence of building leadership competencies. The turnaround plan also requires that schools release and rehire up to 50% of their staff. 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.[13] 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). All Priority schools participate in "diagnostic dialogues"[14] 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

[12] 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.

[13] 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.

[14] 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.

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.[15] 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 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).[16]

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.

## 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

[15] 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).

[16] 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).

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 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: $b_j = 1\left(r_j < r^*\right)$. We are interested in the relationship between the intervention and some school outcome. Suppressing time-related subscripts, we can write:

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

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: $E\left[u_j \neq 0\right]$. 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. For example, we estimate:

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

where $P_g\left(r_j, \gamma\right)$ 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.[17] 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

[17] We employ the data-driven nonparametric commands developed by Calonico, Cattaneo, and Titiunik (2014a, 2014b, 2015).

weights of 0.5, 0.25, and 0.25 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 5th 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. 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 5th percentile is assigned Priority status and every school scoring in the bottom 10 percent of the gap measure is assigned Focus status.[18] However, each year after the initial set of schools were 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. Once a school is labeled a Focus or Priority school, it remains in the

[18] 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.

intervention for up to four years.[19] 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 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 6th to 10th percentile on the composite TTB ranking in 2012 had been assigned Priority status by Fall 2014. Roughly 40 percent of schools in the 11th to 20th 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) $$y_j = \kappa + T_j\theta + P_g\left(r_j, \lambda\right) + \varepsilon_j$$

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

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

[19] In the case of Focus requirements, a school exits Focus status after four years if it makes acceptable progress on several accountability metrics. If it fails to do so, it remains designated as a Focus school (see http://www.michigan.gov/documents/mde/AtAGlance_Overview_Focus_Schools_393918_7.pdf).

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 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. For our "treatment-on-the-treated" estimates, we define *T* as the cumulative number of years a school have been designated for intervention by a given outcome year.

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 barely 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" will lead us to understate the impact of the intervention itself.

### *C. Heterogeneous Effects*

As described in more detail below, we analyze a variety of outcomes including what some might consider intermediate outcomes such as teacher mobility and student composition as well as student achievement. Regardless of outcome, there is reason to suspect that the interventions might have heterogeneous impacts. For example, if Focus schools devote additional effort to helping the lowest-performing students in the school, one might expect the reform to have more positive effects at lower quantiles of the achievement distribution and/or bigger effects on traditionally disadvantaged demographic subgroups.

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. In order 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 disproportionally 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.

## V. Data

We combine four sources of data to study Priority and Focus reforms. We use information from the Michigan Department of Education (MDE) on schools identified as Priority or Focus going into the 2012-2013, 2013-2014, and 2014-2015 academic years. For each cohort, we obtained granular measures of the rating variables that were used to rank schools and assign Priority and Focus statuses. We then incorporate school-level variables that describe teachers and principals based on information collected by the Registry of Education Personnel (REP) at the Center for Educational Performance and Information (CEPI).[20] Finally, we augment these data with some school-level information from the Common Core of Data (CCD),[21] including measures of grade structure and total enrollment; charter, magnet, and Title I status; and geographic location (i.e., urban, suburban, rural/town).

To measure student outcomes, we use administrative, student-level data maintained by the Michigan Consortium for Education Research (MCER) at the Education Policy Initiative (EPI) at the University of Michigan. These data are longitudinal and capture all students in Michigan public schools from 2003 to 2015. We use information on students' gender, race and ethnicity, academic achievement in math and reading, special education status, as well as measures of movement between schools over time. All test scores are standardized by subject-grade-year.

In this analysis, we focus on schools serving students in grades K-8. 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.[22] Finally, we exclude schools that exclusively serve a special needs population, as they were typically not subject to the same mandates as other schools.

Table 1 presents descriptive statistics on all K-8, non-special education public schools in Michigan open in the spring of 2012, along with analogous subsets of Priority and Focus

[20] We use versions of these data cleaned and maintained by the Michigan Consortium of Educational Research (MCER) and Education Policy Initiative (EPI) at the University of Michigan.
[21] For more information on the CCD, please consult https://nces.ed.gov/ccd/.
[22] Table 1 shows that charters make up about 7 percent of all public, K-8, non-special education schools and roughly similar proportions of Priority and Focus schools.

schools. Out of the full group of public K-8 schools in Michigan, about 4 percent were labeled Priority schools and nearly 14 percent Focus schools. The share of black and Hispanic children in Michigan's public elementary and middle schools is about 27 percent, and over half of all students in those schools are considered economically disadvantaged.[23] Only 19 percent of Michigan's K-8 public schools are located in urban settings, while over 43 percent and about 38 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 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 77 percent and 90 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 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: 40 percent versus 55 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. 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

[23] This measure combines students who are eligible for free or reduced-price meals (FARM) with those who are migrants or homeless.

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 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 expect 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 +/- 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. The lines shown in the figure are the estimated regression lines from the parametric specification, with data limited to schools within +/- 0.50 from the cutoff.[24]

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 the risk-set for additional "Priority" treatment.[25] For this reason, these figures exclude any schools that have been previously identified as Priority schools. Again, we see no indication of any discontinuities at the cutoff.

[24] For the parametric models, we limit the sample to +/- 0.50 from the cutoff and include a linear term in the running variable which is allowed to differ on either side of the cutoff. Graphs, results, and conclusions are similar (if more imprecise) when we limit the data window to +/- 0.25.

[25] 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.

In results not presented but available upon request, we find small differences at the cutoff for a few individual school characteristics in the initial year (though not subsequent years), most notably mean teacher experience (point estimate 2.1with standard error of 0.8, relative to control mean of 15.1 years). To control for these small baseline differences, and increase the precision of our estimates, we include a wide range of school characteristics in our preferred specification. Specifically, we include the following control variables, 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); average teacher experience, share of teachers teaching outside of field, share of teachers highly qualified in subject, and the share of teachers from very competitive colleges.[26] 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

[26] We also control for whether a school was previously (i.e., prior to the 2012 cohort) identified as a PLA or SIG school. Main results are robust to excluding these schools from the analytic sample.

cumulative likelihood it is ever flagged for Priority interventions. Figure 3B illustrates that nearly 46 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 School Staffing and Student Composition*

Table 2 presents results from a series of parametric specifications along with their nonparametric analogues, where the outcome of interest is the share of teachers new to a school in year *t* (relative to the prior year). Estimates of the effect of Priority status on teacher mobility are quite stable as we shrink the data window on which we estimate our parametric specification and include controls. Results from our narrowest window (+/- 0.50) with controls mirror nonparametric estimates of the same parameter. This continues to be true for other sets of outcomes and thus, in subsequent tables of results, we present coefficients 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 school enrollment.

Overall, we see little effect of Priority status on teacher mobility. In Table 2, 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. In the subsequent table, we find evidence that this is largely due to the modest increase in enrollment in these schools. Figure 4 illustrates that by 2015 teacher movement due to Priority status had largely stabilized for this initial cohort. We do not find evidence of such increases in the share of new teachers as a consequence of Priority status for cohorts of 2013 and 2014. In addition, in analyses not reported, we fail to find any effect of Priority status on a range of additional measures of school staffing such as whether the principal was new to the school, teacher-student ratio, aide-student ratio, share of teachers teaching out of field, share of highly qualified teachers, or the share of teachers who earned their postsecondary degrees from highly selective institutions.

It is possible that Priority designation influenced the school choices exercised by parents. One might expect parents to pull their children out of Priority schools insofar as they view the label as an explicit sign that the school is underperforming. On the other hand, if parents were already aware of the low quality of the school, they might be attracted to newly designated Priority schools insofar as they believe such schools will be improving. In Table 3, we present estimates of several key measures of student composition.[27] We see effects of Priority status on total school enrollment for the first cohort (of 2012) and little to no effects of Priority designation on the shares of students who are black or economically disadvantaged within a school across all three cohorts. For the 2012 cohort, we find that Priority status increased total enrollment (by about 8 percent per year of Priority status (i.e., the TOT estimate in 2015)). In complementary analyses not shown, we explored whether this increase could be attributed to nearby school closures. If Priority schools were disproportionately surrounded by schools that closed relative to their barely non-Priority counterparts, this could explain the bump in enrollment. We created a measure of the number of schools that closed within 3 miles of each school in our sample during each year in our data. Even after we introduce this measure as an additional control, we continue to see a similar increase in enrollment due to Priority status. Further, this increase is spread across both entry- and non-entry-grade enrollment, with relatively larger effects on non-entry-grade enrollment. Thus, this suggests that perhaps parents in nearby schools viewed the Priority label as an indication of additional state involvement, help, and oversight with hope for improvement.

[27] For interested readers, we present nonparametric estimates of the effect of Priority status on measures of school composition in Appendix Table A1.

*D. Impacts on Academic Achievement*

In Table 4 we present estimates of the effect of Priority status on mathematics achievement – at the mean and at different points in the within-school achievement distribution.[28] For the cohorts of 2012 and 2013, academic tests were given in the early fall in Michigan (i.e., in October of the academic year), only several months after the Priority school list was released. Thus, the first plausible year in which we could expect to see effects of reforms implemented in response to being labeled a Priority school in August of 2012 (2013) is on tests given in October of 2013 (2014). By the time students in schools in the 2014 cohort were assessed, Michigan had switched to giving elementary and middle school tests in the spring of the academic year. Thus, for that one cohort, we present effects on math achievement in the academic year immediately following designation (i.e., 2014-2015). Figure 5 presents graphs by cohort that correspond to the initial effects of Priority designation on short-run, average math scores.[29]

Across the 2012, 2013, and 2014 cohorts, we find little to no evidence of effects on short-run average math performance or performance at different quantiles of within-school achievement distributions. For example, the TOT estimate for the 2012 cohort after two years is -0.019 (0.035). This allows us to rule out moderately sized positive or negative effects. The upper [lower] bound of the 95% confidence interval is 0.05 [-0.09]. So, at best, each year of priority designation is associated with a 0.05 increase in standardized math achievement. Granted, our estimates for later cohorts are relatively imprecise.

[28] For interested readers, we present nonparametric estimates of the effects of Priority status on average math achievement in Appendix Table A2.
[29] The linear parametric RD estimates below each graph come from specifications without control variables.

## 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 school composition, staffing, and student achievement 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 expect 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 A3 and A4 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.[30] For the cohort of 2014, we see little evidence of differences in math achievement, reading achievement, or measures of school composition at the cutoff.

*B. First-Stage Impact on Treatment*

Figure 8 illustrates the extent of dynamic treatment crossover for the Focus treatment.[31] 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 interventions. 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 Staffing and Composition*

Table 5 presents results from a series of parametric specifications along with their nonparametric analogues, where the outcome of interest is the share of teachers new to a school in year *t* (relative to the prior year). Again, since our preferred parametric specification yields results similar to the nonparametric estimates, we only present results from the preferred

[30] For this particular cohort, weighting by school enrollment exacerbates the difference in baseline math performance we estimate at the Focus cutoff, suggesting the difference is driven by larger schools near the threshold. In all subsequent outcome models we control for baseline achievement in math and reading. Nevertheless, we also caution readers in the standalone interpretation of this one cohort.

[31] 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.

parametric model which controls for baseline covariates and weights by enrollment, in subsequent tables. The results in Table 5 suggest that Focus status is associated with lower teacher mobility for the 2012 cohort by the third year after identification and higher teacher mobility for the 2013 cohort by the second year after designation. It is unclear why any effects on teachers would be so delayed for the first cohort of Focus schools.

In Table 6 we explore the effects of Focus designation on the composition of students in schools.[32] Across all three cohorts, we find no evidence that Focus status influenced total enrollment, the share of students who are black, or the share of economically disadvantaged students in a school. 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, the share of students in special education, and the share of students with limited English proficiency.

*D. Impacts on Academic Achievement*

In Table 7, we turn to the effects of Focus designation on measures of academic achievement, with a special focus on the within-school math gap.[33] Since the principal aim of the Focus label was to bring attention to large within-school gaps in achievement, this outcome is of prime importance in assessing the efficacy of any attendant interventions implemented by Focus schools. Figure 9 presents corresponding graphs for each cohort where the outcome is the math gap score.[34] The linear parametric RD estimates below each graph come from specifications

[32] For interested readers, we present nonparametric estimates of the effects of Focus designation on measures of school composition in Appendix Table A3.
[33] We present nonparametric estimates of the effects of Focus status on average math scores and the within-school math gap measure in Appendix Table A4.
[34] This outcome measure is slightly different than the variable that functions as the running variable for Focus assignment. That running variable is calculated over two years of prior math scores (see Appendix B for details). We generate our outcome based on data from each single year and calculate the gap between the top 30 percent of

without control variables. Estimates from our preferred parametric specification presented in Table 7 tell the same qualitative story as the graphs but are relatively more precise.

For two of the three cohorts (i.e., 2012 and 2014), we see some evidence that Focus schools shrunk within-school gaps in math performance more than their barely non-Focus counterparts. Unfortunately, estimates across the achievement distribution suggest that decreases in this 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. For the 2013 cohort, we fail to detect effects of Focus designation on math performance or the map gap score. Though, this is the cohort for which tests of the core RD assumptions using baseline data are mildly concerning. Looking at the cohort for which we can see three years after Focus identification (i.e., the 2012 cohort), we observe declines in average math performance at the mean and across the entire within-school achievement distribution. Taken together, these results suggest that any short-run narrowing of within-school gaps was short-lived and driven by the combination of steady performance of lower-achieving students alongside declines in the performance of higher-achieving students.[35] This is a disappointing finding that runs against the hope of Focus interventions – namely that all students would improve their performance, with larger achievement boosts accruing to initially lower-performing students.[36]

students and bottom 30 percent of students, following the general steps of the state's calculation rules outlined in Appendix B.

[35] 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.

[36] The largest negative effects come from the 2015 testing cycle, in which Michigan conducted its first administration of a new test, the M-STEP. M-STEP replaces the MEAP test used in all prior years. Nonetheless, this transition should have no bearing on our findings since we standardize exam scores within grade-year and our RD approach and estimates should approximate a randomized experiment. For additional information on the M-STEP test, please see http://www.michigan.gov/mde/0,4615,7-140-22709_70117---,00.html.

## 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 2014-2015 (2015). For the 2012 ranking, $\tau$ runs from -3 through +2. For the 2013 and 2014 rankings, respectively, $\tau$ runs from -3 to +1 and -3 to 0. Hence, a school that existed over the entire six-year period would be included 6 + 5 + 4 = 15 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 6 annual observations. Likewise, a school that is designated Priority in 2013 will appear in two ranking cohorts, for a total of 11 observations.[37]

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

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

in which $y_{jt\tau}$ represents the outcome for school *j* in the *t* ranking cohort, τ 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 τ years after the ranking. The estimates of $\theta_\tau^{ITT}$ capture the effect of missing the cutoff on school outcomes τ years later. The equation includes fixed effects for academic year (κ) and year relative to ranking (α). Fixed effects for school-by-cohort (λ) absorb across-school variation. $P_g\left(r_{jt}, \gamma_\tau\right)$ 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 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

[37] 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.

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.

Table 8 presents the TOT estimates on measures of school staffing and student composition. Overall, we fail to find appreciable effects of Priority or Focus status on any of these outcomes. In Table 9, we present results from the dynamic RD setup on measures of math achievement.[38] We find no effect on average math performance, nor do we find noticeable effects on achievement at various quantiles of the within-school achievement distribution. We note that the TOT coefficients representing effects on test scores three years after initial Focus designation are negative, but none rise to conventional levels of statistical significance.

## 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

[38] As described in the cohort-by-cohort approach, the first year in which test scores could have plausibly been affected by the interventions is the second academic year after identification as a Priority or Focus school.

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 some evidence of small effects on candidate mediators such as the share of teachers new to a school and student enrollment. Yet, when we look to impacts of Priority status on measures of achievement, we detect no effects on short-run math or reading performance – neither at the mean nor at various quantiles of the within-school achievement distribution. Overall, we see little effect of Priority designation on measures of school staffing, composition, and achievement.

Our analysis of Focus schools yields similarly uninspiring findings. Although we find some evidence that Focus designation is associated with 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.

Taken together, our findings paint a discouraging picture of the capacity for Priority and Focus reforms to improve outcomes for students and schools. At first blush, one might be optimistic about the accountability-related provisions in the recent reauthorization of the ESEA/NCLB if one thinks that the lack of impacts for Priority schools stems from the fact that the required reform activities were overly prescriptive, divorced from supporting funds, and failed to consider the number, type, and character of local educational and community challenges. The Every Student Succeeds Act (ESEA) 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). Yet, under the waiver-based reforms, states had a large degree of autonomy in fielding interventions in Focus schools and we find little benefit of those myriad interventions on student performance. In addition, 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). Indeed, 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).

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. Perhaps the targeting of a portion of state-level Title I funds toward the lowest-performing schools in a district mixed with a more multifaceted approach to identifying those schools that moves beyond variants of test score measures will prove beneficial. At minimum, our findings that document a general lack of effects generated by Priority- and Focus-school reforms should serve as a cautionary tale.

## REFERENCES


Ahn, T., & Vigdor, J. (2014). The impact of No Child Left Behind's accountability sanctions on school performance: Regression discontinuity evidence from North Carolina. NBER Working Paper Series, No. 20511. Access: http://www.nber.org/papers/w20511

Au, W. (2007). High-stake testing and curricular control: A quantitative metasynthesis. *Educational Researcher, 36(5),* 258-267.

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

Borman, G. D., Hewes, G. M., Overman, L. T., & Brown, S. (2003). Comprehensive school reform and achievement: A meta-analysis. *Review of Educational Research, 73,* 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, Access: http://www.cep-dc.org/cfcontent_file.cfm?Attachment=McMurrerYoshioka%5FReport%5FStatesPerspectivesonWaivers%5F030413%2Epdf

Davidson, E., et al. (2015). Fifty ways to leave a child behind: Idiosyncrasies and discrepancies in states' implementation of NCLB. *Educational Researcher, 44(6),* 347-358.

Dee, T.S. 2012. School turnarounds: Evidence from the 2009 Stimulus. NBER Working Paper Series, No. 17990. Access: 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.

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

Fausset, R., & Blinder, A. (2014, April 14). Atlanta school workers sentenced in test score cheating case, *New York Times*, p. A1.

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.

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.

Imbens, G. W., & Kalyanaraman, K. (2012). Optimal bandwidth choice for the regression discontinuity estimator. *Review of Economic Studies, 79,* 933-959.

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.

Keesler, V. (2016, March 18). Personal communication.

Manna, P. (2007). NCLB in the states: Fragmented governance, uneven implementation. In F. M. Hess, & C. E. Finn Jr. (Eds.), *No remedy left behind: Lessons from a half-decade of NCLB* (pp. 17-42). Washington, DC: AEI Press.

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

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." Access: http://www.ncsl.org/documents/educ/ESSA_summary_NCSL.pdf

Neal, D., & Schanzenbach, D. W. (2010). Left behind by design: Proficiency counts and test-based accountability. *Review of Economics and Statistics, 92*, 263–283.

Peterson, P. E., & Hess, F. M. (2008). Few states set world-class standards: In fact, most render the notion of proficiency meaningless. *Education Next*, Summer, 70-73.

Reback, R. (2008). Teaching to the rating: School accountability and the distribution of student achievement. *Journal of Public Economics*, 92, 1394-1415.

Rothstein, R., Jacobsen, R., & Wilder, T. (2008). *Grading education: Getting accountability right.* New York: Teachers College Press.

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.

Springer, M. G. (2008). The influence of an NCLB accountability plan on the distribution of student test score gains. *Economics of Education Review*, *27*, 556-563.

U.S. Department of Education. (2015, June 7). ESEA Flexibility. Access: http://www2.ed.gov/policy/eseaflex/approved-requests/flexrequest.doc

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;" Access: 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.* Access: https://www2.ed.gov/admins/lead/account/monitoring/reports13/mipartbrpt2014.pdf

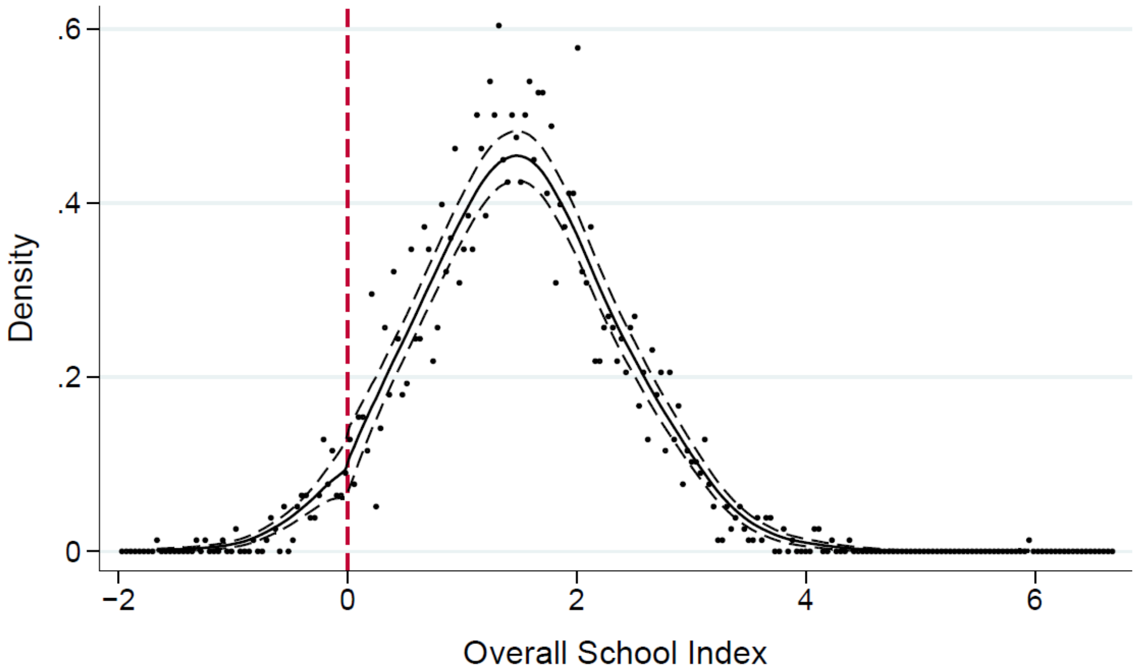
U.S. Department of Education (2010). *Evaluation of the comprehensive school reform program implementation and outcomes: fifth-year report.* Washington DC: Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service.

U.S. Government Accountability Office (2016). *K-12 Education: Education's experiences with flexibility waivers could inform efforts to assist states with new requirements.* Access: http://www.gao.gov/assets/680/678258.pdf

West, M. R., & Peterson, P. E. (2006). The efficacy of choice threats within school accountability systems: Results from legislatively induced experiments. *The Economic Journal, 116(510),* C46-C62.

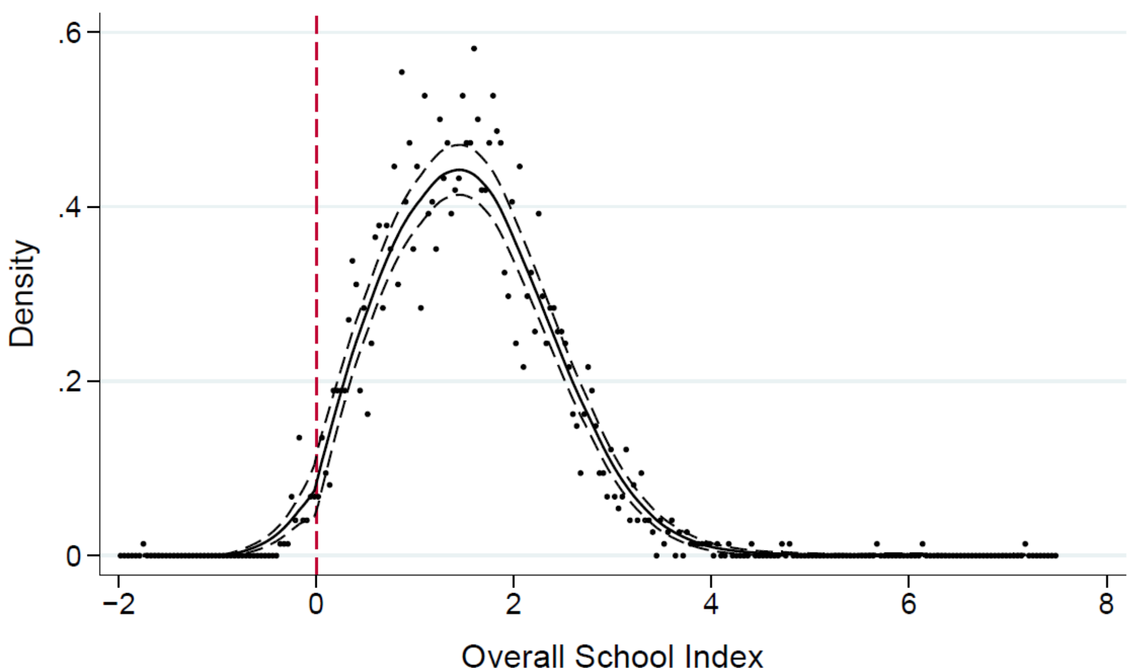
**Figure 1. Distributions of Running Variables: Priority Schools**

**A. 2012 Cohort**



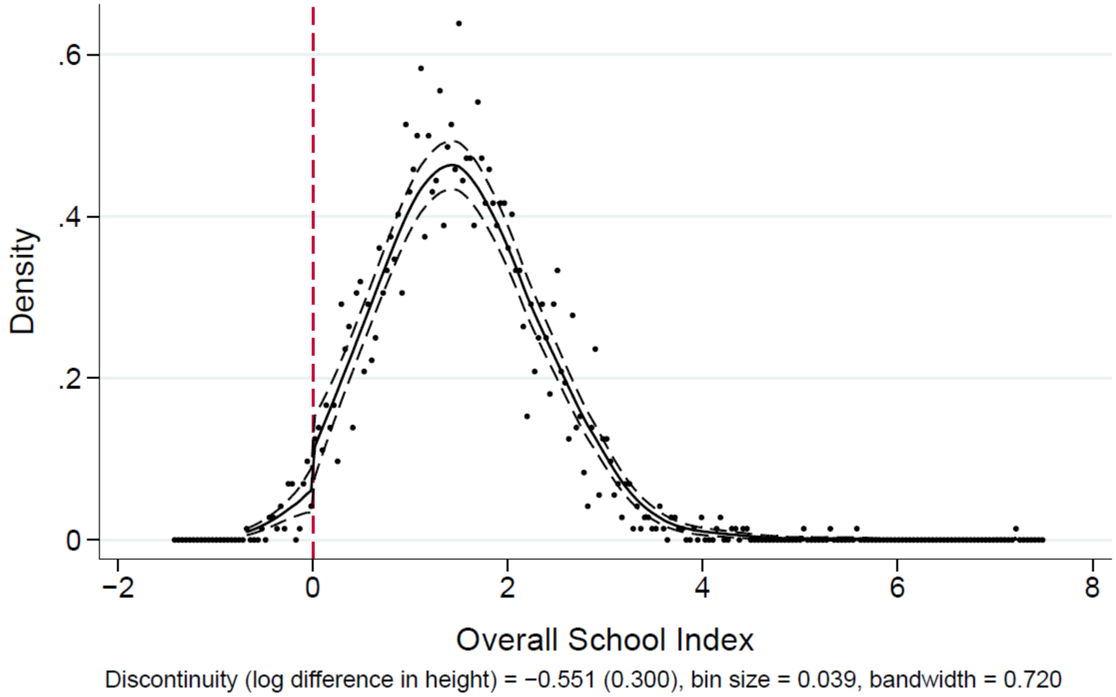

Discontinuity (log difference in height) = −0.083 (0.257), bin size = 0.038, bandwidth = 0.713

**B. 2013 Cohort**




Discontinuity (log difference in height) = −0.064 (0.294), bin size = 0.038, bandwidth = 0.721

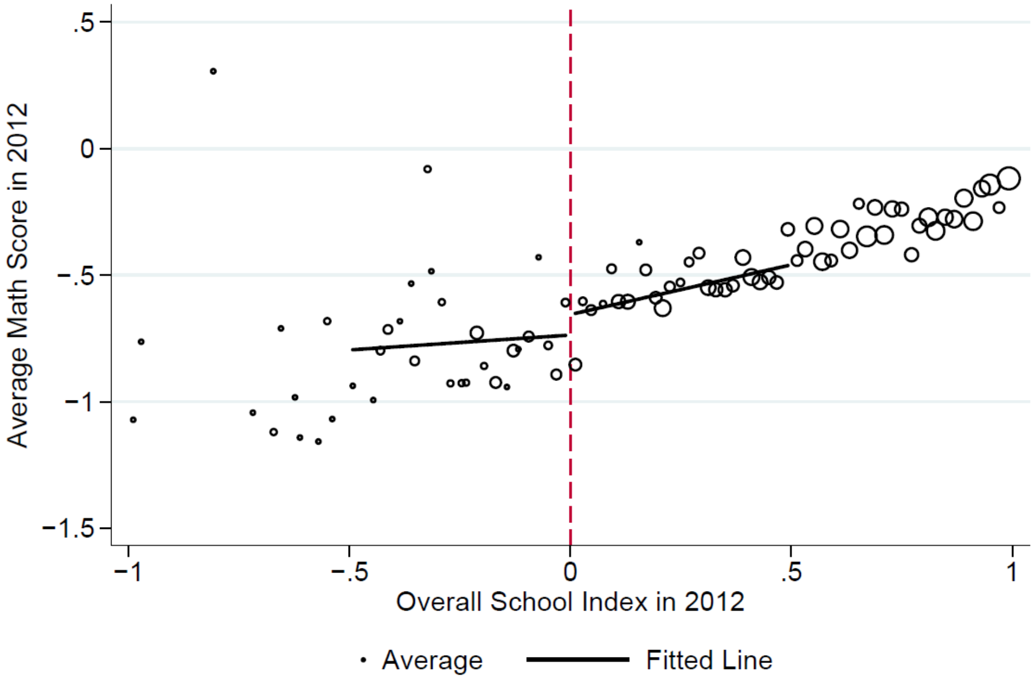
**C. 2014 Cohort**




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.

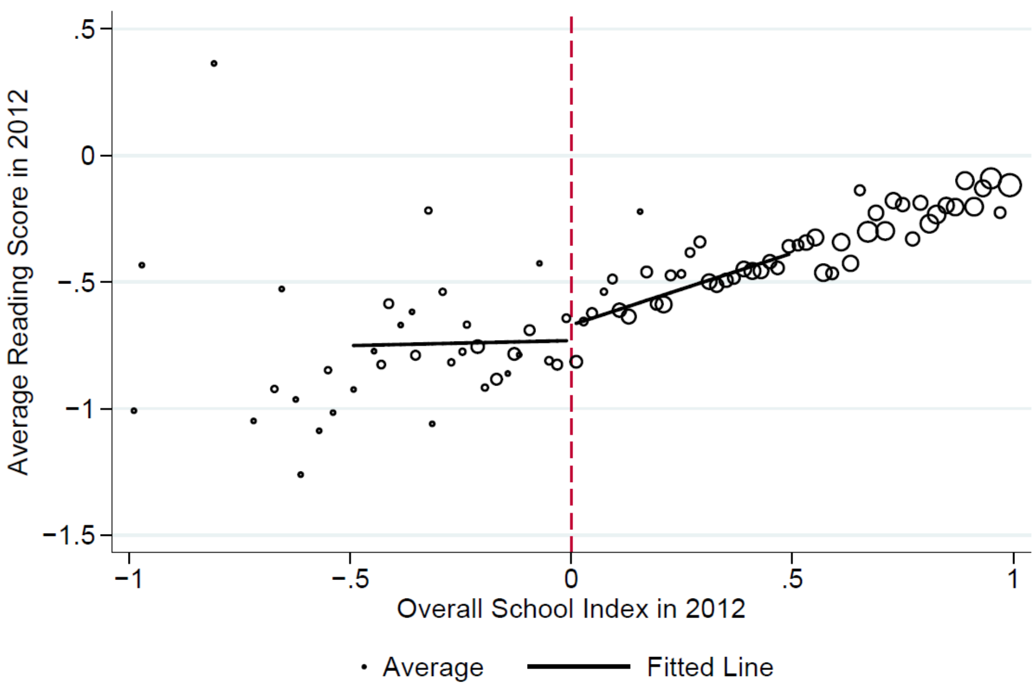
**Figure 2. Baseline Equivalence: Priority Schools, 2012 Cohort**

**A. Average Math Score**



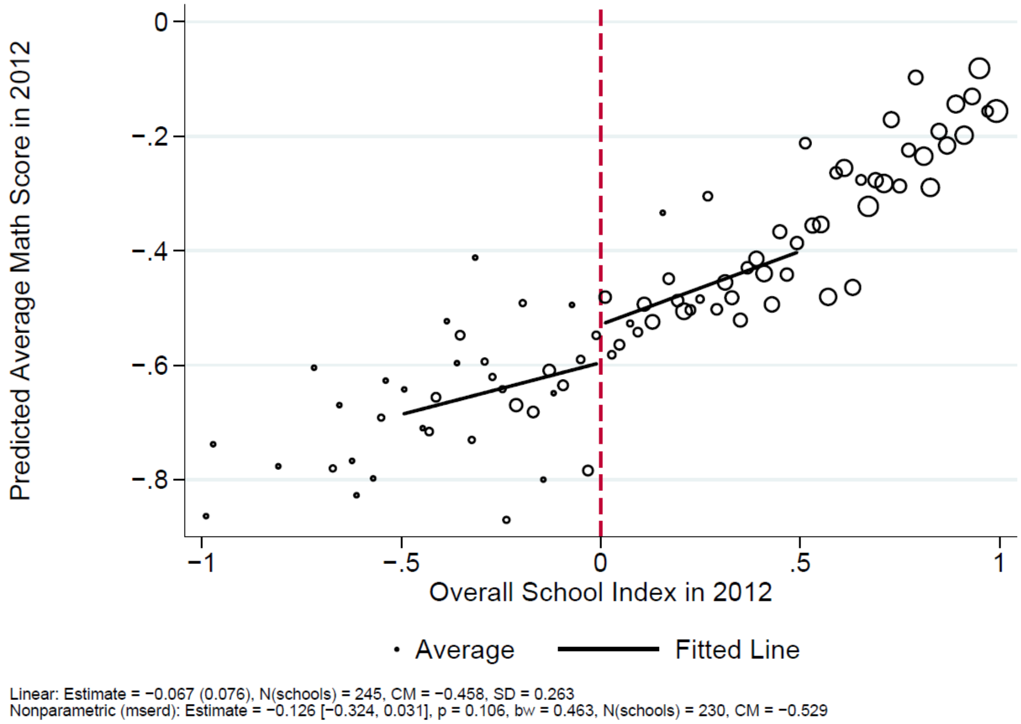

Linear: Estimate = −0.081 (0.083), N(schools) = 245, CM = −0.546, SD = 0.241
Nonparametric (mserd): Estimate = −0.111 [−0.331, 0.080], p = 0.230, bw = 0.483, N(schools) = 236, CM = −0.686

**B. Average Reading Score**

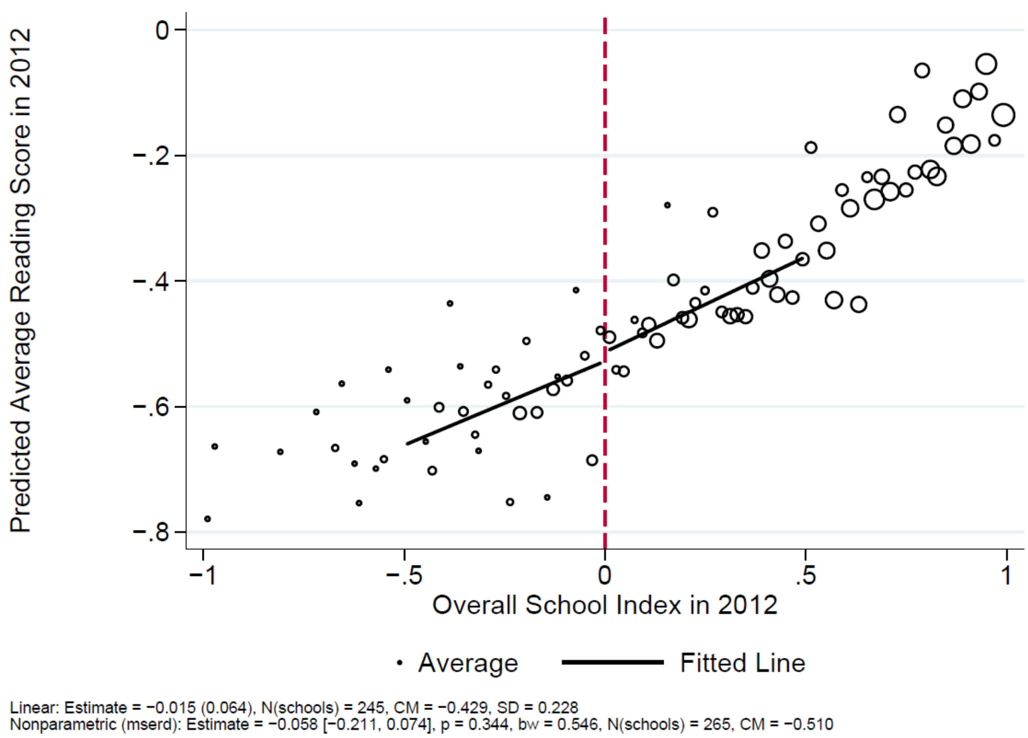



Linear: Estimate = −0.062 (0.080), N(schools) = 245, CM = −0.512, SD = 0.247
Nonparametric (mserd): Estimate = −0.058 [−0.260, 0.166], p = 0.666, bw = 0.409, N(schools) = 193, CM = −0.707

## C. Predicted Average Math Score



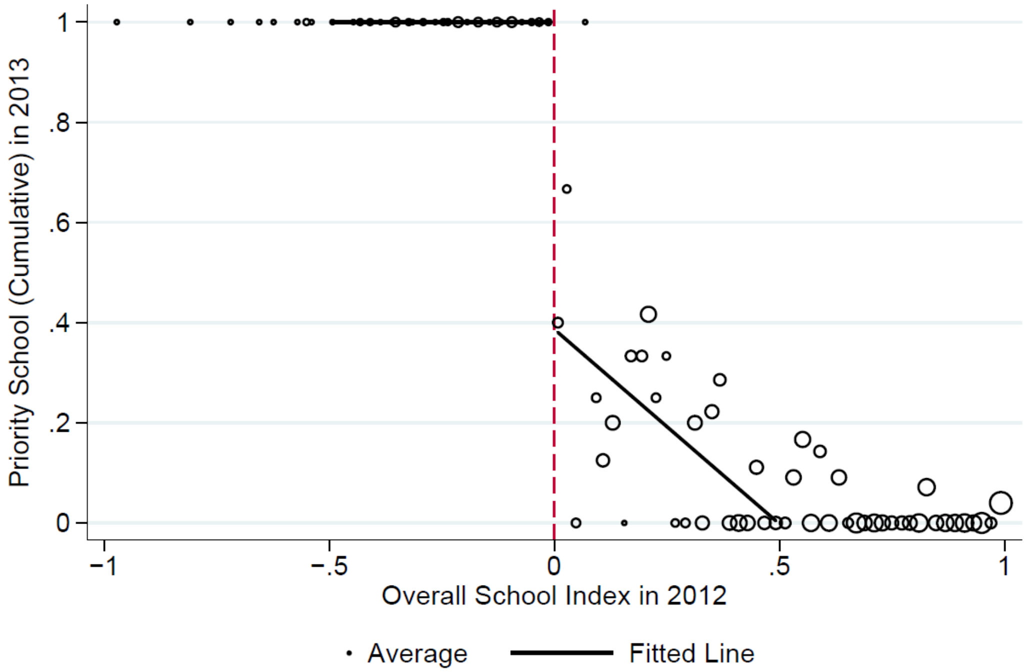

## D. Predicted Average Reading Score




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.

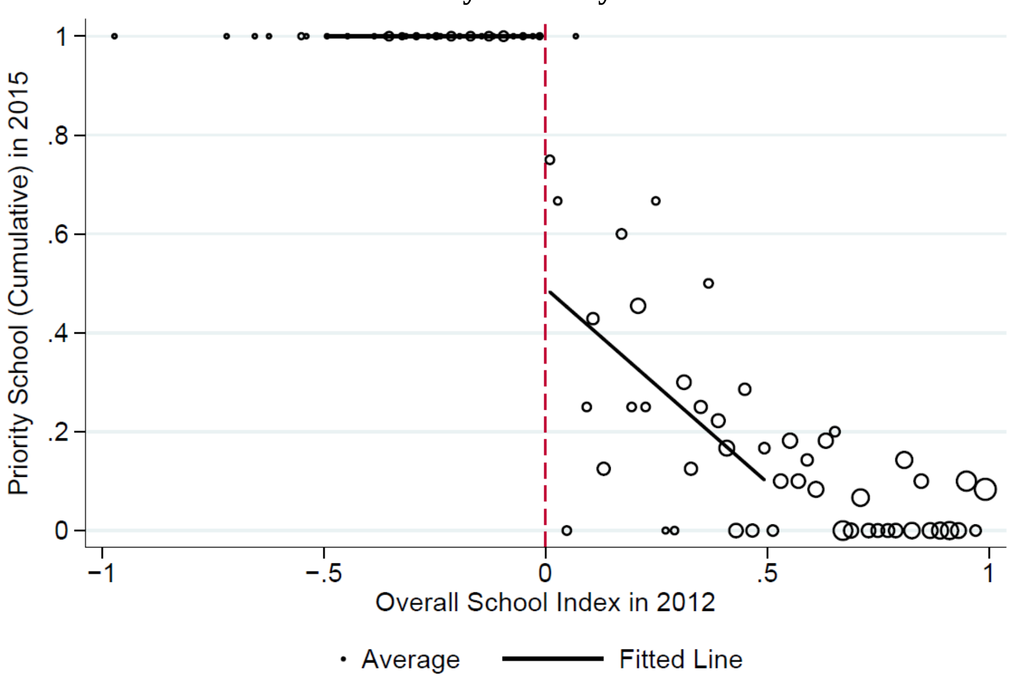
**Figure 3. Dynamic Treatment Crossover: Priority Schools**

**A. Priority Status by 2013**




Linear: Estimate = 0.613*** (0.078), N(schools) = 219, CM = 0.161, SD = 0.368
Nonparametric (mserd): Estimate = 0.662 [0.482, 0.774], p = 0.000, bw = 0.805, N(schools) = 393, CM = 0.338

**B. Priority Status by 2015**

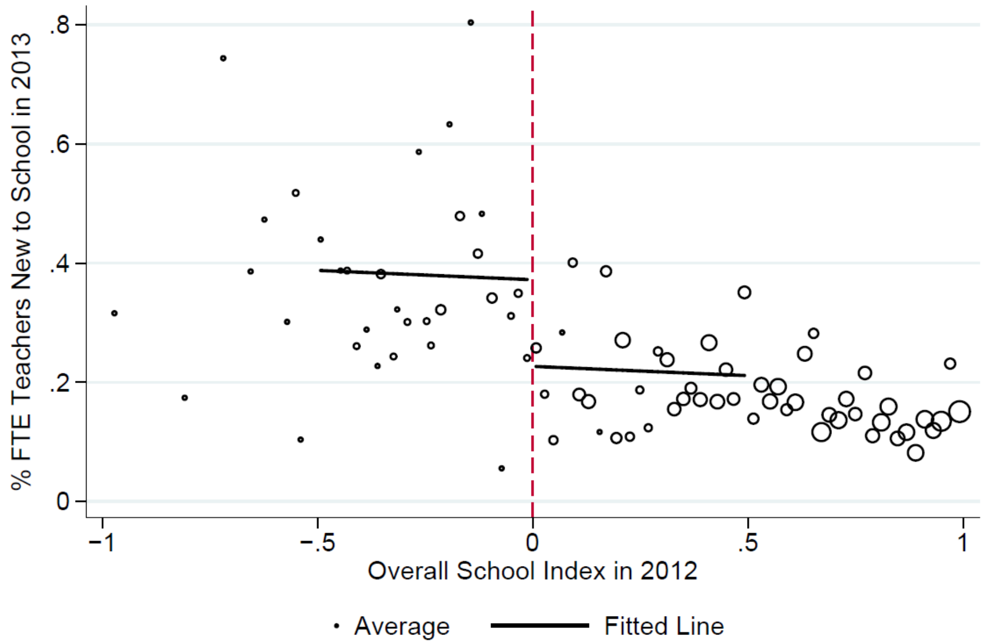



Linear: Estimate = 0.510*** (0.092), N(schools) = 188, CM = 0.262, SD = 0.441
Nonparametric (mserd): Estimate = 0.543 [0.343, 0.667], p = 0.000, bw = 0.838, N(schools) = 376, CM = 0.457

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.

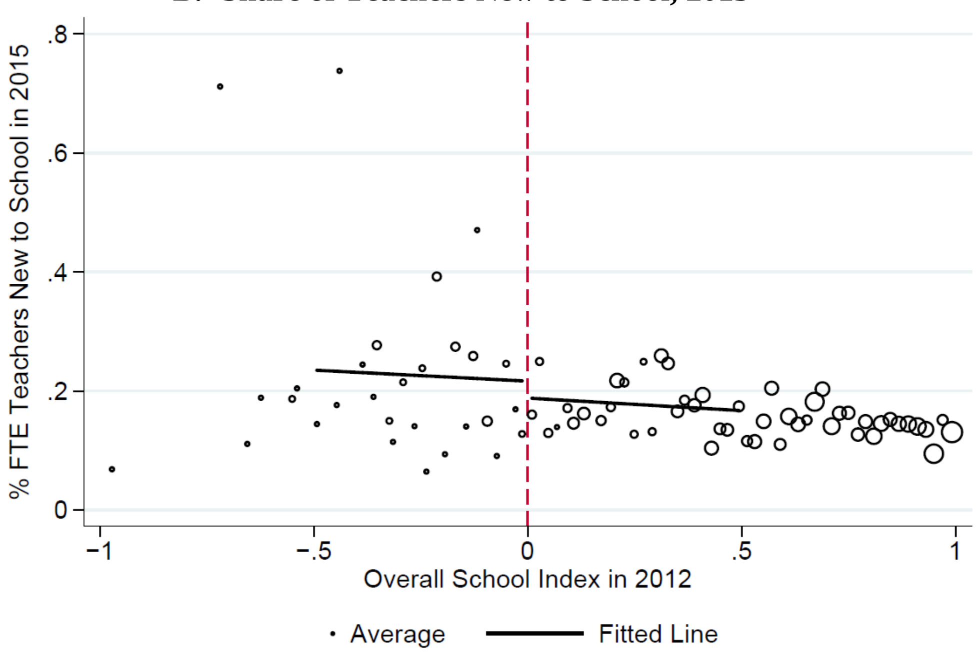
**Figure 4. Effect of Priority Designation on School Staffing, 2012 Cohort**

### A. Share of Teachers New to School, 2013




Linear: Estimate = 0.145** (0.066), N(schools) = 219, CM = 0.218, SD = 0.172
Nonparametric (mserd): Estimate = 0.053 [−0.143, 0.181], p = 0.817, bw = 0.246, N(schools) = 97, CM = 0.201

### B. Share of Teachers New to School, 2015

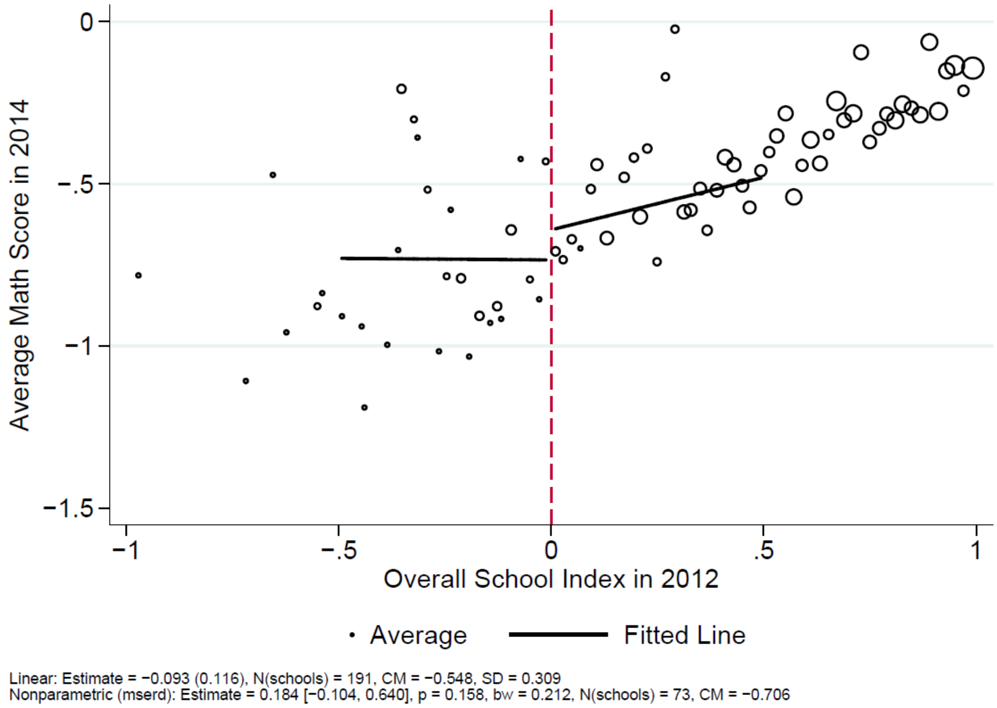



Linear: Estimate = 0.028 (0.057), N(schools) = 190, CM = 0.176, SD = 0.133
Nonparametric (mserd): Estimate = −0.012 [−0.118, 0.081], p = 0.712, bw = 0.326, N(schools) = 112, CM = 0.162

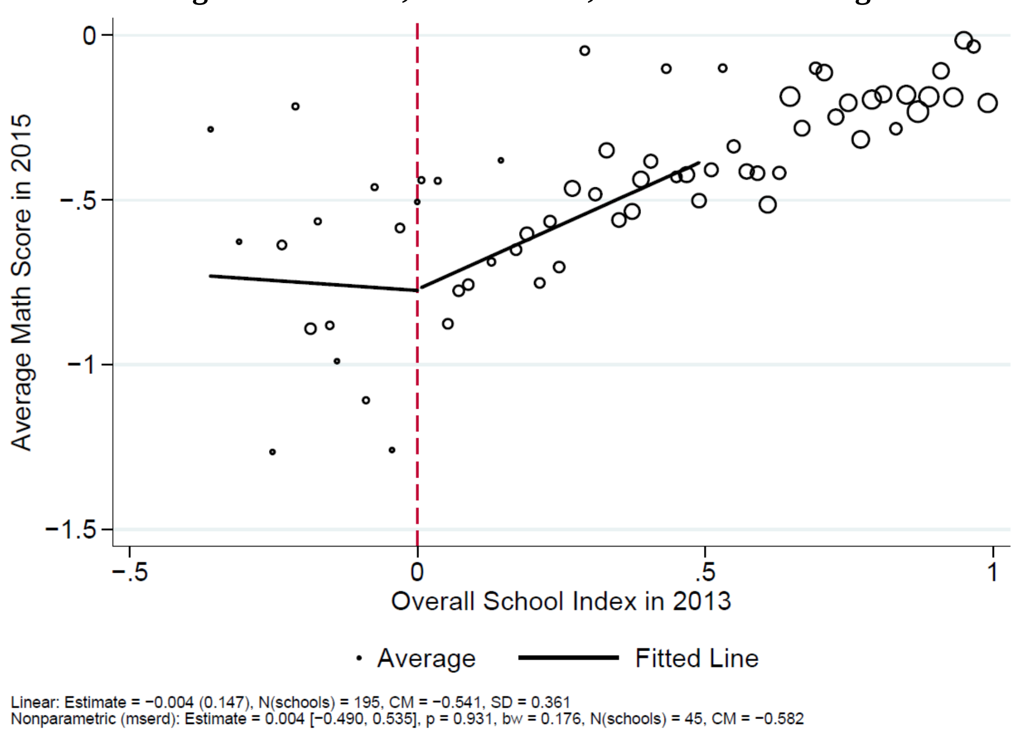
Notes: Parametric, linear specification is weighted by school enrollment. Nonparametric estimates are based on approach of Calonico et al. (2014a, 2014b, 2015). CM = control group mean; SD = standard deviation for control group; bw = bandwidth.

**Figure 5. Effect of Priority Status on Math Achievement**

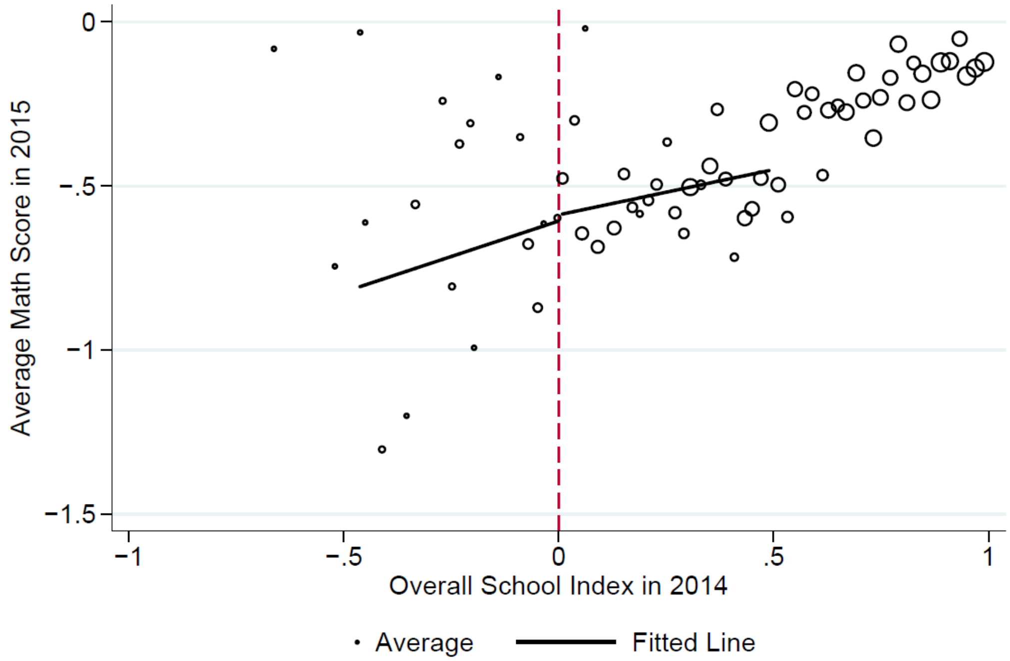
**A. Average Math Scores, 2012 Cohort, 2 Years After Designation**




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



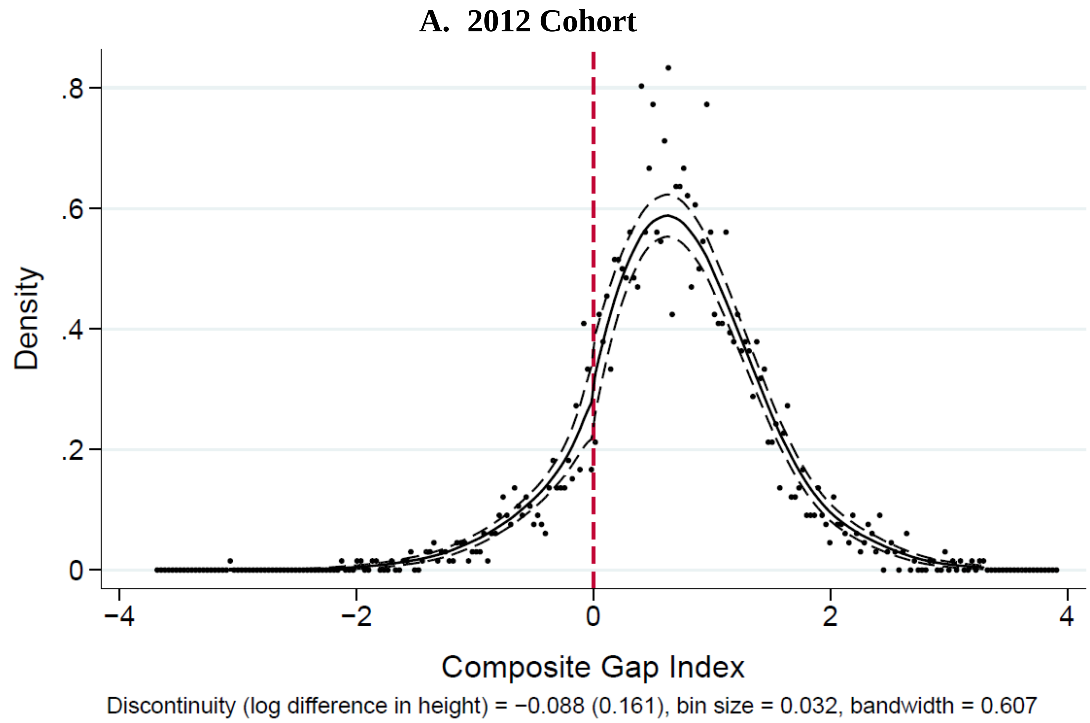
**C. Average Math Scores, 2014 Cohort, 1 Year After Designation**

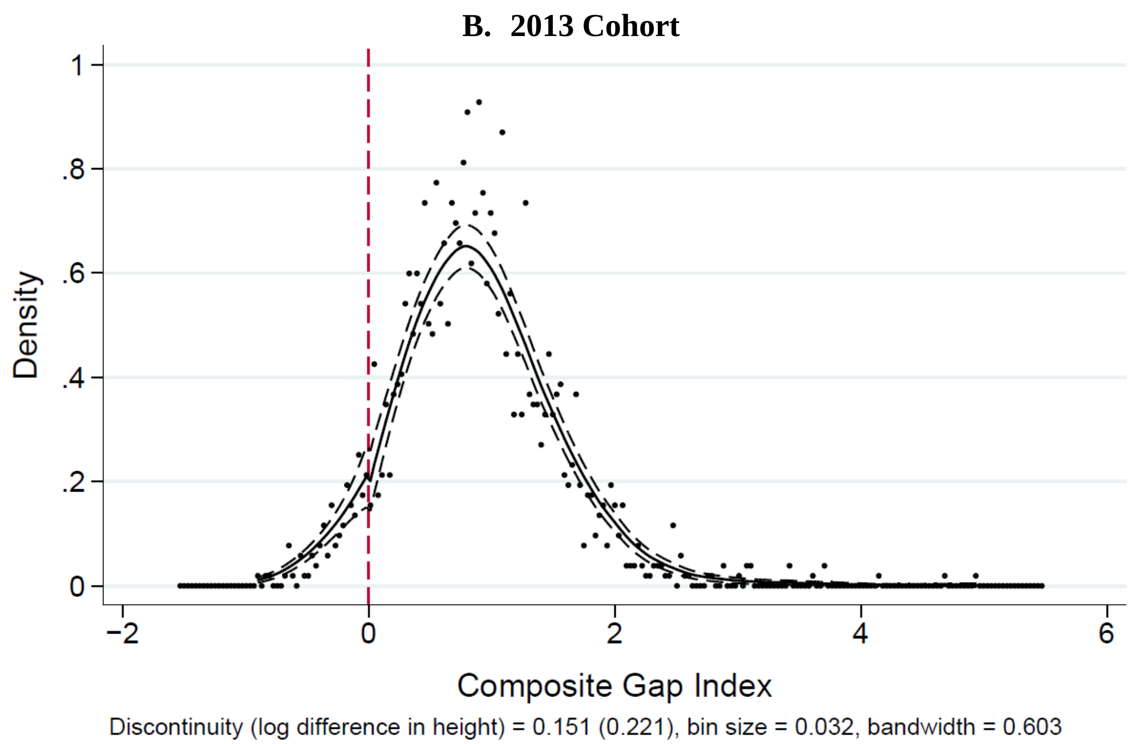



Linear: Estimate = −0.019 (0.138), N(schools) = 195, CM = −0.510, SD = 0.343
Nonparametric (mserd): Estimate = −0.305 [−0.809, 0.125], p = 0.151, bw = 0.183, N(schools) = 59, CM = −0.440

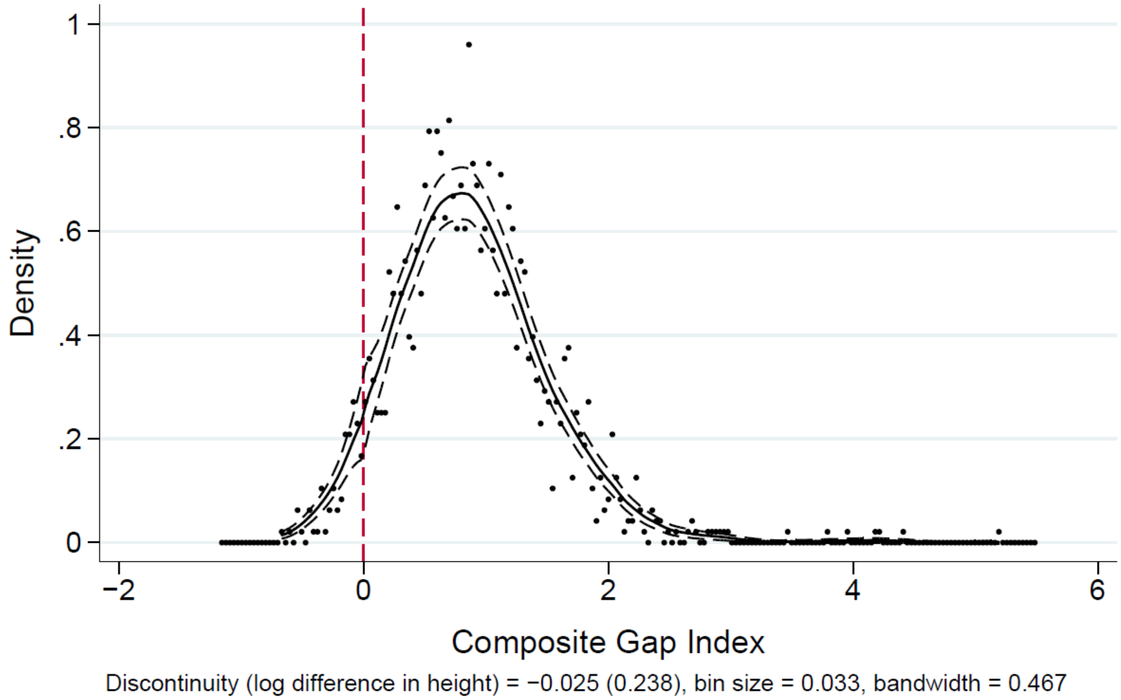
Notes: Parametric, linear specification is weighted by school enrollment. Nonparametric estimates are based on approach of Calonico et al. (2014a, 2014b, 2015). CM = control group mean; SD = standard deviation for control group; bw = bandwidth.

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



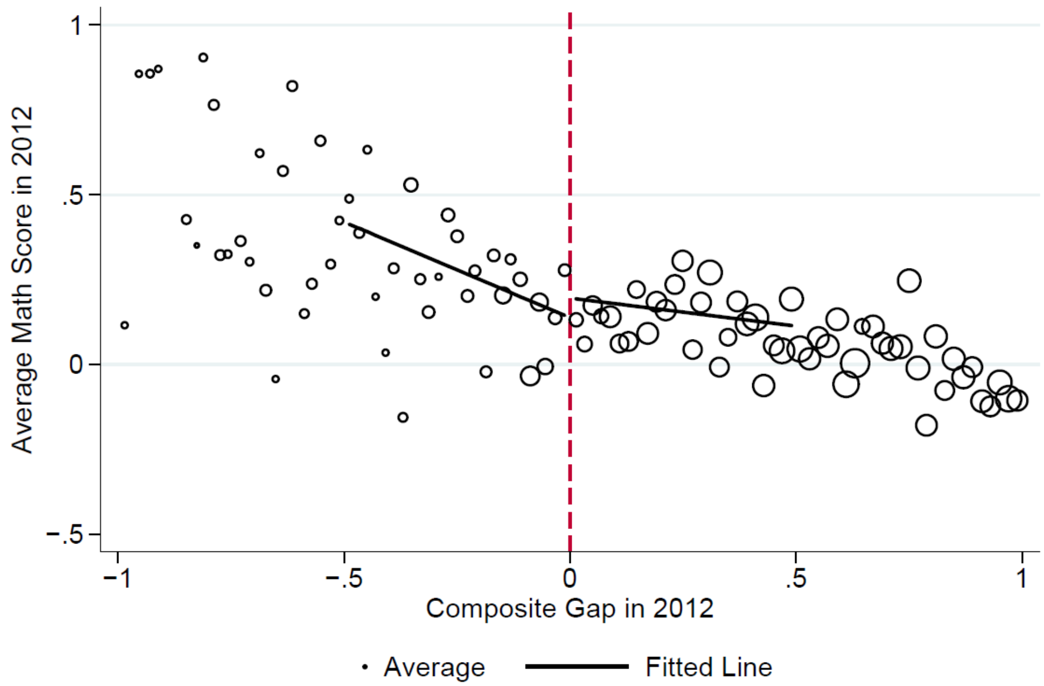
C.  2014 Cohort




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.

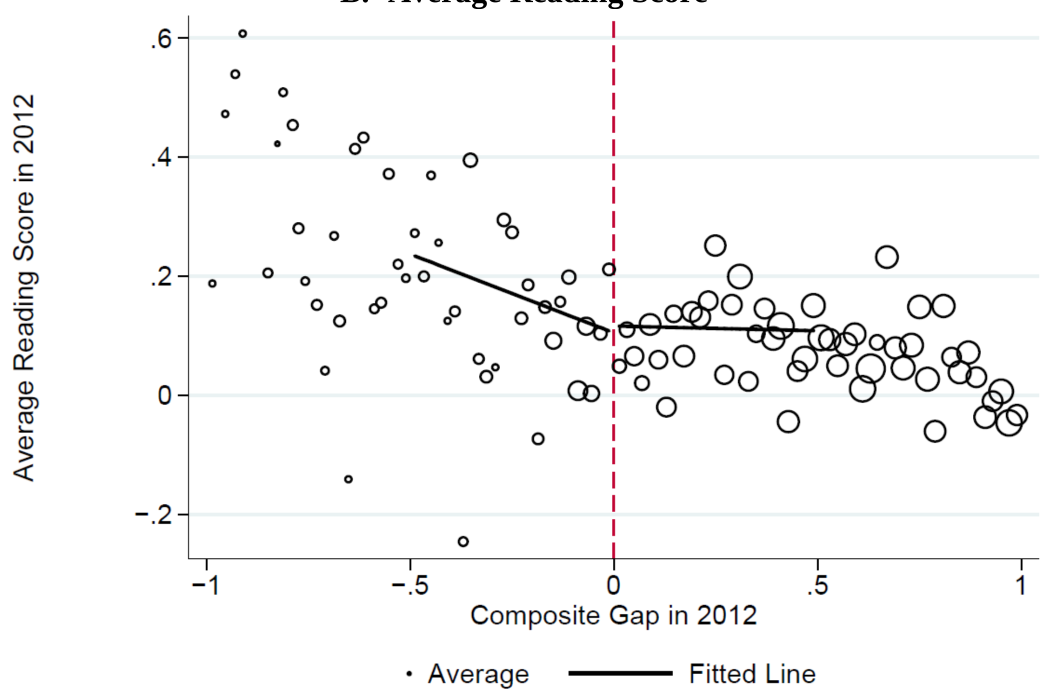
**Figure 7. Baseline Equivalence: Focus Schools, 2012 Cohort**

**A. Average Math Score**



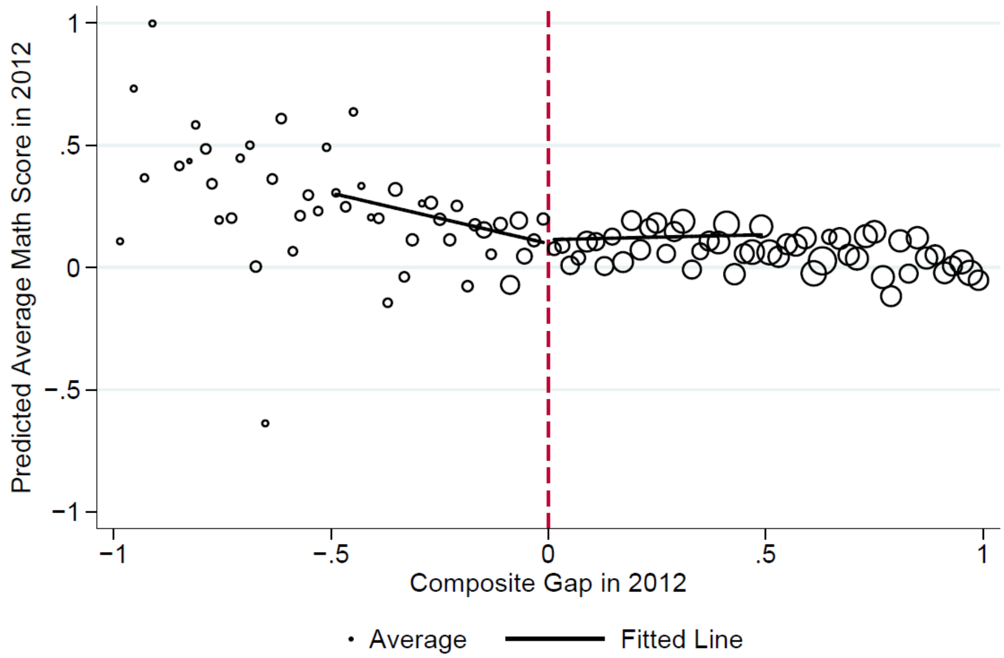

Linear: Estimate = −0.057 (0.074), N(schools) = 650, CM = 0.145, SD = 0.375
Nonparametric (mserd): Estimate = −0.070 [−0.213, 0.090], p = 0.426, bw = 0.623, N(schools) = 834, CM = 0.157

**B. Average Reading Score**



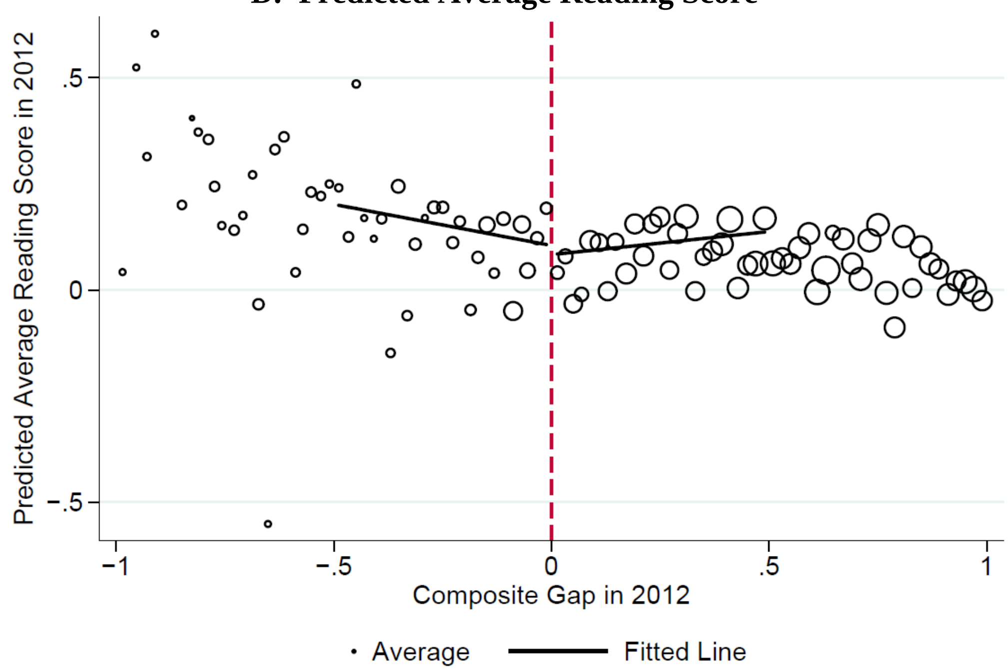

Linear: Estimate = −0.011 (0.057), N(schools) = 650, CM = 0.109, SD = 0.277
Nonparametric (mserd): Estimate = 0.006 [−0.113, 0.152], p = 0.770, bw = 0.481, N(schools) = 620, CM = 0.069

**C. Predicted Average Math Score**




Linear: Estimate = −0.016 (0.057), N(schools) = 650, CM = 0.122, SD = 0.293
Nonparametric (mserd): Estimate = 0.029 [−0.087, 0.180], p = 0.494, bw = 0.455, N(schools) = 581, CM = 0.052

**D. Predicted Average Reading Score**

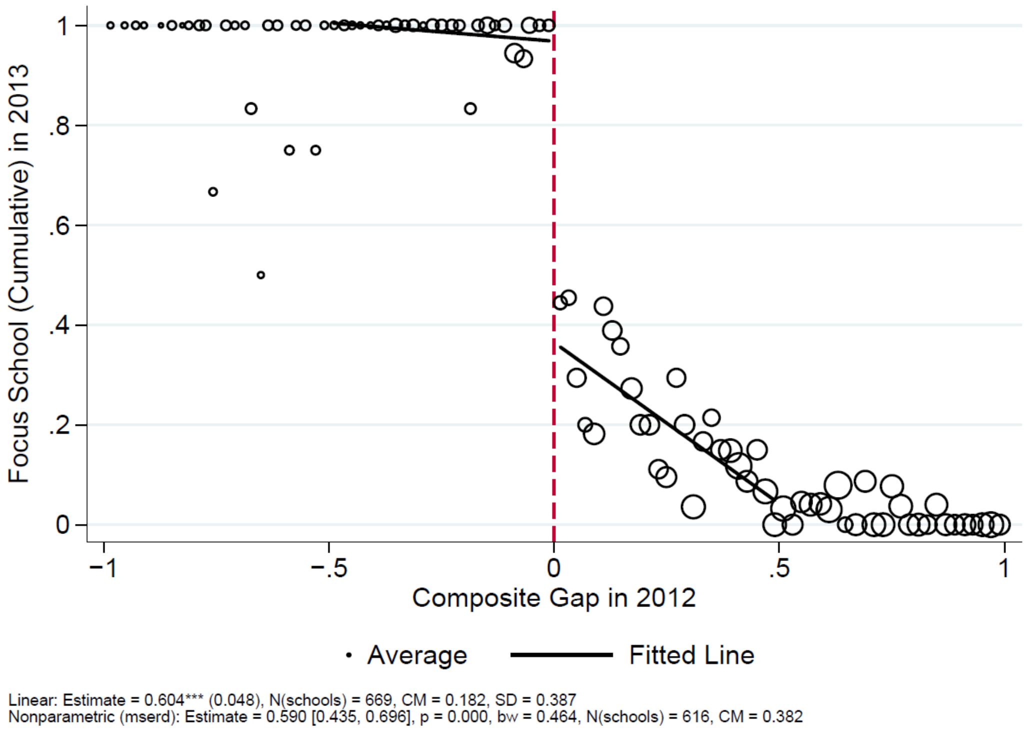



Linear: Estimate = 0.022 (0.051), N(schools) = 650, CM = 0.111, SD = 0.261
Nonparametric (mserd): Estimate = 0.057 [−0.040, 0.189], p = 0.201, bw = 0.445, N(schools) = 565, CM = 0.029

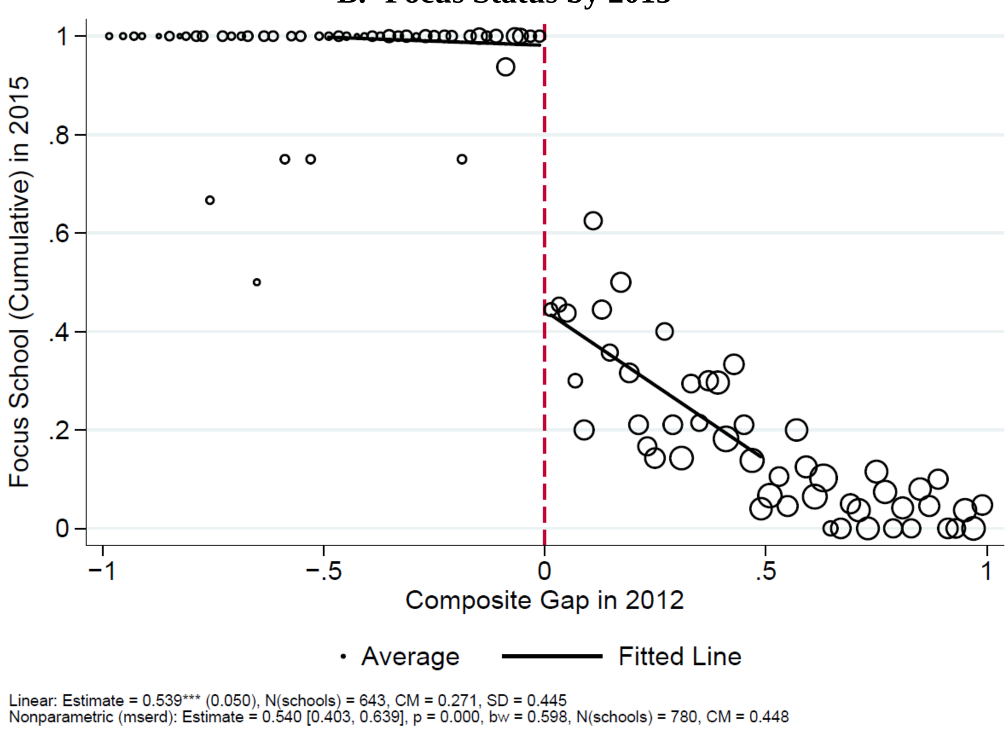
Notes: Parametric, linear specification is weighted by school enrollment. Nonparametric estimates are based on approach of Calonico et al. (2014a, 2014b, 2015). Predicted average scores come from regressions of test scores on the set of school characteristics reported in Table 1. CM = control group mean;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

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

**A. Focus Status by 2013**



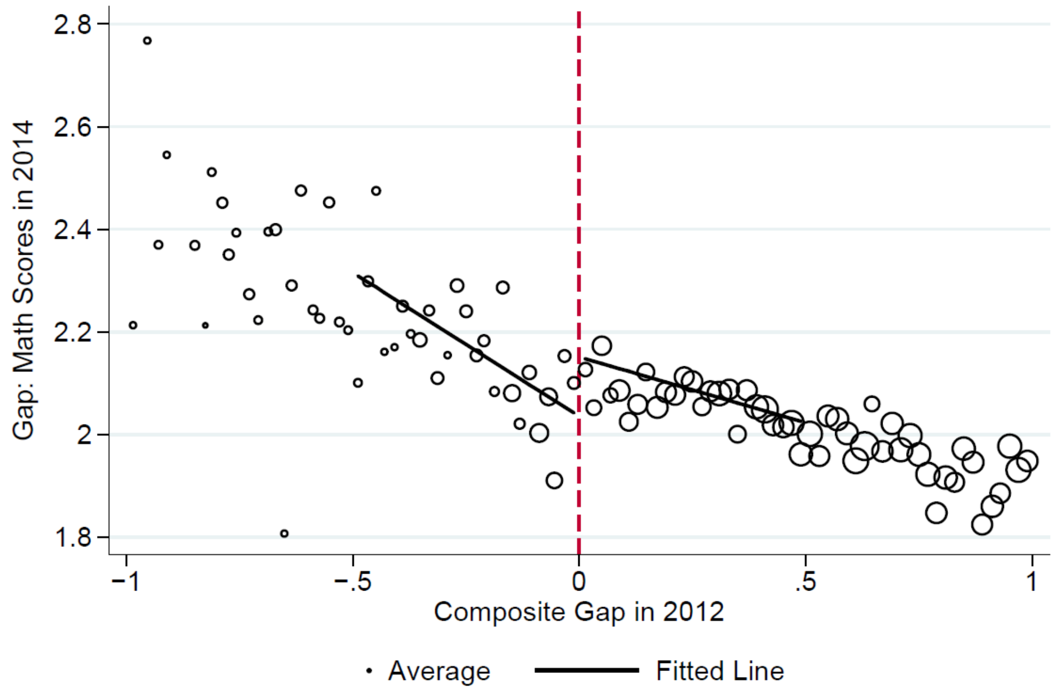

**B. Focus Status by 2015**




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.

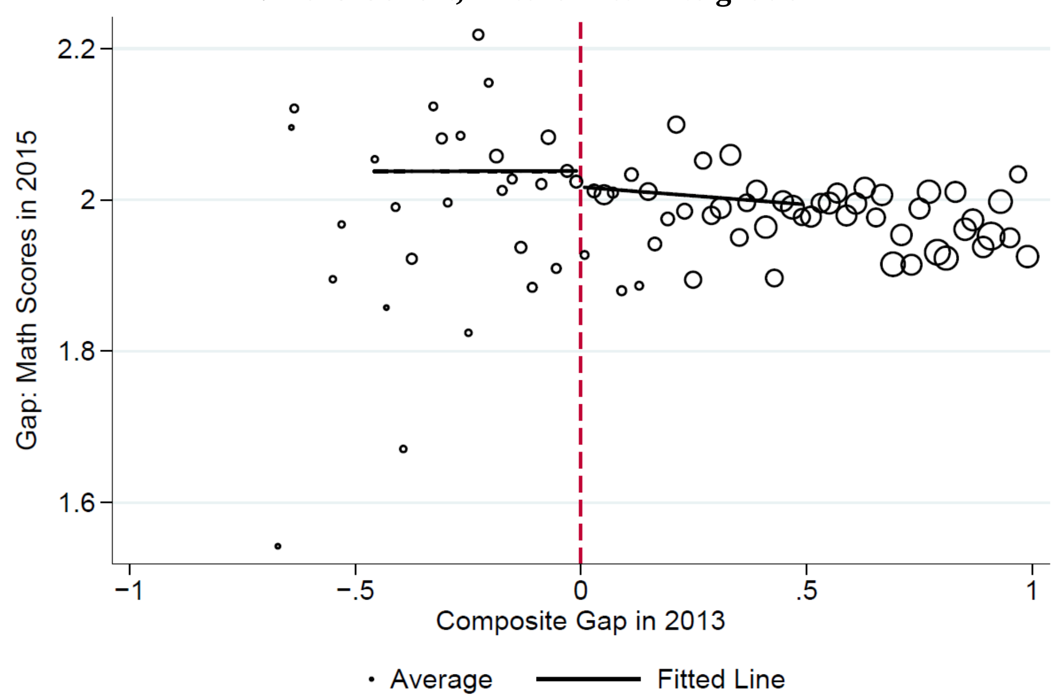
**Figure 9. Effect of Focus Status on Math Achievement Gaps**

**A. 2012 Cohort, 2 Years After Designation**



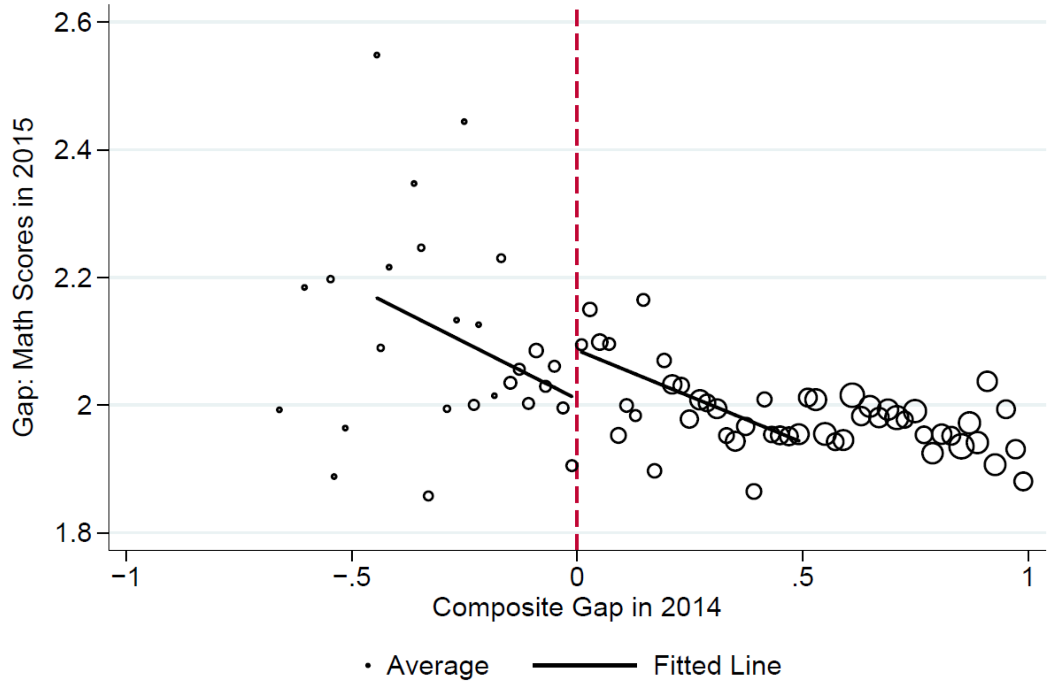

Linear: Estimate = −0.115*** (0.039), N(schools) = 634, CM = 2.077, SD = 0.195
Nonparametric (mserd): Estimate = −0.092 [−0.187, −0.003], p = 0.044, bw = 0.493, N(schools) = 627, CM = 2.106

**B. 2013 Cohort, 2Years After Designation**




Linear: Estimate = 0.021 (0.038), N(schools) = 408, CM = 2.002, SD = 0.162
Nonparametric (mserd): Estimate = 0.047 [−0.146, 0.265], p = 0.571, bw = 0.145, N(schools) = 87, CM = 1.989

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




Linear: Estimate = −0.077* (0.040), N(schools) = 373, CM = 2.008, SD = 0.152
Nonparametric (mserd): Estimate = −0.228 [−0.450, −0.064], p = 0.009, bw = 0.136, N(schools) = 99, CM = 2.160

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.

**Table 1. Descriptive Statistics, 2011-2012**

| Variable | All K-8 Schools Mean (1) | Priority K-8 Schools Mean (2) | Focus K-8 Schools Mean (3) |
|---|---|---|---|
| Priority | 0.041 | 1.000 | -- |
| Focus | 0.138 | -- | 1.000 |
| Reward | 0.116 | -- | -- |
| Average Math Score (Std.) | -0.020 | -0.773 | 0.357 |
| | (0.458) | (0.300) | (0.453) |
| Average Reading Score (Std.) | -0.017 | -0.746 | 0.218 |
| | (0.384) | (0.289) | (0.309) |
| Priority Running Variable | 1.441 | -0.351 | 1.660 |
| | (0.861) | (0.323) | (0.816) |
| Focus Running Variable | 0.709 | 1.418 | -0.475 |
| | (0.753) | (1.038) | (0.453) |
| *Student Characteristics* | | | |
| Share black | 0.195 | 0.776 | 0.136 |
| Share white | 0.691 | 0.142 | 0.708 |
| Share Hispanic | 0.070 | 0.065 | 0.055 |
| Share other | 0.044 | 0.017 | 0.100 |
| Share economically disadvantaged | 0.551 | 0.897 | 0.404 |
| Share LEP | 0.053 | 0.067 | 0.064 |
| *Staff Characteristics* | | | |
| Teacher-student ratio | 0.082 | 0.077 | 0.082 |
| | (0.020) | (0.017) | (0.018) |
| Aide-student ratio | 0.024 | 0.021 | 0.023 |
| | (0.016) | (0.012) | (0.015) |
| Share of teachers in first year at school | 0.127 | 0.239 | 0.121 |
| Average teacher experience (years) | 14.2 | 16.5 | 14.1 |
| | (3.3) | (3.5) | (3.0) |
| *School Characteristics* | | | |
| Total enrollment | 457 | 496 | 507 |
| | (195) | (219) | (203) |
| Elementary | 0.650 | 0.511 | 0.615 |
| Middle | 0.250 | 0.111 | 0.303 |
| Magnet | 0.139 | 0.112 | 0.116 |
| Charter | 0.075 | 0.056 | 0.063 |
| Urban | 0.189 | 0.820 | 0.178 |
| Suburban | 0.434 | 0.124 | 0.543 |
| Rural or Town | 0.376 | 0.056 | 0.280 |
| Closed in Summer of 2012 | 0.026 | 0.167 | 0.007 |
| N(schools) | 2,197 | 90 | 304 |
| N(students) | 995,653 | 43,924 | 152,571 |

Notes: Standard deviations of select continuous variables appear in parentheses below the means. Sample is limited to K-8, non-special-education schools open as of the spring of 2012. 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.

**Table 2. Effects of Priority Designation on Share of Teachers New to the School**

**A. 2012 Cohort**

| Outcomes | (1) | Parametric (2) | (3) | Nonparametric (4) |
|---|---|---|---|---|
| | *Outcome Year = 2013* | | | |
| ITT: impact of Priority designation in 2012 | 0.117* | 0.145** | 0.084* | 0.053 |
| | (0.062) | (0.066) | (0.050) | [-0.143,0.181] |
| | *Outcome Year = 2014* | | | |
| ITT: impact of Priority designation in 2012 | 0.035 | 0.016 | -0.018 | 0.012 |
| | (0.049) | (0.038) | (0.039) | [-0.097,0.142] |
| TOT: impact of spending one more year under Priority designation | 0.022 | 0.010 | -0.011 | 0.007 |
| | (0.030) | (0.023) | (0.022) | [-0.059,0.087] |
| | *Outcome Year = 2015* | | | |
| ITT: impact of Priority designation in 2012 | 0.013 | 0.028 | 0.002 | -0.012 |
| | (0.062) | (0.057) | (0.048) | [-0.118,0.081] |
| TOT: impact of spending one more year under Priority designation | 0.006 | 0.013 | 0.001 | -0.006 |
| | (0.029) | (0.026) | (0.021) | [-0.055,0.038] |

**B. 2013 Cohort**

| Outcomes | (5) | Parametric (6) | (7) | Nonparametric (8) |
|---|---|---|---|---|
| | *Outcome Year = 2014* | | | |
| ITT: impact of Priority designation in 2013 | 0.015 | 0.002 | -0.010 | 0.084 |
| | (0.105) | (0.055) | (0.057) | [-0.034,0.274] |
| | *Outcome Year = 2015* | | | |
| ITT: impact of Priority designation in 2013 | -0.089 | -0.040 | 0.008 | 0.012 |
| | (0.099) | (0.053) | (0.067) | [-0.126,0.199] |
| TOT: impact of spending one more year under Priority designation | -0.061 | -0.026 | 0.005 | 0.006 |
| | (0.067) | (0.034) | (0.041) | [-0.069,0.108] |

**C. 2014 Cohort**

| Outcomes | (9) | Parametric (10) | (11) | Nonparametric (12) |
|---|---|---|---|---|
| | *Outcome Year = 2015* | | | |
| ITT: impact of Priority designation in 2014 | -0.061 | -0.076* | -0.070 | -0.085* |
| | (0.046) | (0.040) | (0.047) | [-0.195,0.004] |
| Data window (left, right) | -1 to 1.5 | -0.5 to 0.5 | -0.5 to 0.5 | data-driven bandwidth |
| Include covariates | yes | no | yes | no |
| Polynomial of running variable | cubic | linear | linear | -- |

Notes: The analytic sample for the 2012 cohort includes all students enrolled in operational K-8, non-special education, non-charter schools identified in the spring of 2012. For subsequent cohorts, we exclude schools identified as Priority in prior years since they leave the risk set for treatment after initial designation for intervention. N(schools) for the cohort of 2012 ranges from 925 in the widest data window to 219 in the narrowest. Corresponding sample sizes for the 2013 and 2014 cohorts are 948 to 204 and 941 to 200, respectively. Across all cohorts and outcome years the data-driven bandwidth selected by the nonparametric approach ranges from 0.14 to 0.27. Parametric models are estimated on school-level data and are weighted by total enrollment. For a full list of covariates, please consult the text. Standard errors clustered by school appear in parentheses. TOT coefficients should be interpreted as the effect of one additional year under Priority status on the outcome of interest. Nonparametric estimates are calculated using a triangular kernel and data-driven bandwidth selection procedure described in Calonico, Cattaneo, and Titiunik (2015). Robust 95-percent confidence intervals are shown in brackets. Robust standard errors appear in parentheses. *** p<0.01, ** p<0.05, * p<0.1. ITT = intent to treat; TOT = treatment on the treated.

**Table 3. Effects of Priority Designation on School Composition**

| | ITT Estimates | | | TOT Estimates | |
|---|---|---|---|---|---|
| | 1 year after | 2 years after | 3 years after | 2 years after | 3 years after |
| Outcome | (1) | (2) | (3) | (4) | (5) |
| *2012 Cohort (N = 219 schools)* | | | | | |
| Log of total enrollment | 0.170*** | 0.158** | 0.169** | 0.098** | 0.079** |
| | (0.059) | (0.077) | (0.081) | (0.043) | (0.035) |
| Share black | -0.002 | -0.012 | -0.015 | -0.007 | -0.007 |
| | (0.007) | (0.012) | (0.016) | (0.007) | (0.007) |
| Share economically disadvantaged | 0.003 | -0.003 | 0.003 | -0.002 | 0.001 |
| | (0.014) | (0.018) | (0.020) | (0.010) | (0.008) |
| *2013 Cohort (N = 204 schools)* | | | | | |
| Log of total enrollment | -0.060 | -0.110 | | -0.073 | |
| | (0.084) | (0.086) | | (0.055) | |
| Share black | -0.024 | -0.053** | | -0.036** | |
| | (0.017) | (0.022) | | (0.014) | |
| Share economically disadvantaged | -0.047** | -0.011 | | -0.007 | |
| | (0.022) | (0.022) | | (0.013) | |
| *2014 Cohort (N = 200 schools)* | | | | | |
| Log of total enrollment | 0.021 | | | | |
| | (0.086) | | | | |
| Share black | -0.008 | | | | |
| | (0.019) | | | | |
| Share economically disadvantaged | -0.012 | | | | |
| | (0.019) | | | | |

Notes: The analytic sample for the 2012 cohort includes all students enrolled in operational K-8, non-special education, non-charter schools identified in the spring of 2012. For subsequent cohorts, we exclude schools identified as Priority in prior years since they leave the risk set for 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. TOT coefficients should be interpreted as the effect of one additional year under Priority status on the outcome of interest. Robust standard errors appear in parentheses. *** p<0.01, ** p<0.05, * p<0.1. ITT = intent to treat; TOT = treatment on the treated.

**Table 4. Effects of Priority Designation on Math Achievement**

| Outcome | ITT Estimates<br>1 year after<br>(1) | 2 years after<br>(2) | 3 years after<br>(3) | TOT Estimates<br>2 years after<br>(4) | 3 years after<br>(5) |
|---|---|---|---|---|---|
| *2012 Cohort* | | *2012 Cohort (N = 215 schools)* | | | |
| Average math score (std) | | -0.031 | 0.104 | -0.019 | 0.049 |
| | | (0.061) | (0.069) | (0.035) | (0.030) |
| Math score, 10th percentile | | -0.031 | 0.073 | -0.019 | 0.034 |
| | | (0.040) | (0.079) | (0.023) | (0.035) |
| Math score, 25th percentile | | -0.016 | 0.080 | -0.010 | 0.038 |
| | | (0.048) | (0.068) | (0.027) | (0.030) |
| Math score, 50th percentile | | -0.045 | 0.104 | -0.028 | 0.049* |
| | | (0.062) | (0.067) | (0.035) | (0.029) |
| Math core, 75th percentile | | -0.036 | 0.125 | -0.022 | 0.058* |
| | | (0.083) | (0.081) | (0.048) | (0.034) |
| Math score, 90th percentile | | 0.004 | 0.126 | 0.002 | 0.059 |
| | | (0.113) | (0.092) | (0.065) | (0.039) |
| | | *2013 Cohort (N = 199 schools)* | | | |
| Average math score (std) | | -0.039 | | -0.026 | |
| | | (0.100) | | (0.062) | |
| Math score, 10th percentile | | -0.045 | | -0.030 | |
| | | (0.113) | | (0.070) | |
| Math score, 25th percentile | | -0.024 | | -0.016 | |
| | | (0.100) | | (0.062) | |
| Math score, 50th percentile | | -0.050 | | -0.033 | |
| | | (0.098) | | (0.061) | |
| Math core, 75th percentile | | -0.073 | | -0.049 | |
| | | (0.106) | | (0.065) | |
| Math score, 90th percentile | | -0.016 | | -0.011 | |
| | | (0.135) | | (0.084) | |
| | | *2014 Cohort (N = 193 schools)* | | | |
| Average math score (std) | -0.091 | | | | |
| | (0.081) | | | | |
| Math score, 10th percentile | -0.071 | | | | |
| | (0.083) | | | | |
| Math score, 25th percentile | -0.045 | | | | |
| | (0.077) | | | | |
| Math score, 50th percentile | -0.092 | | | | |
| | (0.078) | | | | |
| Math core, 75th percentile | -0.073 | | | | |
| | (0.094) | | | | |
| Math score, 90th percentile | -0.165 | | | | |
| | (0.101) | | | | |

Notes: The analytic sample for the 2012 cohort includes all students enrolled in operational K-8, non-special education, non-charter schools identified in the spring of 2012. For subsequent cohorts, we exclude schools identified as Priority in prior years since they leave the risk set for 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. TOT coefficients should be interpreted as the effect of one additional year under Priority status on the outcome of interest. Robust standard errors appear in parentheses. *** p<0.01, ** p<0.05, * p<0.1. ITT = intent to treat; TOT = treatment on the treated.

**Table 5. Effects of Focus Designation on Share of Teachers New to the School**

**A. 2012 Cohort**

| Outcomes | (1) | Parametric (2) | (3) | Nonparametric (4) |
|---|---|---|---|---|
| *Outcome Year = 2013* | | | | |
| ITT: impact of Focus designation in 2012 | 0.035 | 0.009 | 0.019 | 0.022 |
| | (0.030) | (0.020) | (0.019) | [-0.033,0.084] |
| *Outcome Year = 2014* | | | | |
| ITT: impact of Focus designation in 2012 | -0.016 | -0.002 | 0.001 | -0.004 |
| | (0.022) | (0.017) | (0.017) | [-0.052,0.030] |
| TOT: impact of spending one more year under Focus designatio | -0.011 | -0.001 | 0.001 | 0.003 |
| | (0.014) | (0.011) | (0.010) | [-0.030,0.019] |
| *Outcome Year = 2015* | | | | |
| ITT: impact of Focus designation in 2012 | -0.057*** | -0.035** | -0.042** | -0.058*** |
| | (0.019) | (0.017) | (0.016) | [-0.109,-0.024] |
| TOT: impact of spending one more year under Focus designatio | -0.027*** | -0.016** | -0.019*** | -0.021*** |
| | (0.009) | (0.008) | (0.007) | [-0.047,-0.009] |

**B. 2013 Cohort**

| Outcomes | (5) | Parametric (6) | (7) | Nonparametric (8) |
|---|---|---|---|---|
| *Outcome Year = 2014* | | | | |
| ITT: impact of Focus designation in 2013 | -0.015 | 0.014 | 0.010 | -0.052 |
| | (0.023) | (0.028) | (0.025) | [-0.154,0.028] |
| *Outcome Year = 2015* | | | | |
| ITT: impact of Focus designation in 2013 | 0.021 | 0.051*** | 0.060*** | 0.030 |
| | (0.023) | (0.018) | (0.016) | [-0.055,0.102] |
| TOT: impact of spending one more year under Focus designatio | 0.015 | 0.035*** | 0.043*** | 0.011 |
| | (0.016) | (0.013) | (0.011) | [-0.042,0.062] |

**C. 2014 Cohort**

| Outcomes | (9) | Parametric (10) | (11) | Nonparametric (12) |
|---|---|---|---|---|
| *Outcome Year = 2015* | | | | |
| ITT: impact of Focus designation in 2014 | 0.021 | 0.019 | 0.017 | 0.018 |
| | (0.028) | (0.020) | (0.021) | [-0.055,0.092] |
| Data window (left, right) | -1 to 1.5 | -0.5 to 0.5 | -0.5 to 0.5 | data-driven bandwidth |
| Include covariates | yes | no | yes | no |
| Polynomial of running variable | cubic | linear | linear | -- |

Notes: The analytic sample for the 2012 cohort includes all students enrolled in operational K-8, non-special education, non-charter schools identified in the spring of 2012. 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. N(schools) for the cohort of 2012 ranges from 1,735 in the widest data window to 674 in the narrowest. Corresponding sample sizes for the 2013 and 2014 cohorts are 1,379 to 434 and 1,247 to 385, respectively. Across all cohorts and outcome years the data-driven bandwidth selected by the nonparametric approach ranges from 0.15 to 0.55. Parametric models are estimated on school-level data and are weighted by total enrollment. For a full list of covariates, please consult the text. Standard errors clustered by school appear in parentheses. TOT coefficients should be interpreted as the effect of one additional year under Focus status on the outcome of interest. Nonparametric estimates are calculated using a triangular kernel and data-driven bandwidth selection procedure described in Calonico, Cattaneo, and Titiunik (2015). Robust 95-percent confidence intervals are shown in brackets. Robust standard errors appear in parentheses. *** p<0.01, ** p<0.05, * p<0.1. ITT = intent to treat; TOT = treatment on the treated.

**Table 6. Effects of Focus Designation on School Composition**

| | ITT Estimates | | | TOT Estimates | |
|---|---|---|---|---|---|
| | 1 year after | 2 years after | 3 years after | 2 years after | 3 years after |
| Outcome | (1) | (2) | (3) | (4) | (5) |
| | | | *2012 Cohort (N = 674 schools)* | | |
| Log of total enrollment | 0.013 | 0.039 | 0.038 | 0.024 | 0.017 |
| | (0.028) | (0.031) | (0.033) | (0.018) | (0.014) |
| Share black | -0.001 | 0.001 | 0.003 | 0.000 | 0.001 |
| | (0.003) | (0.004) | (0.005) | (0.002) | (0.002) |
| Share economically disadvantaged | 0.011 | 0.006 | 0.006 | 0.004 | 0.003 |
| | (0.008) | (0.009) | (0.011) | (0.006) | (0.005) |
| | | | *2013 Cohort (N = 434 schools)* | | |
| Log of total enrollment | 0.004 | -0.010 | | -0.007 | |
| | (0.033) | (0.038) | | (0.026) | |
| Share black | 0.004 | 0.002 | | 0.002 | |
| | (0.005) | (0.006) | | (0.004) | |
| Share economically disadvantaged | -0.000 | -0.006 | | -0.004 | |
| | (0.014) | (0.014) | | (0.010) | |
| | | | *2014 Cohort (N = 385 schools)* | | |
| Log of total enrollment | -0.001 | | | | |
| | (0.032) | | | | |
| Share black | -0.001 | | | | |
| | (0.008) | | | | |
| Share economically disadvantaged | 0.026* | | | | |
| | (0.016) | | | | |

Notes: The analytic sample for the 2012 cohort includes all students enrolled in operational K-8, non-special education, non-charter schools identified in the spring of 2012. 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. TOT coefficients should be interpreted as the effect of one additional year under Focus status on the outcome of interest. Robust standard errors appear in parentheses. *** p<0.01, ** p<0.05, * p<0.1. ITT = intent to treat; TOT = treatment on the treated.

**Table 7. Effects of Focus Designation on Math Achievement**

| | ITT Estimates | | | TOT Estimates | |
|---|---|---|---|---|---|
| | 1 year after | 2 years after | 3 years after | 2 years after | 3 years after |
| Outcome | (1) | (2) | (3) | (4) | (5) |
| | | *2012 Cohort (N = 650 schools)* | | | |
| Average math score (std) | | -0.003 | -0.059* | -0.002 | -0.026* |
| | | (0.030) | (0.032) | (0.018) | (0.014) |
| Math gap score | | -0.080*** | -0.019 | -0.049*** | -0.008 |
| | | (0.029) | (0.027) | (0.017) | (0.012) |
| Math score, 10th percentile | | 0.020 | -0.060* | 0.012 | -0.027* |
| | | (0.026) | (0.034) | (0.015) | (0.015) |
| Math score, 25th percentile | | 0.028 | -0.056* | 0.017 | -0.025* |
| | | (0.029) | (0.033) | (0.017) | (0.014) |
| Math score, 50th percentile | | 0.017 | -0.046 | 0.010 | -0.021 |
| | | (0.031) | (0.033) | (0.018) | (0.014) |
| Math core, 75th percentile | | -0.019 | -0.063* | -0.011 | -0.028* |
| | | (0.037) | (0.036) | (0.022) | (0.016) |
| Math score, 90th percentile | | -0.055 | -0.088** | -0.033 | -0.039** |
| | | (0.044) | (0.040) | (0.026) | (0.017) |
| | | *2013 Cohort (N = 415 schools)* | | | |
| Average math score (std) | | -0.025 | | -0.018 | |
| | | (0.048) | | (0.033) | |
| Math gap score | | 0.012 | | 0.009 | |
| | | (0.033) | | (0.022) | |
| Math score, 10th percentile | | -0.029 | | -0.021 | |
| | | (0.051) | | (0.036) | |
| Math score, 25th percentile | | -0.012 | | -0.008 | |
| | | (0.050) | | (0.034) | |
| Math score, 50th percentile | | -0.036 | | -0.026 | |
| | | (0.049) | | (0.034) | |
| Math core, 75th percentile | | -0.024 | | -0.017 | |
| | | (0.051) | | (0.036) | |
| Math score, 90th percentile | | -0.017 | | -0.012 | |
| | | (0.051) | | (0.035) | |
| | | *2014 Cohort (N = 373 schools)* | | | |
| Average math score (std) | -0.027 | | | | |
| | (0.044) | | | | |
| Math gap score | -0.082** | | | | |
| | (0.040) | | | | |
| Math score, 10th percentile | 0.011 | | | | |
| | (0.055) | | | | |
| Math score, 25th percentile | -0.005 | | | | |
| | (0.046) | | | | |
| Math score, 50th percentile | -0.027 | | | | |
| | (0.046) | | | | |
| Math core, 75th percentile | -0.069 | | | | |
| | (0.047) | | | | |
| Math score, 90th percentile | -0.063 | | | | |
| | (0.048) | | | | |

Notes: The analytic sample for the 2012 cohort includes all students enrolled in operational K-8, non-special education, non-charter schools identified in the spring of 2012. 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. TOT coefficients should be interpreted as the effect of one additional year under Focus status on the outcome of interest. Robust standard errors appear in parentheses. *** p<0.01, ** p<0.05, * p<0.1. ITT = intent to treat; TOT = treatment on the treated.

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

| | Priority | | | Focus | | |
|---|---|---|---|---|---|---|
| | 1 year after | 2 years after | 3 years after | 1 year after | 2 years after | 3 years after |
| Outcome | (1) | (2) | (3) | (4) | (5) | (6) |
| *A. Staffing* | | | | | | |
| Share of teachers new to school | 0.047 | 0.038 | 0.047 | 0.019 | 0.003 | -0.058 |
| | (0.057) | (0.065) | (0.102) | (0.021) | (0.022) | (0.032) |
| *B. Composition* | | | | | | |
| Log of total enrollment | -0.013 | 0.033 | 0.157 | -0.008 | -0.008 | 0.013 |
| | (0.043) | (0.090) | (0.140) | (0.020) | (0.036) | (0.050) |
| Share black | -0.007 | -0.023 | -0.024 | 0.005 | 0.010 | 0.012 |
| | (0.009) | (0.020) | (0.034) | (0.003) | (0.006) | (0.009) |
| Share economically disadvantaged | 0.000 | 0.020 | 0.013 | 0.006 | -0.001 | -0.008 |
| | (0.014) | (0.023) | (0.037) | (0.007) | (0.012) | (0.017) |

Notes: Each row within each panel (Priority, Focus) represents a separate specification, and reports effects of Priority or Focus designation on outcomes one, two, and three 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 Math Achievement: Evidence from a Dynamic RD Approach**

| | Priority | | Focus | |
|---|---|---|---|---|
| | 2 years after | 3 years after | 2 years after | 3 years after |
| Outcome | (1) | (2) | (3) | (4) |
| Average math score (std) | -0.023 | 0.006 | -0.009 | -0.052 |
| | (0.086) | (0.157) | (0.043) | (0.064) |
| Math gap score | -0.098 | -0.044 | -0.050 | 0.052 |
| | (0.097) | (0.157) | (0.041) | (0.074) |
| Math score, 10th percentile | 0.032 | 0.020 | 0.006 | -0.103 |
| | (0.094) | (0.179) | (0.040) | (0.065) |
| Math score, 25th percentile | 0.017 | 0.038 | 0.020 | -0.058 |
| | (0.085) | (0.142) | (0.042) | (0.061) |
| Math score, 50th percentile | -0.055 | -0.011 | -0.001 | -0.039 |
| | (0.086) | (0.170) | (0.047) | (0.067) |
| Math core, 75th percentile | -0.081 | -0.012 | -0.011 | -0.018 |
| | (0.105) | (0.183) | (0.052) | (0.078) |
| Math score, 90th percentile | -0.065 | -0.014 | -0.008 | -0.022 |
| | (0.124) | (0.193) | (0.058) | (0.089) |

Notes: Each row within each panel (Priority, Focus) represents a separate specification, and reports effects of Priority or Focus designation on outcomes one, two, and three 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.

**Appendix Table A1. Effects of Priority Designation on School Composition: Nonparametric Estimates**

| | ITT Estimates | | | TOT Estimates | |
|---|---|---|---|---|---|
| | 1 year after | 2 years after | 3 years after | 2 years after | 3 years after |
| Outcome | (1) | (2) | (3) | (4) | (5) |
| | | | *2012 Cohort* | | |
| Log of total enrollment | 0.182 | 0.221 | 0.148 | 0.136 | 0.069 |
| | [-0.259,0.565] | [-0.153,0.645] | [-0.268,0.557] | [-0.094,0.399] | [-0.124,0.260] |
| Nonparametric bandwidth | 0.31 | 0.35 | 0.37 | 0.35 | 0.37 |
| N(schools) | 111 | 125 | 134 | 125 | 134 |
| Share black | 0.135 | -0.101 | -0.094 | -0.064 | -0.046 |
| | [-0.246,0.450] | [-0.516,0.181] | [-0.510,0.190] | [-0.333,0.119] | [-0.257,0.099] |
| Nonparametric bandwidth | 0.31 | 0.23 | 0.24 | 0.23 | 0.24 |
| N(schools) | 111 | 82 | 81 | 82 | 81 |
| Share economically disadvantaged | 0.014 | -0.003 | -0.013 | -0.002 | -0.006 |
| | [-0.067,0.102] | [-0.087,0.085] | [-0.116,0.080] | [-0.054,0.053] | [-0.055,0.038] |
| Nonparametric bandwidth | 0.40 | 0.52 | 0.39 | 0.52 | 0.39 |
| N(schools) | 164 | 200 | 139 | 200 | 139 |
| | | | *2013 Cohort* | | |
| Log of total enrollment | 0.069 | -0.047 | | -0.027 | |
| | [-0.293,0.607] | [-0.469,0.398] | | [-0.277,0.239] | |
| Nonparametric bandwidth | 0.18 | 0.19 | | 0.019 | |
| N(schools) | 47 | 56 | | 56 | |
| Share black | -0.049 | -0.012 | | -0.007 | |
| | [-0.443,0.411] | [-0.404,0.469] | | [-0.220,0.256] | |
| Nonparametric bandwidth | 0.13 | 0.13 | | 0.13 | |
| N(schools) | 35 | 33 | | 33 | |
| Share economically disadvantaged | -0.083 | -0.012 | | -0.007 | |
| | [-0.265,0.073] | [-0.154,0.152] | | [-0.090,0.090] | |
| Nonparametric bandwidth | 0.16 | 0.20 | | 0.20 | |
| N(schools) | 42 | 61 | | 61 | |
| | | | *2014 Cohort* | | |
| Log of total enrollment | 0.313 | | | | |
| | [-0.141,0.760] | | | | |
| Nonparametric bandwidth | 0.18 | | | | |
| N(schools) | 60 | | | | |
| Share black | 0.179 | | | | |
| | [-0.366,0.812] | | | | |
| Nonparametric bandwidth | 0.18 | | | | |
| N(schools) | 62 | | | | |
| Share economically disadvantaged | -0.045 | | | | |
| | [-0.147,0.078] | | | | |
| Nonparametric bandwidth | 0.18 | | | | |
| N(schools) | 61 | | | | |

Notes: The analytic sample for the 2012 cohort includes all students enrolled in operational K-8, non-special education, non-charter schools identified in the spring of 2012. 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. TOT coefficients should be interpreted as the effect of one additional year under Priority status on the outcome of interest. Nonparametric estimates are calculated using a triangular kernel and data-driven bandwidth selection procedure described in Calonico, Cattaneo, and Titiunik (2015). Robust 95-percent confidence intervals are shown in brackets.*** p<0.01, ** p<0.05, * p<0.1. ITT = intent to treat; TOT = treatment on the treated. The N(schools) represent the number of effective observations used given the data-driven bandwidth selected.

**Appendix Table A2. Effects of Priority Designation on Math Achievement: Nonparametric Estimates**

| | ITT Estimates | | | TOT Estimates | |
|---|---|---|---|---|---|
| | 1 year after | 2 years after | 3 years after | 2 years after | 3 years after |
| Outcome | (1) | (2) | (3) | (4) | (5) |
| | | | *2012 Cohort* | | |
| Average math score (std) | | 0.184 | 0.198 | 0.120 | 0.098 |
| | | [-0.104,0.640] | [-0.125,0.724] | [-0.081,0.431] | [-0.073,0.369] |
| Nonparametric bandwidth | | 0.21 | 0.23 | 0.21 | 0.23 |
| N(schools) | | 73 | 81 | 73 | 81 |
| | | | *2013 Cohort* | | |
| Average math score (std) | | 0.004 | | 0.002 | |
| | | [-0.490,0.535] | | [-0.281,0.306] | |
| Nonparametric bandwidth | | 0.18 | | 0.18 | |
| N(schools) | | 45 | | 45 | |
| | | | *2014 Cohort* | | |
| Average math score (std) | -0.305 | | | | |
| | [-0.809,0.125] | | | | |
| Nonparametric bandwidth | 0.18 | | | | |
| N(schools) | 59 | | | | |

Notes: The analytic sample for the 2012 cohort includes all students enrolled in operational K-8, non-special education, non-charter schools identified in the spring of 2012. For subsequent cohorts, we exclude schools identified as Priority in prior years since they leave the risk set for treatment after initial designation for intervention. TOT coefficients should be interpreted as the effect of one additional year under Priority status on the outcome of interest. Nonparametric estimates are calculated using a triangular kernel and data-driven bandwidth selection procedure described in Calonico, Cattaneo, and Titiunik (2015). Robust 95-percent confidence intervals are shown in brackets.*** p<0.01, ** p<0.05, * p<0.1. ITT = intent to treat; TOT = treatment on the treated. The N(schools) represent the number of effective observations used given the data-driven bandwidth selected.

**Appendix Table A3. Effects of Focus Designation on School Composition: Nonparametric Estimates**

| | ITT Estimates | | | TOT Estimates | |
|---|---|---|---|---|---|
| | 1 year after | 2 years after | 3 years after | 2 years after | 3 years after |
| Outcome | (1) | (2) | (3) | (4) | (5) |
| | | | *2012 Cohort* | | |
| Log of total enrollment | 0.112 | 0.114 | 0.123 | 0.063 | 0.033 |
| | [-0.063,0.361] | [-0.064,0.359] | [-0.060,0.381] | [-0.031,0.215] | [-0.028,0.147] |
| Nonparametric bandwidth | 0.35 | 0.37 | 0.33 | 0.45 | 0.5 |
| N(schools) | 455 | 474 | 419 | 592 | 646 |
| Share black | -0.003 | -0.005 | -0.005 | -0.002 | -0.001 |
| | [-0.072,0.072] | [-0.073,0.066] | [-0.074,0.067] | [-0.044,0.038] | [-0.033,0.028] |
| Nonparametric bandwidth | 0.43 | 0.47 | 0.47 | 0.57 | 0.59 |
| N(schools) | 580 | 617 | 610 | 758 | 768 |
| Share economically disadvantaged | -0.050 | -0.072 | -0.067 | -0.025 | -0.018 |
| | [-0.156,0.028] | [-0.184,0.008] | [-0.179,0.013] | [-0.093,0.015] | [-0.073,0.013] |
| Nonparametric bandwidth | 0.45 | 0.41 | 0.43 | 0.58 | 0.57 |
| N(schools) | 592 | 539 | 558 | 772 | 747 |
| | | | *2013 Cohort* | | |
| Log of total enrollment | -0.017 | 0.047 | | 0.150 | |
| | [-0.350,0.344] | [-0.223,0.383] | | [-0.042,0.397] | |
| Nonparametric bandwidth | 0.33 | 0.28 | | 0.127 | |
| N(schools) | 258 | 201 | | 85 | |
| Share black | 0.014 | -0.023 | | -0.014 | |
| | [-0.124,0.131] | [-0.166,0.083] | | [-0.101,0.080] | |
| Nonparametric bandwidth | 0.35 | 0.29 | | 0.19 | |
| N(schools) | 277 | 209 | | 125 | |
| Share economically disadvantaged | -0.062 | -0.094 | | -0.080* | |
| | [-0.253,0.085] | [-0.263,0.018] | | [-0.187,0.020] | |
| Nonparametric bandwidth | 0.22 | 0.23 | | 0.14 | |
| N(schools) | 159 | 162 | | 91 | |
| | | | *2014 Cohort* | | |
| Log of total enrollment | -0.020 | | | | |
| | [-0.312,0.266] | | | | |
| Nonparametric bandwidth | 0.20 | | | | |
| N(schools) | 136 | | | | |
| Share black | -0.078 | | | | |
| | [-0.278,0.069] | | | | |
| Nonparametric bandwidth | 0.12 | | | | |
| N(schools) | 92 | | | | |
| Share economically disadvantaged | 0.098* | | | | |
| | [-0.028,0.247] | | | | |
| Nonparametric bandwidth | 0.21 | | | | |
| N(schools) | 149 | | | | |

Notes: The analytic sample for the 2012 cohort includes all students enrolled in operational K-8, non-special education, non-charter schools identified in the spring of 2012. 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. TOT coefficients should be interpreted as the effect of one additional year under Focus status on the outcome of interest. Nonparametric estimates are calculated using a triangular kernel and data-driven bandwidth selection procedure described in Calonico, Cattaneo, and Titiunik (2015). Robust 95-percent confidence intervals are shown in brackets.*** $p<0.01$, ** $p<0.05$, * $p<0.1$. ITT = intent to treat; TOT = treatment on the treated. The N(schools) represent the number of effective observations used given the data-driven bandwidth selected.

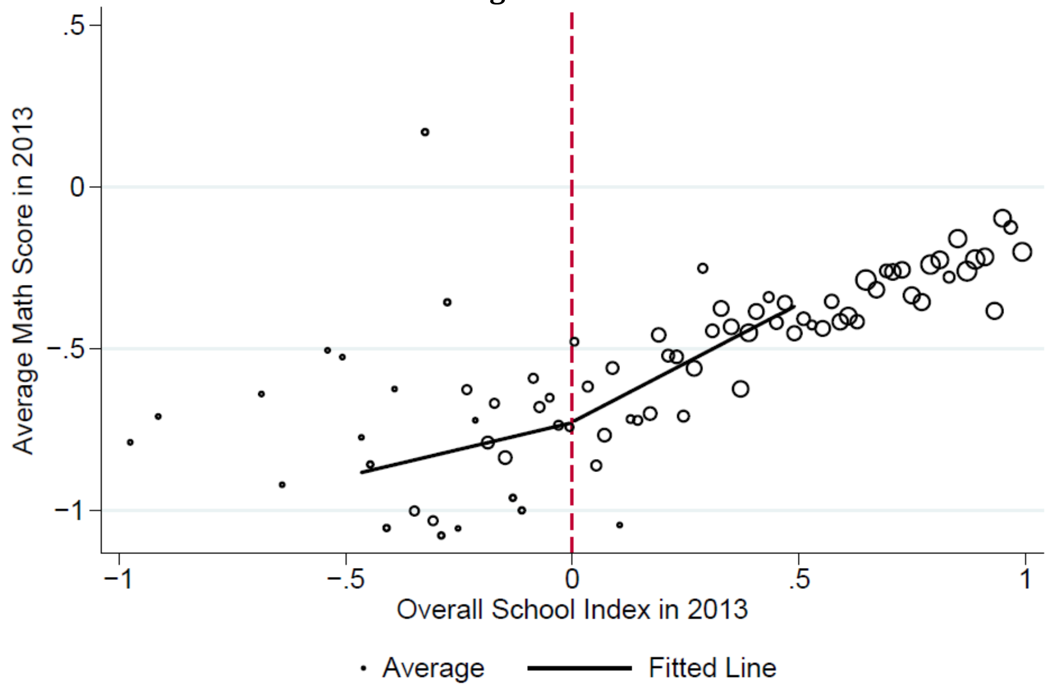
**Appendix Table A4. Effects of Focus Designation on Math Achievement: Nonparametric Estimates**

| | ITT Estimates | | | TOT Estimates | |
|---|---|---|---|---|---|
| | 1 year after | 2 years after | 3 years after | 2 years after | 3 years after |
| Outcome | (1) | (2) | (3) | (4) | (5) |
| | | | *2012 Cohort* | | |
| Average math score (std) | | -0.040 | -0.062 | -0.015 | -0.028 |
| | | [-0.182,0.128] | [-0.213,0.116] | [-0.120,0.099] | [-0.099,0.054] |
| Nonparametric bandwidth | | 0.58 | 0.50 | 0.47 | 0.49 |
| N(schools) | | 747 | 625 | 585 | 612 |
| Math gap score | | -0.092** | -0.006 | -0.057** | -0.003 |
| | | [-0.187,-0.003] | [-0.071,0.067] | [-0.120,0.000] | [-0.034,0.029] |
| Nonparametric bandwidth | | 0.49 | 0.52 | 0.45 | 0.53 |
| N(schools) | | 627 | 662 | 553 | 674 |
| | | | *2013 Cohort* | | |
| Average math score (std) | | 0.104 | | 0.124 | |
| | | [-0.092,0.364] | | [-0.023,0.302] | |
| Nonparametric bandwidth | | 0.26 | | 0.12 | |
| N(schools) | | 177 | | 77 | |
| Math gap score | | 0.047 | | 0.025 | |
| | | [-0.146,0.265] | | [-0.098,0.155] | |
| Nonparametric bandwidth | | 0.15 | | 0.12 | |
| N(schools) | | 87 | | 74 | |
| | | | *2014 Cohort* | | |
| Average math score (std) | -0.116 | | | | |
| | [-0.410,0.163] | | | | |
| Nonparametric bandwidth | 0.16 | | | | |
| N(schools) | 112 | | | | |
| Math gap score | -0.228*** | | | | |
| | [-0.450,-0.064] | | | | |
| Nonparametric bandwidth | 0.14 | | | | |
| N(schools) | 99 | | | | |

Notes: The analytic sample for the 2012 cohort includes all students enrolled in operational K-8, non-special education, non-charter schools identified in the spring of 2012. 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. TOT coefficients should be interpreted as the effect of one additional year under Focus status on the outcome of interest. Nonparametric estimates are calculated using a triangular kernel and data-driven bandwidth selection procedure described in Calonico, Cattaneo, and Titiunik (2015). Robust 95-percent confidence intervals are shown in brackets.*** p<0.01, ** p<0.05, * p<0.1. ITT = intent to treat; TOT = treatment on the treated. The N(schools) represent the number of effective observations used given the data-driven bandwidth selected.

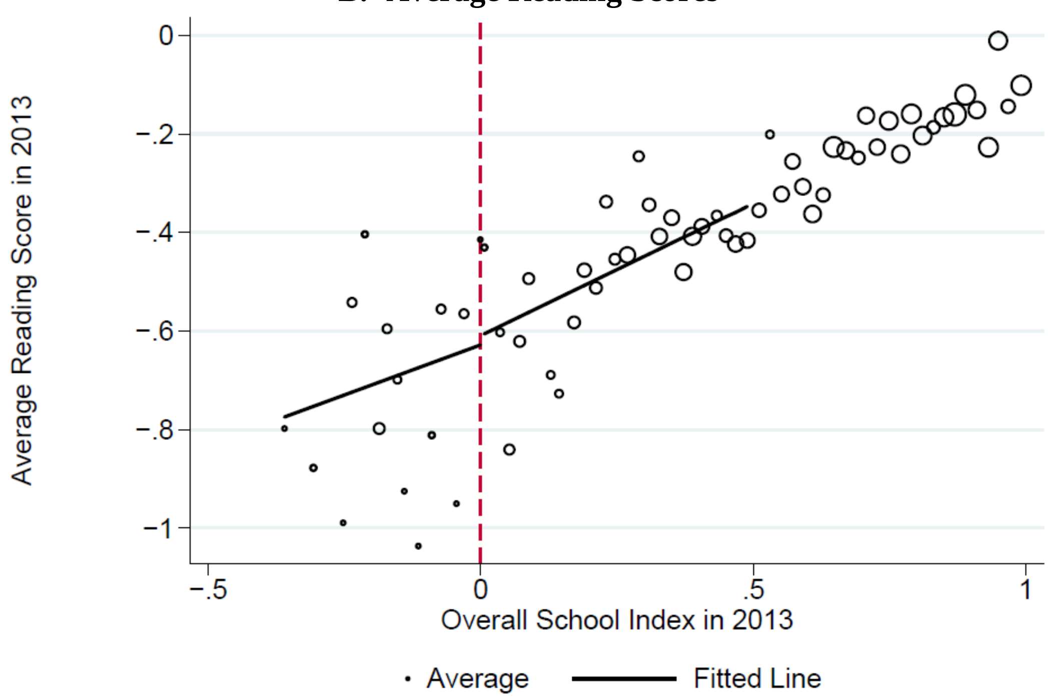
**Appendix Figure A1. Baseline Equivalence: Priority Schools, 2013 Cohort**

**A. Average Math Scores**



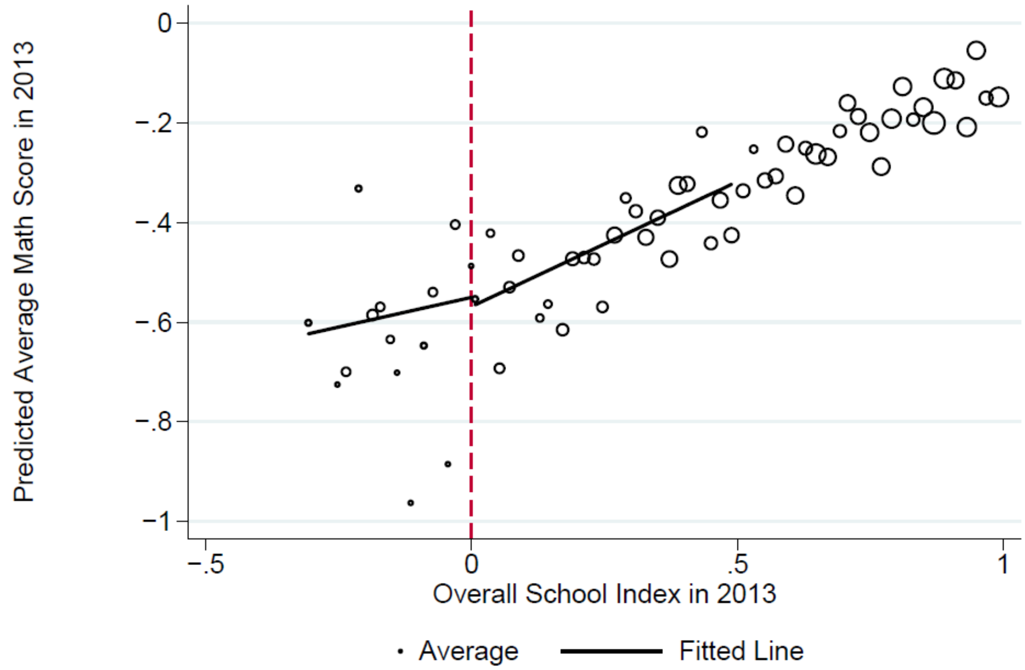

Linear: Estimate = −0.002 (0.078), N(schools) = 253, CM = −0.522, SD = 0.274
Nonparametric (mserd): Estimate = 0.013 [−0.186, 0.253], p = 0.765, bw = 0.379, N(schools) = 191, CM = −0.733

**B. Average Reading Scores**



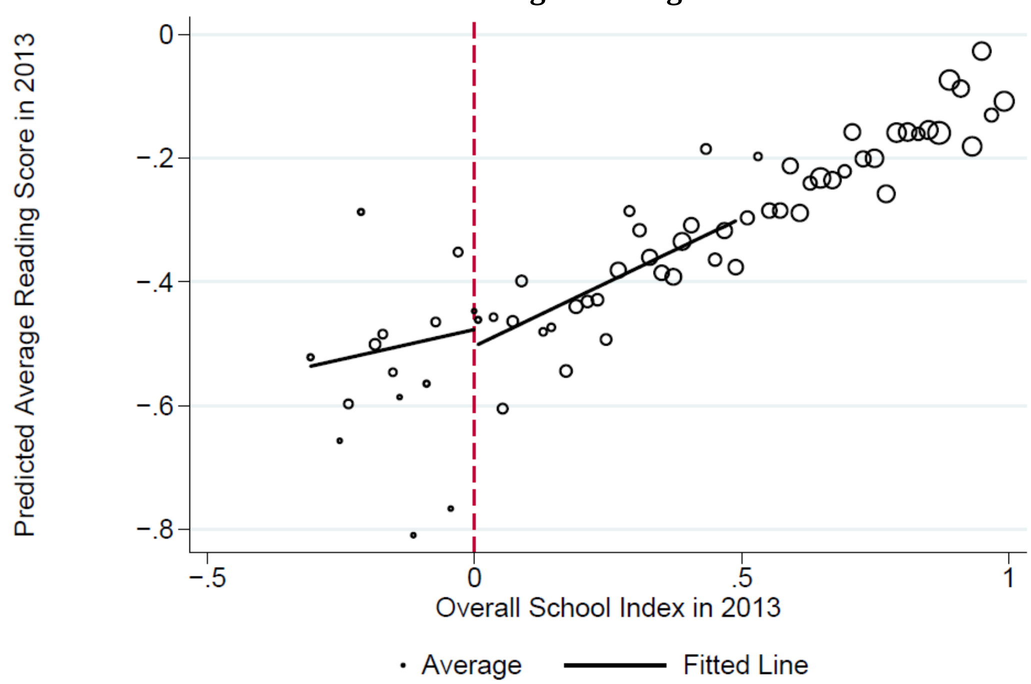

Linear: Estimate = −0.019 (0.094), N(schools) = 219, CM = −0.452, SD = 0.230
Nonparametric (mserd): Estimate = 0.075 [−0.388, 0.533], p = 0.757, bw = 0.146, N(schools) = 42, CM = −0.555

**C. Predicted Average Math Scores**




Linear: Estimate = 0.018 (0.096), N(schools) = 214, CM = −0.421, SD = 0.241
Nonparametric (mserd): Estimate = 0.115 [−0.241, 0.566], p = 0.429, bw = 0.169, N(schools) = 48, CM = −0.538

**D. Predicted Average Reading Scores**

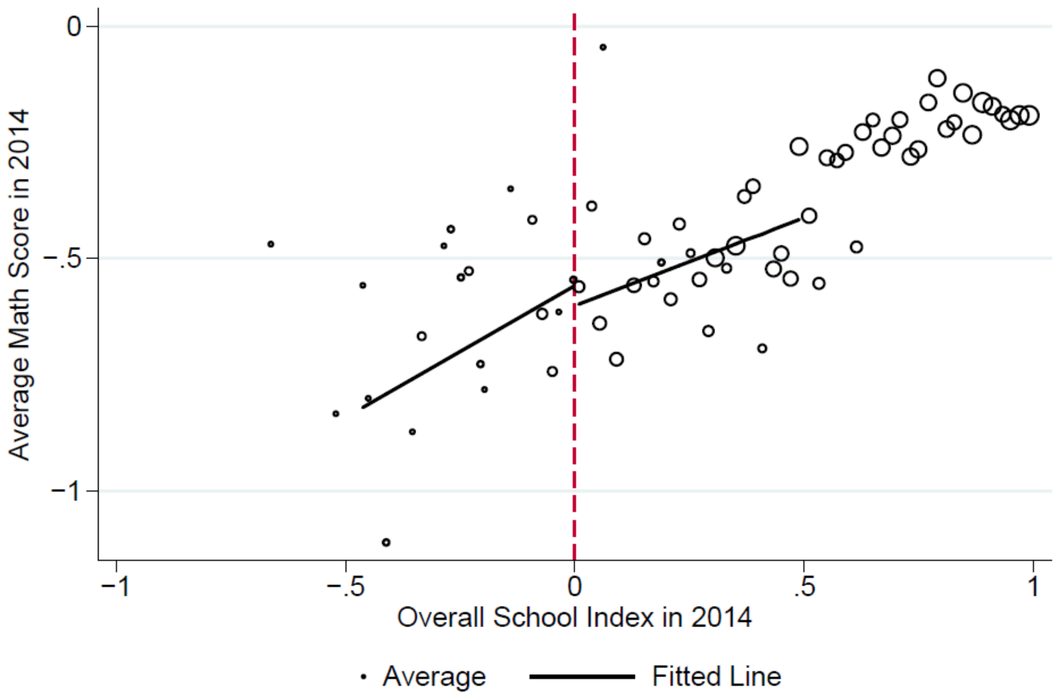



Linear: Estimate = 0.027 (0.081), N(schools) = 214, CM = −0.382, SD = 0.209
Nonparametric (mserd): Estimate = 0.137 [−0.173, 0.524], p = 0.324, bw = 0.152, N(schools) = 43, CM = −0.502

Notes: Analytic sample excludes schools identified as Priority in 2012. Parametric, linear specification is weighted by school enrollment. Nonparametric estimate is 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.

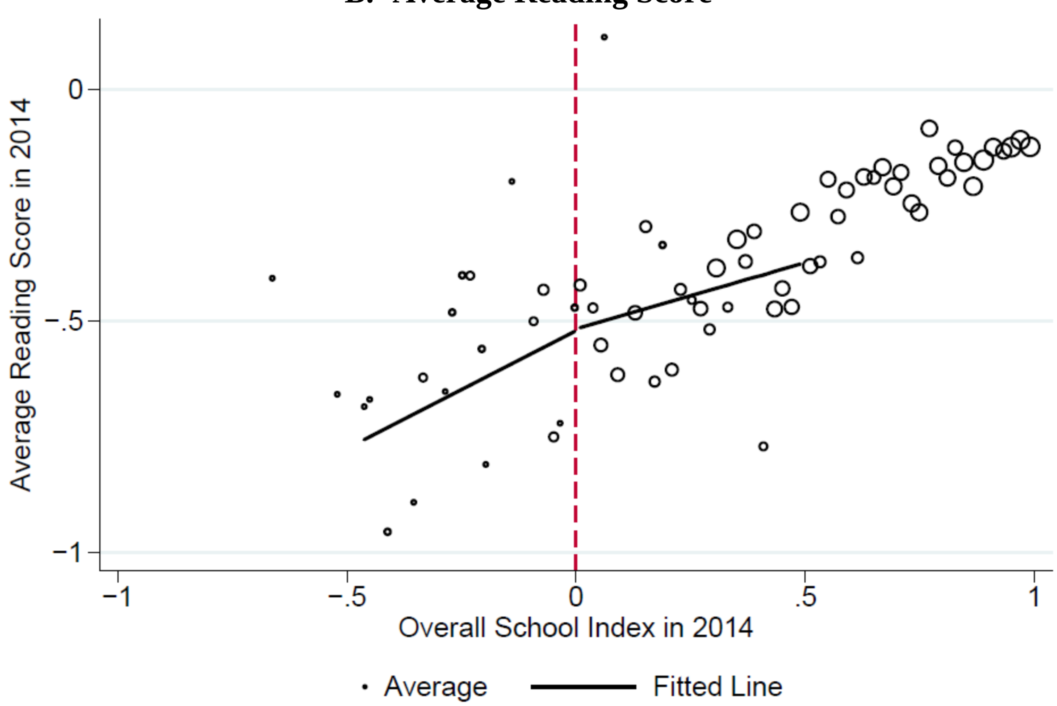
**Appendix Figure A2. Baseline Equivalence: Priority Schools, 2014 Cohort**

**A. Average Math Score**



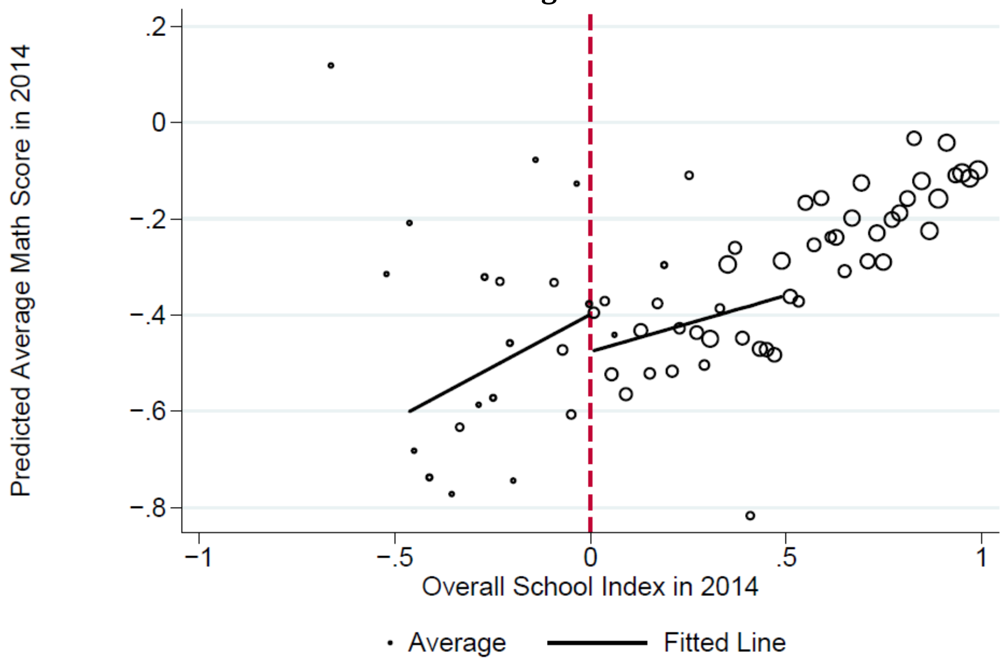

Linear: Estimate = 0.043 (0.083), N(schools) = 205, CM = −0.494, SD = 0.260
Nonparametric (mserd): Estimate = −0.131 [−0.410, 0.092], p = 0.215, bw = 0.170, N(schools) = 59, CM = −0.519

**B. Average Reading Score**



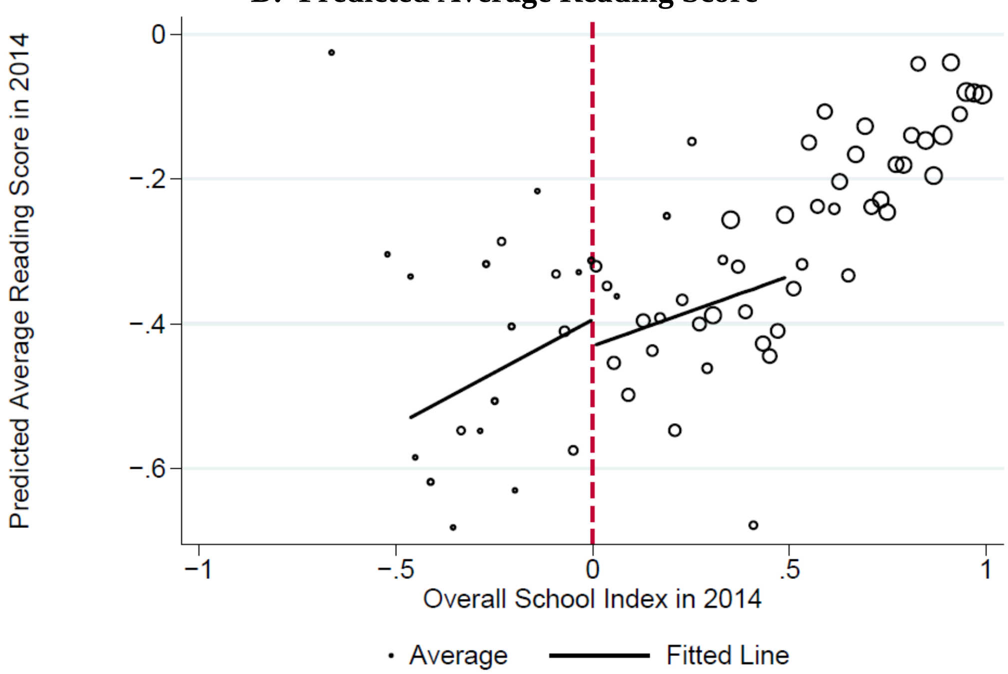

Linear: Estimate = −0.004 (0.092), N(schools) = 205, CM = −0.436, SD = 0.237
Nonparametric (mserd): Estimate = −0.183 [−0.535, 0.129], p = 0.231, bw = 0.179, N(schools) = 60, CM = −0.447

**C. Predicted Average Math Score**




Linear: Estimate = 0.077 (0.090), N(schools) = 203, CM = −0.410, SD = 0.256
Nonparametric (mserd): Estimate = −0.054 [−0.472, 0.314], p = 0.694, bw = 0.173, N(schools) = 58, CM = −0.404

**D. Predicted Average Reading Score**

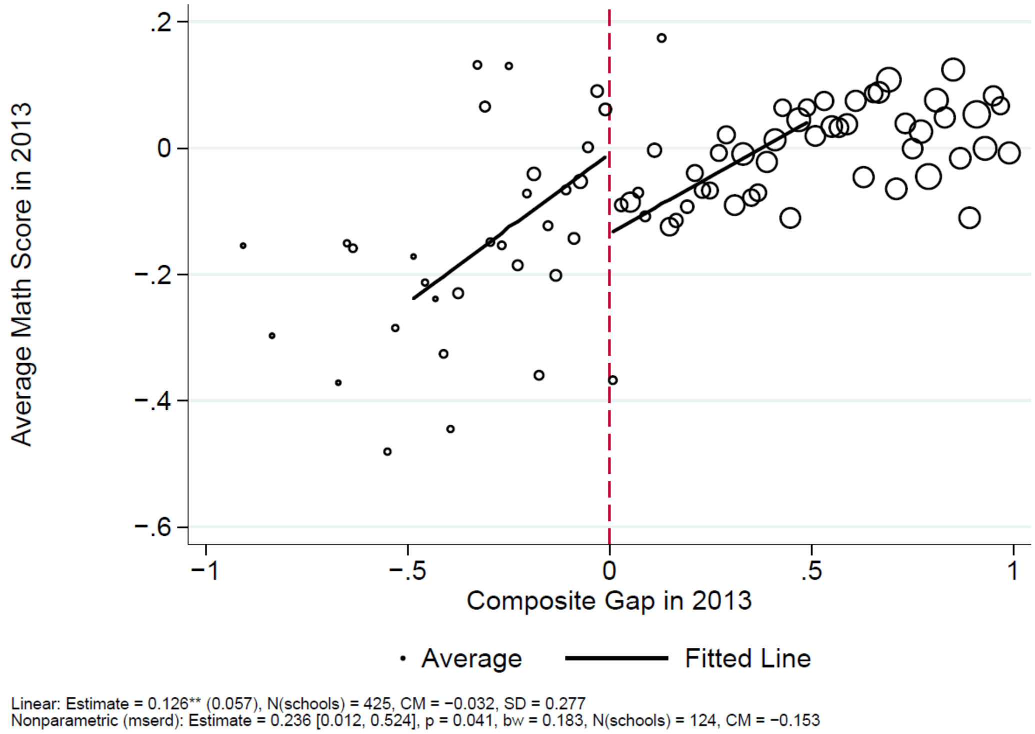



Linear: Estimate = 0.036 (0.074), N(schools) = 203, CM = −0.376, SD = 0.220
Nonparametric (mserd): Estimate = −0.081 [−0.428, 0.237], p = 0.574, bw = 0.171, N(schools) = 58, CM = −0.338

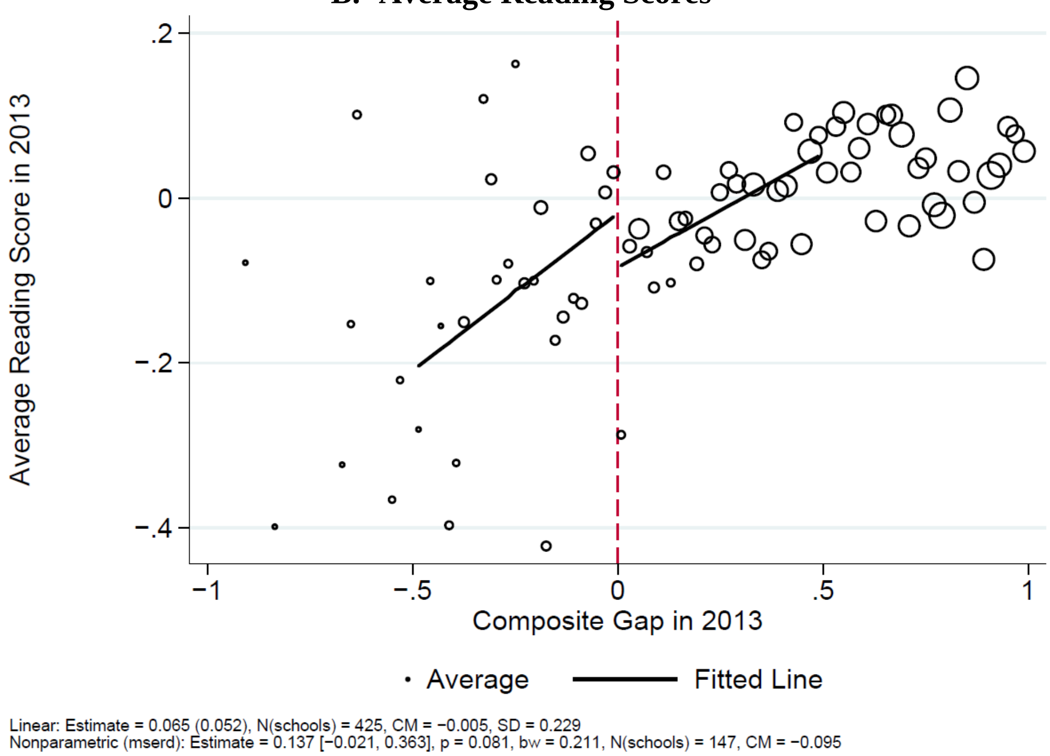
Notes: Analytic sample excludes schools identified as Priority in 2013 or 2012. Parametric, linear specification is weighted by school enrollment. Nonparametric estimate is based on approach of Calonico et al. (2014a, 2014b, 2015). Predicted average scores come from regressions of test scores on the set of school characteristics reported in Table 1. CM = control group mean; SD = standard deviation for control group; bw = bandwidth.

**Appendix Figure A3. Baseline Equivalence: Focus Schools, 2013 Cohort**

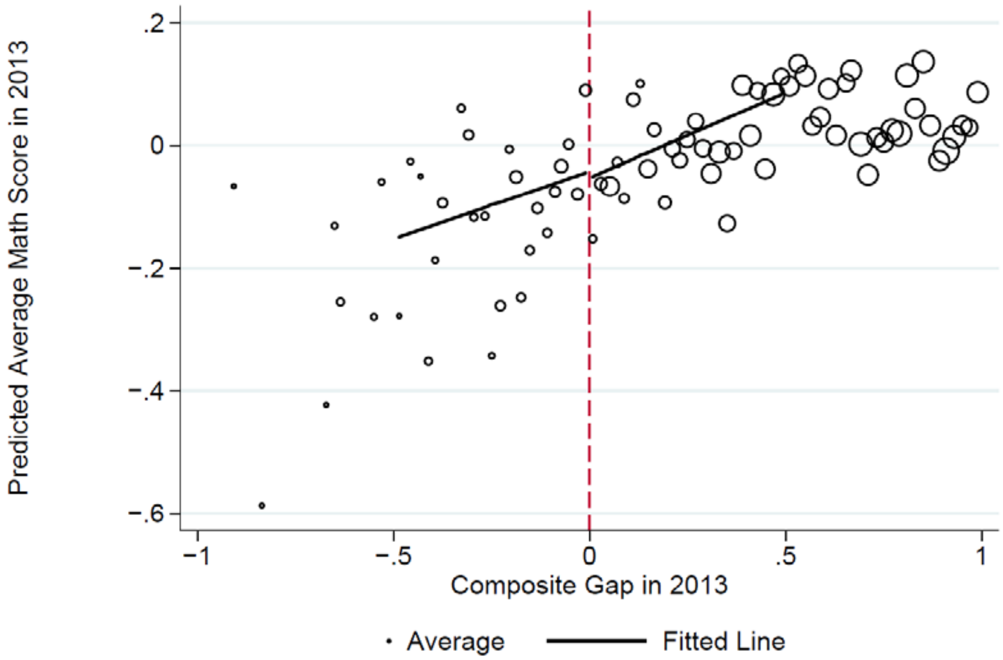
**A. Average Math Scores**




**B. Average Reading Scores**

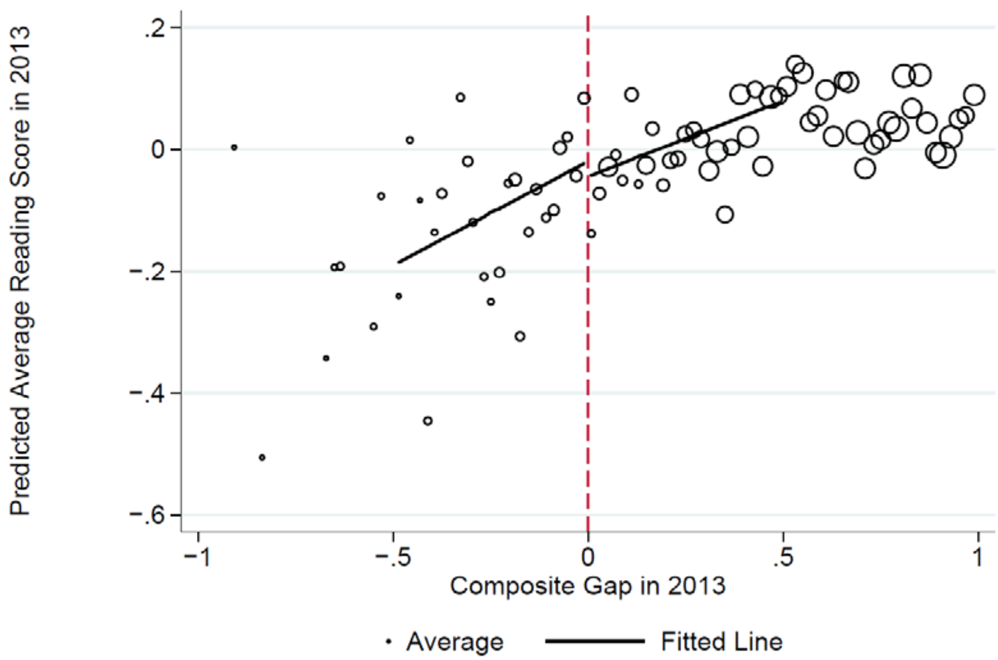


### C. Predicted Average Math Scores




Linear: Estimate = 0.010 (0.060), N(schools) = 420, CM = 0.026, SD = 0.253
Nonparametric (mserd): Estimate = 0.094 [−0.088, 0.353], p = 0.239, bw = 0.243, N(schools) = 171, CM = −0.086

### D. Predicted Average Reading Scores

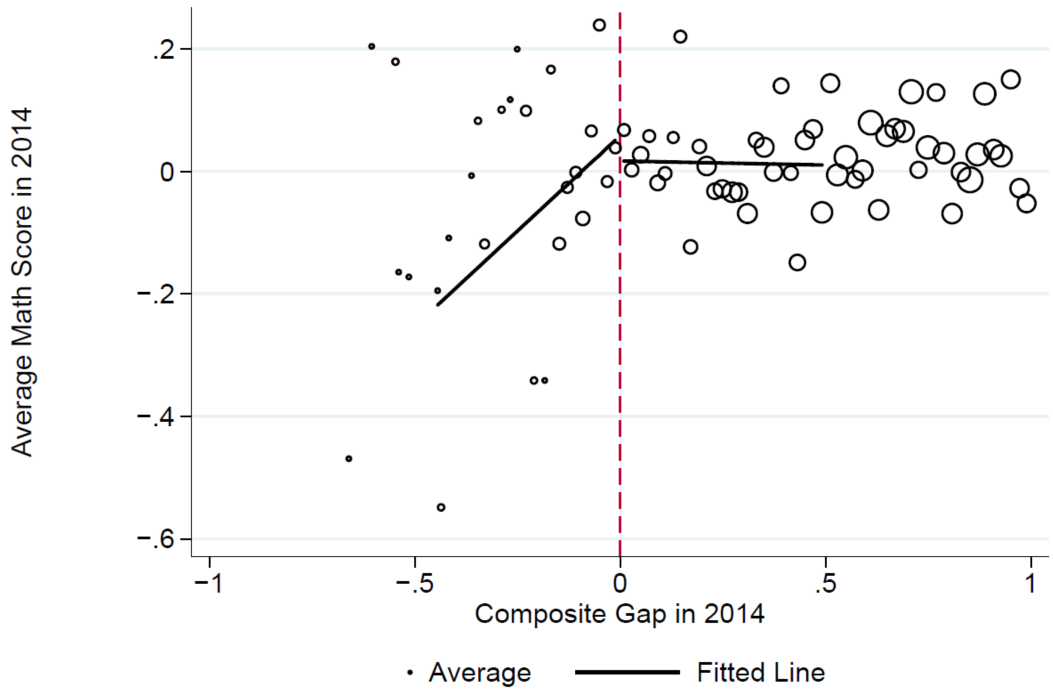



Linear: Estimate = 0.025 (0.057), N(schools) = 420, CM = 0.027, SD = 0.234
Nonparametric (mserd): Estimate = 0.094 [−0.065, 0.323], p = 0.193, bw = 0.247, N(schools) = 173, CM = −0.062

Notes: Analytic sample excludes schools identified as Focus or Priority in 2012. Parametric, linear specification is weighted by school enrollment. Nonparametric estimate is 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.

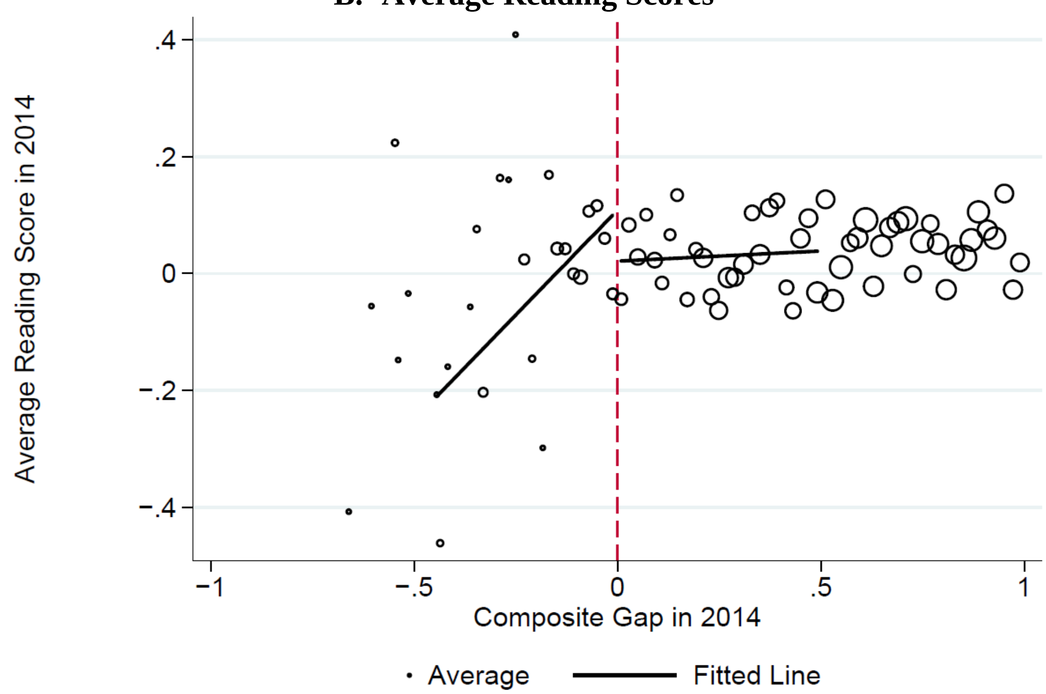
**Appendix Figure A4. Baseline Equivalence: Focus Schools, 2014 Cohort**

**A. Average Math Scores**



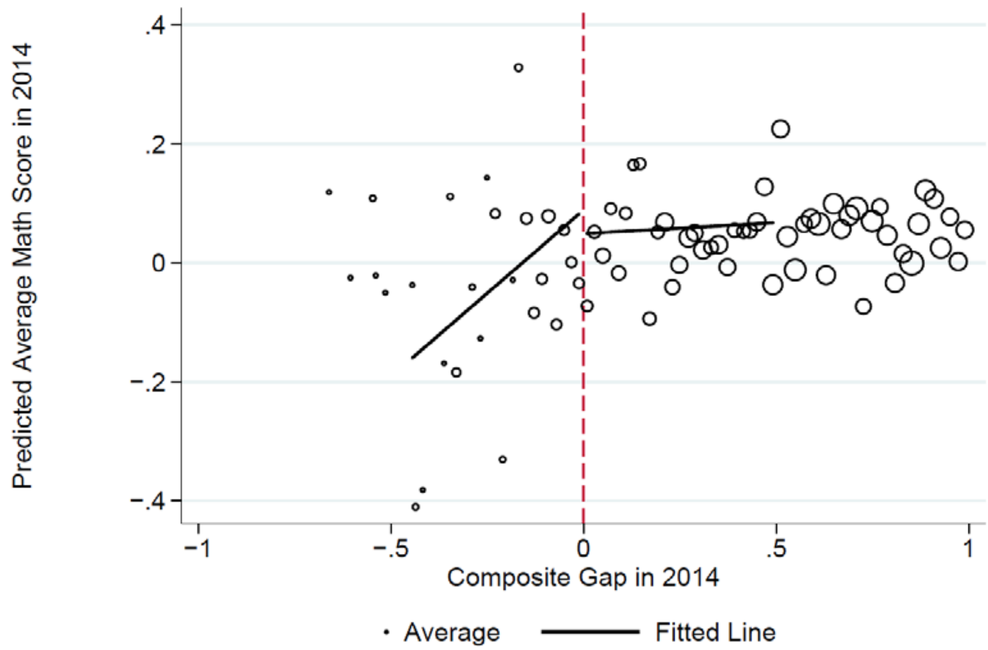

Linear: Estimate = 0.041 (0.076), N(schools) = 378, CM = 0.006, SD = 0.311
Nonparametric (mserd): Estimate = 0.022 [−0.329, 0.380], p = 0.888, bw = 0.173, N(schools) = 123, CM = 0.033

**B. Average Reading Scores**



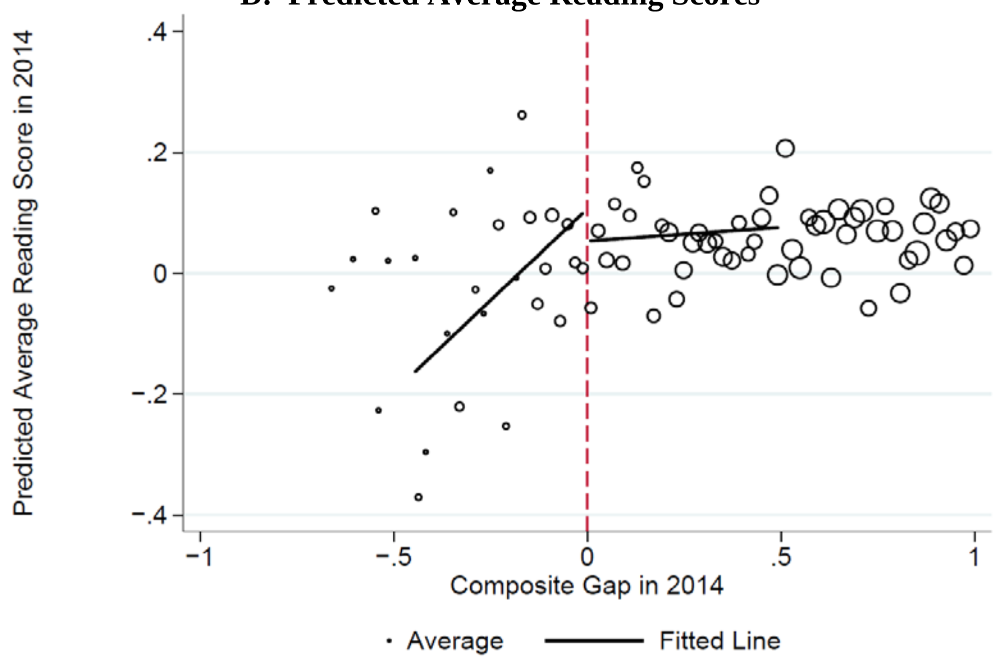

Linear: Estimate = 0.087 (0.054), N(schools) = 378, CM = 0.025, SD = 0.242
Nonparametric (mserd): Estimate = −0.012 [−0.249, 0.227], p = 0.926, bw = 0.188, N(schools) = 130, CM = 0.024

## C. Predicted Average Math Scores




Linear: Estimate = 0.040 (0.057), N(schools) = 378, CM = 0.056, SD = 0.248
Nonparametric (mserd): Estimate = −0.001 [−0.191, 0.218], p = 0.900, bw = 0.174, N(schools) = 123, CM = −0.040

## D. Predicted Average Reading Scores




Linear: Estimate = 0.053 (0.052), N(schools) = 378, CM = 0.062, SD = 0.232
Nonparametric (mserd): Estimate = 0.010 [−0.170, 0.219], p = 0.803, bw = 0.177, N(schools) = 125, CM = −0.021

Notes: Analytic sample excludes schools identified as Focus or Priority in 2013 or 2012. Parametric, linear specification is weighted by school enrollment. Nonparametric estimate is 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.

## Appendix B. Calculation of Top-to-Bottom (TTB) Index and the Identification of Priority and Focus Schools in Michigan

Overview

The TTB is a performance index that was developed in order to implement a system of differentiated accountability and supports under Michigan's approved waiver from NCLB. The index ranks public schools on student performance in mathematics, reading, writing, science, and social studies, as well as graduation rates (for high schools). Performance in each tested subject (i.e., content area) is measured in three ways: level of achievement, growth in achievement over time, and the gap in achievement between the highest 30 percent and lowest 30 percent of students within each school.

Which schools get a TTB ranking?

Schools must have at least 30 full-academic-year (FAY) students in either the elementary/middle or high school span (or both) with test scores over two years in at least two subjects.[1] Traditional public schools as well as charter schools that meet this criterion are ranked.

Which students are included in a school's TTB ranking?

All FAY public school students with valid test scores on regular or alternate assessments within eligible schools are included. Thus, special education students who traditionally take alternate exams are included. Homeschooled and private school students with state test scores are not included.

How does Michigan calculate the overall TTB score?

We describe the rules that identified the first cohort of Priority and Focus schools in August of 2012. To calculate the TTB index for each school, the Michigan Department of Education (MDE) used students' test scores[2] over a minimum of two prior years.

After assigning each student to her main school for each academic year, the state calculated standardized scale scores for each student in each subject (and test type) in which she was tested by year and grade level. These standardized scores provide a measure of how well a student performed relative to her peers across the state who took the same test in the same year.

For each subject area in which a school had the requisite number of tested students over a period of two prior years, the state created a "content area index" that was a function of prior test scores, change in student performance, and the within-school gap between the top 30 percent and bottom 30 percent of students. Below, we describe how the state calculated each of these components for a given subject area (e.g., mathematics):

[1] Closed schools are provided a ranking if adequate historical data exist. A school classified as both an elementary/middle and high school has ranks calculated for both sets of grades and the final rank is an average of the two.

[2] Michigan's exams include the Michigan Educational Assessment Program (MEAP) tests given in mathematics, reading, writing, science, and social studies across grades 3 through 9; MEAP-Access, which is the alternate assessment given to students with special education needs; and the Michigan Merit Examination (MME), given in 11th grade, which includes the ACT.

- Two-year average scores (weight = 50%): For each school, the state calculated the average of student-level standardized scale scores in subject X in year t-1 (i.e., the prior year) and year t-2 (i.e., two years prior). Along with these means, the state counted the number of students tested in subject X in the school for each of those prior two years. Then the state calculated a weighted average of these two means and standardized the resulting average across comparable schools (i.e., elementary/middle or high schools). That value became the school-level standardized score for subject X.

- Growth in student performance (weight = 25%): The state used a two-pronged approach to assess improvements in performance. For a given school, the choice of approach for each subject/content area rested on whether students were tested in that subject over multiple years (e.g., math tests in grades 3, 4, and 5).

  For subjects in which adjacent-year testing occurred (e.g., math and reading in elementary and middle school grades), the state used a minimum of three years of historical data to classify students into different groups of "performance level changes," which were based on a student's movement across sub-categories of the four main proficiency-level categories: not proficient, partially proficient, proficient, and advanced.

  The state divided each proficiency-level category into three sub-categories: e.g., within proficient there is low-proficient, middle-proficient, and high-proficient. Each sub-category was constructed so that the range of included scale scores was less than the standard error of all scale scores within the broader proficiency category.

  Then the state convened a panel of experts to associate movements by students across those 12 sub-proficiency-level categories with the following descriptors (in reality, this was a "policy judgment" of what constituted "significant improvement," etc.):[3]
  - Significant decline = decrease of 3 or more sub-performance-level categories
  - Decline = decrease of 1 or 2 sub-performance-level categories
  - Maintain = no change in sub-performance-level category
  - Improvement = increase of 1 or 2 sub-performance-level categories
  - Significant improvement = increase of 3 or more sub-performance-level categories

  The state counted up students who fell into each of the above improvement categories according to the following table, which delineates between students who were previously proficient and those who were not:

[3] The one exception to this set of movement rules was if a student started out in the "advanced" proficiency category and exhibited a decline across the sub-levels of that "advanced" category, thus still remaining "advanced" – in this case the student was classified as "maintaining" and not as "decline" or "significant decline."

| Previously Proficient? | Performance Level Change | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Year t-1 | | | | | Year t -2 | | | | |
| | SD | D | M | I | SI | SD | D | M | I | SI |
| Yes | | | | | | | | | | |
| No | | | | | | | | | | |

For a school and subject, the state totaled the number of students with performance level change values separately by prior proficiency status and the applied the following weights to those counts:

| Previously Proficient? | SD | D | M | I | SI |
|---|---|---|---|---|---|
| No | -2 | -1 | 0 | 1 | 2 |
| Yes | -2 | -1 | 1 | 1 | 2 |

For example, the number of students who were previously proficient and made "significant declines" was multiplied by -2.

Then the state added up all of the weighted "performance level change" counts (across the two years in the table of Ns above) and divided that total by the sum of students with "performance level change" values across those same two years.

Finally, the state standardized the resultant average across schools of the same type (elementary/middle or high school) and this standardized value became the "change/improvement index value" for a given subject and school.

In cases where adjacent-year testing did not occur (i.e., for all calculations in high school grades as well as in subjects such as science, social studies, and writing), the state used a minimum of two and maximum of four years of prior, student-level standardized test scores (for subject X) to calculate yearly means. If the school had three or four years of data attached to it for subject X, the state fit a simple linear regression line to the mean scores and recorded the slope. If the school only had two years of prior scores with which to work, the state calculated the simple difference and took the result as the improvement value. The standardized value (across all comparable schools) of this change/slope became the "improvement index" for subject X for a given school.

Thus, the improvement index calculated for a given subject for a school could either be based on the standardized, weighted change in students' performance levels over three prior years or the standardized slope of average, standardized scale scores across two to four prior years.[4]

[4] Therefore, within the calculations for a school's "math context area index," the improvement component could be based on changes in students' performance levels – but for the same school, the improvement index component of the "science content area index" might be based on the slope approach.

- Gaps in student performance (weight = 25%): Once each student within a school had a standardized score (i.e., a z-score), the state arrayed all students attached to the school by z-score (within a subject area) for each of two prior years. Next the state took the average standardized scale score of the bottom 30 percent of students across those two years and subtracted from it the average standardized scale score of the top 30 percent of students across those two prior years to yield a subject-specific gap score. The resultant difference was standardized across comparable schools (i.e., elementary/middle or high school) and became the "achievement gap index" for subject X.

For each subject area, the state linearly combined scores for the above components according to weights in parentheses (0.5 for level achievement, 0.25 for growth, and 0.25 for gaps) to arrive at a "content area index" value, and standardized this value across schools.[5] For each school, the state took a simple linear combination of the standardized content-area indices (equally weighting each subject area) to arrive at the overall TTB index score.

If the school is a high school, its TTB index also included a subcomponent that was a function of prior four-year graduation rates: First, the state calculated a two-year average of a school's four-year graduation rate and standardized the resultant average across all schools. Second, the state computed the change (or slope) in four-year graduation rates based on a minimum of two years and a maximum of four years of historical graduation data, and standardized that slope across schools. The first part of this high school graduation index was weighted by two-thirds and the second part by one-third. The standardized value of their linear combination is the high school graduation index score. When calculating the overall TTB index score, the high school graduation index was weighted by 10 percent, with the remaining 90 percent evenly apportioned among the number of subjects for which a school had content area index values.

How does Michigan use the TTB ranking to identify Priority Schools?
Using the overall TTB index scores for all schools, the state ranked schools from highest to lowest. Next the state identified the subset of schools with TTB scores in the bottom 5 percent of this overall distribution. Federal guidance required that states identify the lowest performing 5 percent of Title I schools. Thus, once Michigan had identified the bottom 5 percent of all public schools, it counted the subset of those low-performing schools that were Title I (participating or eligible) and ensured that the resulting number was equal to or greater than 5 percent of the total number of Title I schools in the state in that year. In Michigan, the bottom 5 percent of all public schools included 5 percent of the stock of Title I public schools in the state.

How does Michigan use the TTB ranking to identify Focus Schools?
Michigan used one particular component of the overall TTB ranking to identify Focus schools: the achievement gap index. For each school, the state took a simple average of all available subject-area-specific achievement gap indices[6] to generate a composite gap index.

[5] In cases where performance level change information is not available for a school and in cases where the most recent year's proficiency rate is at or above 90%, the state omitted the "growth" part of the content area and weighted the remaining two components: level achievement and gaps.

[6] To identify the first cohort of Focus schools (in 2012), the state calculated the "average gap" by including all available subjects regardless of the number of FAY students with test scores (as long as the school had already met the criterion of having at least 30 FAY students in two subjects over two years). In all subsequent cohorts, the state only averaged subject-specific gap scores over subjects with at least 30 FAY students.

The state ranked schools by this composite gap index. Since federal waiver guidance required states to identify 10 percent of non-Priority, Title I schools with the largest achievement gaps as Focus, the state moved up the distribution of the composite gap index until it reached a value below which a number of non-Priority Title I schools equivalent to 10 percent of the population of Title I schools fell. Michigan labeled all schools, regardless of Title I status, below that cutoff as Focus schools.

Additional Changes to TTB Calculations for Subsequent Cohorts of Priority and Focus Schools
The most significant changes for this year were related to the identification of Focus schools. First, the state normalized students' scale scores before standardizing and examining within-school gaps between the average for the top 30 percent of students and the bottom 30 percent of students. Second, the state capped z-scores at 2 and -2. The intent of these adjustments was to reduce the capacity of outliers (especially in smaller schools) to disproportionately influence measures of gaps. Third, there was an "audit" function where a school that was flagged as Focus was removed if the bottom 30 percent of students in that school performed better than the state average in at least two subjects, and if that school's overall TTB ranking was above the 75th percentile.

## Appendix B: References

Michigan Department of Education (2012a). "Top-to-Bottom Ranking, Priority, Focus, and Rewards Schools Identification Business Rules: 2011-2012." Technical Report.

Michigan Department of Education (2012b). "2012 Top-to-Bottom Ranking: Understanding How the Ranking is Calculated." Presentation.

Michigan Department of Education. "Top-to-Bottom School Rankings: Historical Ranking Information." Website: https://www.michigan.gov/mde/0,4615,7-140-22709_56562---00.html

Personal Communication with Venessa Keesler (March 18, 2016), Joseph Martineau (March 31, 2106) and Alex Schwarz (April 14, 2016)