The Rhode to Turnaround:
The Impact of Waivers to No Child Left Behind on School Performance

Shaun M. Dougherty
University of Connecticut

Jennie Weiner
University of Connecticut

Abstract: Using data from Rhode Island, and deploying a fuzzy regression discontinuity design, this study capitalizes on a natural experiment in which schools, in accordance with the No Child Left Behind (NCLB) waivers, were sorted into performance categories based on a continuous performance measure. The lowest performing schools were then mandated to implement interventions, with more interventions required in the lowest performing schools. We find that schools required to implement few interventions performed no differently relative to schools that had no interventions required. Among lower-performing schools, those required to adopt more interventions did worse than schools mandated to implement fewer, including higher student mobility. This suggests larger negative effects for schools receiving the most negative labels, and the potential need to better understand what leads to these worsening outcomes in response to accountability requirements.

Running Head: Rhode to Turnaround

The Rhode to Turnaround: The Impact of Waivers to No Child Left Behind on School Performance

Introduction

In December, 2015 the US government passed the long awaited reauthorization to the Elementary and Secondary Education Act (ESEA). Titled the Every Student Succeeds Act (ESSA), one of its frequently touted aspects is its orientation towards shifting a large portion of school oversight and control back to the states (Smarick, 2015). In so doing, the law is seen by many as creating opportunities for states to modify, reinvent or reinforce interventions for underperforming schools in ways that will better serve their needs. However, and despite the general enthusiasm for this approach, questions remain regarding exactly how states and/or districts should work with their schools to improve them (Honig et al., 2010; Lee, Louis & Anderson, 2012; McGuinn, 2012). One reason for such questions is that it is still unclear whether and to what degree earlier federally driven intervention models from policies like Race to the Top and the ESEA waivers should be replicated (e.g., Dee, 2012; Stuit, 2012; etc.).

On one hand, there is a substantial body of work that concludes that K-12 accountability systems can have positive effects on high-stakes state test scores, even among schools that are labeled lowest performing and subject to accountability pressure (Carnoy & Loeb, 2002; Chakrabarti 2010; Chiang 2009; Dee & Jacob, 2011; Hanushek and Raymond 2004; Rockoff & Turner, 2010; Rouse et al., 2013; Winters & Cowen, 2012). More recent research by Papay (2015) and others (Dee, 2012; De la Torre et al., 2013; Player & Katz, 2013; Strunk, Marsh, Hashim, Mecenas-Bush, Weinstein, forthcoming; Zimmer, Henry, & Kho, 2016) suggests that even more recent turnaround policies spurred by RTTT may also serve to enhance student performance. While these studies appear to build a positive case for such an approach, they do not necessarily

provide insight into why or how these effects are achieved. For example, it could be that a well-developed state and district accountability or reform models were instrumental in these schools success (e.g., in Massachusetts). Alternatively, it may be that interventions selected to improve outcomes were implemented with greater fidelity or in more fruitful combinations. In any case, the potential for variability in the supports chosen and their implementation suggest a need for further research to assess generalizability.

Reinforcing this need for additional research is work by Stuit (2012) and others who suggest that turnaround efforts have been lackluster at best. Some have theorized that one reason for these more mixed results is the poor construction of the interventions themselves (Duke, 2012; McQuillan & Salomon -Fernandez, 2008), their tendency to take a one-size fits all approach (Finn & Winkler, 2010) and that they focus solely on educator's behaviors rather than systemic issues like intergenerational poverty or racism (Peck & Reitzug, 2012). Others have pointed to a lack of coherence regarding reform interventions at the state or district level (Hemmer, Madsen & Torres, 2012) and a need for greater capacity among these entities to support these interventions adequately at the school level (Bottoms & Schmidt-Davis, 2010; Mendels & Mitgang, 2013). If such observations are correct then returning to ESSA (or a similarly decentralized system), a move from greater federal to state oversight, may serve to exacerbate rather than mitigate these issues. Such considerations again point to the need for additional work assessing the impact of federal intervention policies before generating new localized policies and is the focus of the current study.

In this paper we focus on Rhode Island, which, as part of its ESEA waiver application, identified not only which schools were lower performing, but also how many and which interventions an underperforming school or district would implement. This approach to reform, combined with the fact that Rhode Island was one of few states to maintain its state testing regime

after waiver approval, provides a novel opportunity to further enhance our understanding of the impact of accountability policy when schools, identified as underperforming, must participate in a set of mandated interventions.

Using student- and school-level data from Rhode Island, we compare the subsequent performance of students in schools just below and above the thresholds used to mandate differing amounts of prescriptive interventions to estimate the causal impact of these mandates on student learning and mobility. Deploying a fuzzy regression-discontinuity design, we model the effects of a school being just subject to the requirements associated with falling into a performance tier that required intervention, relative to schools in the tier just above it. Importantly, we are able to differentiate the potentially different effects of being in the lowest two tiers of schools identified as underperforming. Using this method, we assess the impact of being required to implement different amounts of prescriptive school-improvement requirements on the subsequent educational outcomes of students in schools on the margin of these requirements. We also add descriptive data on the type of interventions elected by schools in different performance levels to try and gain some insight into the potential mechanisms of our results.

We find that, in the first two years of the policy, schools with the fewest mandated interventions (i.e., the highest performing of the low performers) do no differently in math or reading than their peer schools that just avoided being required to implement these mandates. Additionally, we find no evidence that being identified as needing intervention causes greater student mobility out of these schools. Conversely, we find suggestive evidence that schools in the lower two performance tiers – those schools that face more and the most restrictive interventions – see lower subsequent performance relative to schools with fewer mandated interventions, and likely face higher rates of student mobility than similar schools with slightly stronger performance.

These results suggests that more interventions might not always be better and may have unintended consequences that impact a school's long term ability to improve.

Background

Background on ESEA Waivers and their Impact

In 2009, President Obama and the Secretary of Education, Arne Duncan, announced a new federal initiative called Race to the Top (RTTT). Meant to "spur systemic reform to improve teaching and learning in America's schools," (U.S. Department of Education) the initiative provides states and, more recently, districts the opportunity to submit comprehensive reform plans to compete for funding. The initiative has been far-reaching and as of 2013, 19 states had received over $4 billion to implement their plans, including the State of Rhode Island.

One of the key features of RTTT has been a focus on chronically low performing schools and initiatives aimed at reforming them. Specifically, to be awarded RTTT funds, Local Education Agencies needed to engage their lowest performing schools in implementing one of four intervention models:

- *Turnaround:* the school is completely redesigned, including replacing the principal and at least half the staff or the school is replaced with a new school or schools.

- *Restart Model*: The school is either converted to a charter school or replaced with a charter that serves the same students.

- *Transformation*: The school engages in a number of redesign features including a rigorous new teacher evaluation system though can retain all staff other than principal, who must be replaced.

- *School closure*: Close the school and enroll the students who attended the school in higher achieving schools.

As Rhode Island and other states and districts that received RTTT funds began grappling with implementation, the deadlock regarding reauthorizing No Child Left Behind persisted. This left a void in terms of federal accountability policy; a void that the DOE filled by encouraging state education agencies to apply for waivers to some of the specific requirements of the law. Unlike former waivers associated with NCLB, these new waivers would be part of a competitive process contingent on states adopting a variety of reform-oriented policies. Among these were instituting college and career ready standards as well as creating a system of interventions for the state's lowest performing schools. Thirty-nine states, including Rhode Island, applied and received such waivers in 2012.

Research on the impact of the interventions associated with RTTT and now ESEA waivers is decidedly mixed. A number of researchers have found little if any impact of current turnaround efforts (Ladd & Heissel, 2016; Stuit 2010; 2012, etc.), and have gone as far as to argue that such efforts may even serve to exacerbate current inequities (Trujillo, 2012). Alternatively, a set of other studies finds evidence of positive impacts of intervention efforts when the interventions or requirements are more severe. For instance, research by Schueler (2015) suggests that turnaround interventions can be successful. Their study of district-wide turnaround via state takeover in Lawrence, Massachusetts found that, in the first two years, these efforts produced large gains in math achievement and modest positive gains in ELA, particularly for ESL students. Yet, the authors questioned whether such gains will be sustainable over time and the degree to which the mechanism of state takeover combined with the tightly constructed and coherent portfolio model may have contributed to these positive effects. Similarly, work by Ruble (2015) in New Orleans where external contractors took over traditional public schools, showed that such efforts produced positive student achievement. Stunk and colleagues (forthcoming) also found positive effects on

reading test scores (but not math) in Los Angeles in response to reconstitution of the teaching force in the lowest-performing schools in the district. And Zimmer, Henry, and Kho (2016) generated evidence that turnaround efforts in Tennessee produced positive effects on student test scores in settings that allowed for flexibility in governance and innovative approaches to school improvement. However, in each of these cases of improved outcomes in the face of mandated interventions, and consistent with prior work from the Chicago Public Schools (De la Torre, Allensworth, Jagesic, Sebastian, Salmonowicz, Meyers, & Gerdeman, 2013), the success of the interventions occurred concurrently with a substantial overhaul to how the schools in the study were structured and operated, creating some questions regarding the generalizability of the results.

Shifting now to the impact of these policy reform efforts in more traditional public school settings, such as those found in Rhode Island, recent work by Papay (2015), also in Massachusetts, found that, over a period four years, schools across the state that had participated in turnaround practices (i.e., what the state deemed level 4 schools) dramatically improve student achievement across the board, closed achievement gaps and, positively changed their overall performance trajectories. However, the author points out that these findings far exceed the results from other state's efforts and thus there may be something about the Massachusetts context in particular (e.g., its long reform history and strong infrastructure), that contributed to these findings.

Additionally, Dee (2012) also found more modest positive impacts of school improvement grants (SIG) for schools in California that were among the lowest performance, but not those at the just higher threshold of performance. Taken together, and highlighted by Papay (2015), this research suggests a need for additional research on the impact of these types of intervention models across contexts, including an eye to the specific mechanisms that might be influencing these potentially varied results.

School Intervention in Rhode Island

After successfully competing for RTTT funds, in 2012 Rhode Island was awarded an ESEA waiver. Within the waiver, Rhode Island introduced a six-tiered classification system for schools from lowest to highest performing (i.e., priority, focus, warning, typical, commended, and leading). A school's classification was derived from overall student proficiency and distinction, participation on state assessments, gap closures, progress towards predetermined 2017 targets, and general student progress[1]. Based on these elements all schools in the state were given a score and fell along a continuum of performance. Schools are identified for a particular performance category based on cutoffs designated on an overall measure of performance, the Composite Index Score (CIS). While the cutoffs for thresholds to determine categories were designated on the CIS, schools could also fall into these categories by receiving low scores on individual elements of the total score. The total score was comprised of: proficiency levels, performance within subgroups, student performance growth, and participation rates on the testing regime, as well as high-school graduation rate (applicable only to high schools) (See Figure 1 to observe the relationship between the CIS and school classification.

Once classified, the lowest three tiers of schools (i.e., priority, focus, and warning) were required to implement targeted interventions,[2] the nature of these interventions closely mirrored those outlined in RTTT and included options for closure and restart. However, as true across the U.S. in relation to RTTT (Author & Author, 2016), few, if any, of R.I.'s lowest performing schools were closed or converted into charter schools. Rather most schools identified as low performing

---

[1] High schools are also held accountable for graduation rates.
[2] In their RTTT application RI committed to having all schools implement what they name their 3 core improvement strategies: common core, a new educator evaluation system and data-based decision-making. These interventions were reaffirmed during the waiver process.

engaged in what the state called the Flex model which was closely modeled after school transformation under RTTT.

All schools choosing the Flex option were required to implement the state's core improvement strategies as outlined in RTTT, and the state-mandated interventions presented in the states' flex menu (Appendix A). In particular, the state's lowest performing schools (i.e., priority schools) were required to implement four Tier I strategies (i.e., those directly modifying the school's leadership, infrastructure, content and professional support system) and 2 Tier II strategies (i.e., a mixture of lighter touch often programmatic strategies) while focus schools needed to implement two Tier I strategies and 2 Tier II strategies. Warning schools, those closest to the performance threshold excluding them from intervention were required only to implement four interventions in total though these could be drawn from sources outside the flex menu as long as they were an "empirically-proven strategy of equal intensity" (RI DOE).

In our analysis, we focused on the margin of being identified in two of the three tiers of schools (i.e., warning and focus) that are required to undergo interventions as a result. First, we explore how schools, just identified as being in the warning category and hence mandated to implement interventions, performed compared to peer schools just identified as typical and hence not mandated to implement interventions. Of perhaps greater policy interest is the impact of being just labeled as a focus school (i.e., the category representing the second to lowest performance tier) relative to peer schools in the higher performing warning status as focus schools were required to implement far more interventions than warning schools and were monitored more closely by the state as they did.[3]

---

[3] The classification of rules for priority schools – those in the lowest performance category – do not permit us to include them in our analyses. Specifically, the identification of these schools did not rely on a fixed cutoff score like those that allow for strong identification of causal effects as for the other categories, and assumptions necessary to support other quasi-experimental designs also fail when trying to identify such effects in Rhode Island.

Methods

Through a research partnership with the Rhode Island Department of Education, we link publically available classification and accountability data with student-level data to provide one of the first estimates of the causal impact of such policies for schools on the margin of being identified in these different tiers. Using data from the first two years of the implementation of the new classification system (as well as the two years prior), we examine the policy impact on several relevant student outcomes across two discontinuities in the CIS (i.e., performance-level cutoffs) used to assign schools to performance categories and mandated interventions. We measure the initial impact of a school experiencing the interventions on its resultant state standardized test scores in math and reading as well as measures of student mobility out of the lowest-performing schools. We analyze schools near the margins of two cutoffs, the first between typical and warning schools where warning schools has to implement a few interventions, and the margin between warning and schools labeled as focus and priority that must implement interventions in greater number and intensity. In the former set of analyses, the typical schools constitute a true control in that they face no required interventions, while in the latter analyses we estimate the difference in subsequent outcomes between schools that face lighter requirements relative to those that face more substantial required changes.

Data & Sample

Our dataset was built using Rhode Island administrative data from the school years spanning 2010-2011 through 2013-2014. Data span the two years prior to a waiver being granted into the first two years under the waiver where some schools had been identified as failing and therefore subject to mandatory intervention. Our sample includes data from 274 schools for which we have student-level records. In our first set of analyses we focus on students in schools and

grades that were near the cutoffs to determine whether a school's performance index placed them in the warning category versus just above that cutoff and in the typical category. In our preferred specification there are 110 schools with over 15,000 students enrolled each year. Our latter analysis on the lowest performing focus and priority schools relative to warning schools includes 42 schools with over 11,000 students in each year. Our sample focuses on students in schools that serve grades three through eight and is restricted to those who can be observed across the three focal years of analysis, the year in which a school's performance was identified, and the two subsequent years of test scores.

In Table 1 we show descriptive characteristics of the students in the schools across all performance categories in the baseline year of the policy 2012. This table clearly illustrates the differences in average prior student performance across categories which is expected since performance levels, growth, and differences by subgroup comprise the underlying measure used to make assignment to these categories. This is further illustrated in Figure 1 where performance category is on the vertical axis and the continuous composite measure of school performance, the CIS, on the horizontal access. The dashed vertical line denotes the cutoff used to mandate interventions between typical and warning schools. The delineation between warning and focus and priority schools is made based on a subscore of the CIS, the ten points apportioned to proficiency level, a fact whose use we explain further below. Any schools that fall into the bottom three categories but have scores above the 50-point CIS threshold were identified on subscores (components of the overall CIS). These subscores are related to proficiency levels, performance within student subgroups, student performance growth, and participation rates.

Measures & Outcomes

Running Head: Rhode to Turnaround

   The accountability measures under NCLB and the subsequent waivers were all designed

to improve student learning as measured by required state test scores. Pursuant to these goals, our

primary outcomes of interest are student scores on the math (*MATH*) and reading (*READ*) New

England Common Assessment Program (NECAP), the test mandated for accountability purposes

by Rhode Island. We also defined two binary measures of mobility in response to the policy. First,

we defined whether students made out-of-sequence school changes (*MOVE*), to document students

changing schools at unanticipated times in response to the policy.[4] We hypothesize that higher-

ability students are more likely to leave schools with lower performance levels. This would likely

result in differentially lower test scores in those lower-category schools, but might also constrain

the variance by truncating the top portion of the distribution (Chiang, 2009; Figlio & Rouse, 2006).

We also define a measure of attrition (*ATTRITE*) from the data to indicate whether a student leaves

the data set entirely between the baseline policy year and one of the two subsequent years. This

outcome is used to satisfy the methodological requirements (as required by Schochet et al., 2010)

for our analytical approach, and also to explore whether movement out of public schools was

differential on the margin of a school's classification.

Analytic Strategy

   The exogenous choice of cut-points to identify the performance categories, allows for a

high-quality evaluation of the impact of this policy on student educational outcomes. Using a fuzzy

regression-discontinuity design (Imbens & Lemieux, 2008), we model the effects of being just

below the performance thresholds that carry prescriptive school-improvement requirements on the

---

[4] In earlier drafts we defined school-by-grade measures of variation in test score outcomes as additional outcomes that might capture changes in the distribution of scores in response to mobility (see Chiang, 2009, or Figlio & Rouse 2006 for similar approaches). These analyses provided imprecise evidence and so we preferred the individual-level measures as outcomes.

subsequent educational outcomes of students in schools relative to students in schools that just missed being identified as needing these interventions. We follow a modeling strategy similar to those used in several recent papers (Chiang 2009; Papay, Murnane, & Willett 2011; Rockoff & Turner 2010; Rouse, Hannaway, Goldhaber, & Figlio 2013; Winters & Cowen 2012). The treatment is fuzzy in this case because the policy requires that any school below a certain threshold adopt interventions (where aspects of participation differs by the exact performance category), but exceptions to the overall assignment rule allow schools with higher total score, but low sub-scores to be subject to intervention. We estimate reduced-from, intent-to-treat effects, of being below aggregate score thresholds to compare schools just facing mandated interventions compared to those who just missed such a requirement. In conjunction with a rectangular kernel and multiple choices of bandwidth (including the optimal bandwidth suggested by Imbens and Kalyanaraman, 2012) we estimate these effects using local-linear regression and specifying the following reduced-from statistical model:

$$Y_{isg} = \alpha_0 + \alpha_1 BELOW_{sg} + \alpha_2 f(RV)_{sg} + \alpha_3 BELOW \times RV_{sg} + \boldsymbol{X'_{isd}\beta} + \gamma_g + \varepsilon_{isg}$$

In this model, Y_{isgt} is our generic outcome for student i in school s in grade g, RV_{sg} is our running or forcing variable – accountability score or subscore - used to classify school performance, $BELOW_{sg}$ is a binary indicator of falling below the threshold that identifies schools as requiring intervention, $BELOW \times RV_{sg}$ is an interaction term that allows the relationship between the forcing variable and outcome to differ for treated and non-treated schools, $\boldsymbol{X'_{isg}}$ is a vector of student-level covariates to improve statistical precision (As shown in the appendix, the core results are not impacted when including or excluding covariates, though including them improves statistical precisions), γ_g is a set of grade-level fixed effects, and ε_{isg} is the error term clustered at

the school level since the forcing variable is continuous (Lee & Card, 2008). The coefficient of interest is α_1 which represents the population causal effect of being just required to implement mandatory interventions as a function of the continuous measure of school performance relative to students in similar schools that just missed this designation.[5] We demonstrate that our data satisfy the requirements of a regression-discontinuity (RD) design as suggested by Schochet et al. (2010). As required for an RD design to be valid, we demonstrate that schools on this margin, but opposite sides of the cutoff, were equivalent on observable dimensions and therefore (presumably) equal on unobservable dimensions as well. Our identification of the effects of mandated interventions further relies on schools inability to manipulate their position relative to the performance category cutoffs, making this restriction to our sample necessary to preserve internal validity while sacrificing external validity.

Results

Validity of Research Design

To establish the validity of our RD design we began by testing that manipulation of the position of a school relative to the cutoff established for mandated intervention by either the schools or the state did not occur. In Figure 2, we present two histograms. In Panel A we present CIS performance scores in bins of width one with a dashed line overlaid representing the threshold used to identify warning schools. We complement this with Panel B that shows the distribution of the proficiency sub-score used to distinguish schools between the warning and focus/priority performance categories. Though the distribution below the cutoff is thinner for the focus/ priority schools, because the distributions just on one side of the cutoff or the other seem to differ

---

[5] We test the sensitivity of our estimates to our choice of bandwidth and the use of a linear specification of the forcing variable. Our results are robust to different specifications and are available upon request.

continuously, we argue that these figures provide evidence that the position of schools with respect to the cutoff have not been manipulated.[6]

We also present in Table 2 evidence that there is no difference in the observable characteristics of students in schools on either side of the cutoff. Estimates in Table 2 show the results of fitting equation 1 with the listed covariates in the column headings as the outcome. We find no evidence of meaningful differences between schools at the threshold and thus conclude that our research design is valid and can provide potentially unbiased estimates of the effect of the mandated interventions under NCLB waivers at these two distinct thresholds. There is one statistically-significant difference at the five-percent level of significance, which in 18 tests could be attributed to Type I error. We complement Table 2 with Figures A1 and A2, which demonstrate covariate balance graphically and provide visual evidence that using a linear model to estimates these potential differences is appropriate.

To demonstrate that the variables used to assign schools to performance categories really do create a discontinuity in the probability of having to implement mandated interventions we present Figure 3. In the two panels of this figure we demonstrate that falling just below the threshold used to differentiate performance categories creates a large jump that a school is identified as falling into the lower-performing category. It is precisely this exogenous variation in classification which identifies the effects of being mandated to implement interventions.

Reduced-form Estimates

---

[6] In fact, because the accountability scores are multi-faceted it would be impossible for schools to manipulate their position relative to a previously unknown cutoff. In theory, policymakers could have, presumably, chosen a cutoff score to manipulate where schools fell, but it does not appear to have occurred in this case.

Our analysis suggests that there is no clear evidence of an impact of being just classified as needing intervention programs on math or reading test scores at the warning threshold, but that having to implement more interventions may negatively impact test performance for students in focus schools. Interestingly, there is no identifiable difference in the mobility of students out of lower-performing schools at the margin of school classification.

Figures 4 and 5 present visual evidence of the effects on test scores and mobility of being in a school that was just categorized as falling into warning and focus, respectively. Evidence of differences in performance, if they exist, should be evident from clear visual differences in the trends on either side of the cutoff at the cutoff itself. Though the visual evidence in Figure 4 does not clearly indicate a difference in average test scores for schools near, but on opposite sides of the warning threshold, there are some suggestive differences in Figure 5 between schools near the focus threshold that suggest that focus schools perform worse than peer schools in the warning category after being mandated to implement interventions, particularly in the second year of the policy implementation. This suggestive difference in performance accords with similar evidence of a potentially increased probability of moving out of a school that was labeled in the focus category relative to warning schools.

Our estimates in Table 3 provide statistical evidence to support the visual evidence that there was no clear effect on test scores or mobility for students in warning schools. However, the small magnitude of these point estimates and their lack of statistical significance in the pre- (2012) and post-policy (2013, 2014) years supports the conclusion of no effect.[7] However, the large standard errors suggest that we cannot strictly rule out potential impacts on test scores. The second

---

[7] Standard errors for these estimates are large enough that we cannot conclusively rule out some potential for a small positive effect, but in that case the effect should be suggested, but is absent, in the visual evidence.

panel of Table 3 demonstrates that there were suggestive negative effects on math scores, clearer

negative effects on reading test scores for students in focus schools that just missed being classified

as warning schools. There is also imprecise, but suggestive evidence that students in these schools

were more likely to change schools relative to similar peers in the warning category.[8] Importantly,

for the internal validity of our comparisons, we do not find evidence of differential attrition from

the sample in either set of estimates based on whether students were enrolled in school receiving

greater or lesser degrees of sanctions.

To ensure that our findings were not driven by our choice of analytic window (bandwidth)

or our preferred linear specification, we both halved and doubled our preferred bandwidth, in

addition to applying the IK optimal bandwidth, and fit both linear and quadratic specifications

(Gelman & Imbens, 2014) of our forcing variables to test whether effects are sensitive to our

assumption of local linearity (Imbens & Lemieux, 2008). The results presented in Table 4 suggest

that our findings are not driven by model specification.[9]

Discussion & Significance

Earlier evidence on the effects of school turnaround are mixed. When turnaround has been

successful it has often occurred in districts in which state takeover occurred (e.g., Lawrence,

Massachusetts; New Orleans, Louisiana) (Ruble, 2015; Schueler, 2015). In California (Dee, 2012)

and Massachusetts (Papay, 2015) there is also some suggestive evidence that states with

infrastructure and capacity to support turnaround efforts, as well as a history of enforcing policy

---

[8] Using the prior year as a placebo test rules out that any perceived policy effects were in fact differences between groups of schools that predated the policy.

[9] In appendix Table A1 we include results without student-level covariates. Though point estimates are somewhat different, the core results are not impacted. In Table A2 we include only students who were in the same schools across all three years of the panel and reach similar conclusions for warning schools – results for Focus schools are similar and are available from the authors upon request.

requirements with fidelity, can generate positive impacts through turnaround efforts. Taken together, such findings reinforce arguments that turnaround requires not just school level interventions but must also impact what (Peurach & Neumerski, 2015) and others (Hopkins & Spillane, 2015; Hopkins, et al., 2013) call the larger "educational infrastructure" (i.e., structures and resources used by educational systems to enhance instructional practice).

In Rhode Island, we find no evidence of a positive impact on student test scores among schools that were just identified as having to undertake mandatory interventions under the state's ESEA waiver. Given that, in this context, where neither has state takeover occurred nor has the state's approach to reform often been highlighted as a model for other states regarding enforcement, such findings reinforce that it may indeed be context and coherence that serve to determine whether school turnaround will be effective. However, although, as we will discuss in more detail later, in Rhode Island we see little evidence of high intensity or potentially high-leverage interventions being chosen among schools required to adopt interventions, it is important to note that we also have relatively little insight into the processes that lead to improved outcomes in New Orleans, Lawrence, or Massachusetts more generally. In all existing papers on this topic, authors have not observed the implementation of these turnaround efforts and so can only speculate as to the mechanisms that produced change. In Rhode Island we face the same reality with little on-the-ground knowledge of how interventions were implemented and how the chosen interventions cohered with other elements of local practice and policy.

That said, as we have already hinted to above, one might conclude that because the success stories in New Orleans, Los Angeles, Tennessee, Lawrence, and Massachusetts appeared contingent on large-scale and what some might consider aggressive implementation of reform mandates, that similar approaches would have been needed in Rhode Island to produce better

outcomes. Indeed it is true that, historically, small attempts in the direction of more stringent approaches have shown little success in Rhode Island. For example, when in 2010, the high school in Central Falls, Rhode Island was initially identified as in need of turnaround under the Race to the Top grant process and all the teachers were fired as a result, over 90 percent of the same individuals were rehired before the start of the next school year, though with some changes to work rules. These rehires were made in part because of union pressure and a sense that firings would be politically untenable (Zezima, 2010). In this context, taking a high-stakes turnaround approach to changing educators was not successful, nor did it seem to adequately match the context in which union power and a close knit community made a large-scale change in staff difficult if not impossible. This may suggest that imposing stronger, but standardized requirements for turnaround (e.g., more rules about firing teachers or principals) will not be sufficient to lead organizational change and improvement.

Furthermore, there is ample research to suggest that coherence of activities is a cornerstone of organizational improvement (See Newmann et al., 2001 for a review). As such, it would not make sense to simply impose a standard set of requirements and expect better outcomes. It is worth noting that among the many selections that schools in Rhode Island made for their mandated interventions, nearly all of them resembled the guidance that the Federal government provided to states who applied for ESEA waivers. This suggests that perhaps Rhode Island, among other states, took more of a compliance based approach to initially implementing Waiver requirements, and gave little though to how, and under what conditions the menu of intervention options might be applicable or appropriate for their schools. Moreover, and adding perhaps to larger questions regarding the degree of coherence within the state's turnaround efforts, for schools required to take on a larger number of mandated interventions, we find evidence that their students fare worse on

standardized measures of reading than their peers in schools that were just required to implement fewer interventions.

In considering why greater number of interventions caused less positive results, a few possibilities arise. First, it may be the case that though intended to empower schools and districts to make localized choices about the types of interventions best suited for the context, they either lacked the capacity or wherewithal to select those likely to catalyze change. For example, there is strong evidence to suggest that schools, and indeed organizations more broadly (Argyris & Schon, 1996), often choose reforms that are more technical in nature focusing on surface level changes rather than engaging in more substantive initiatives focused on capacity building (Author, 2016). Additionally, others studies have explored the impact of prior accountability programs (e.g., NCLB) to sanctions including loss of funding suggest that some improvement may be evident when accountability was binding and required substantial changes in educational organization or practice (Dee 2012; Papay, 2015; Strunk et al, forthcoming; Zimmer, Henry, & Kho, 2016), but this is not entirely consistent across all contexts (Heissel & Ladd 2016).

Connected to this prior point, these negative effects may have more to do with implementation than selection. Though the answers lay outside the scope of this study, schools may have wanted to implement the interventions with fidelity but, like many schools engaged in improvement efforts, they lacked the necessary district and state support to do so effectively (Honig, 2010; 2012). Additionally, given teachers' and school leaders' tendency to feel overwhelmed and to focus on immediate concerns rather than taking a systemic approach to improvement (Hargreaves & Shirley, 2009), the fact that most of these schools served the state's most vulnerable populations, as well as research to suggest that in turnaround efforts, it is better

to focus on fewer more focused interventions (e.g., Hofer, 1980), it is perhaps not a surprise that asking these schools to do more did not produce positive outcomes.

Third, these results may be related to a lack of capacity of the restructuring agencies themselves. In the case of Rhode Island for example, the interventions and flex menu was constructed by the state, not the schools. Moreover, though schools hypothetically were given the opportunity to select some interventions, districts were encouraged to identify district-wide initiatives to which all schools would be held accountable. As such, the negative results associated with more interventions may be a result of a district being less effective at identifying appropriate strategies for reform than the schools themselves. This reality reinforces what many researchers (e.g., Peck & Reitzug 2014) see as a problematic assumption of the policy - that underperforming schools are underperforming in the same way and hence require the same intervention. Such an arrangement was also meant to facilitate the district to provide increased infrastructure and support for the selected interventions. Our findings, therefore may also suggest that district capacity to support these interventions was not optimal during initial stages of intervention. As Honig (2012) points out, weak district oversight would likely increase the potential lag time for improved performance at the school sites.

Regardless of the specific mechanism producing these results, an analysis of the actual choice of interventions by warning schools supports work by Harris et al., 2013 and others (e.g., Klar et. al, 2015; Ovenden-Hope, 2014) that we would not expect changes in outcomes based on the short-term rollout of these interventions. Among the 53 different interventions chosen by the 39 warning schools, all were mandated to implement few total interventions (fewer than 2 on average) and chose interventions that may not be expected to change student performance perceptibly when compared to typical schools. Interventions choices were clustered among

emphasizing cultural competence and its relationship to student attendance and discipline, the implementation of progress monitoring systems to track performance of struggling students, and contracting with an outside vendor to help support building leaders. In addition, ten of the 53 chosen interventions fell into an approved other category making it unclear how to evaluate nearly 20 percent of choices.

Among the nine schools identified in the focus category 52 interventions were chosen. This reflects a clear difference in the total number of interventions required for focus schools relative to warning schools. The three most frequently chosen interventions (each selected by 8 of 9 school) were the establishment of common planning time for teachers, the establishment of a system of peer support for teachers, and a systematic review of course offerings and course alignment. These choices emphasize developing teacher capacity, and while they may not induce immediate changes in student performance, are certainly consistent with longer-term capacity building. That said, such efforts are difficult, often unsustainable, and require a number of school-level resources and supports that may not be available to schools that are chronically underperforming (Thoonen et al. 2012).

On the other hand, even if interventions were harmful due to a lack of fidelity, support or mismatch to need, we might not expect the effects to show up in the very first year of the policy, since there would have been limited time to deploy them and impact student experience (e.g. 2012-2013). This may explain why the suggestively negative effects we see in reading test scores are not pronounced until the second year of the policy. Short-term effects could also be driven by higher ability or more motivated students moving out of low-performing schools, causing average performance to drop mechanically (barring substantial improvement by the remaining students). This story is consistent with the lower test scores, and suggestively higher rates of mobility among

students in focus schools, but lack of precision in the estimates for those students moving out of these schools does not allow for strong inferences. If higher-performing students were more mobile out of focus schools then scores for students in focus schools could go down, relative to warning schools, even if warning scores were unchanged. Our data don't allow us to establish conclusively that our proposed mechanism explains our results, however, it seems plausible and we find no evidence that would directly contradict such a possibility.

Conclusion

As the Every Student Succeeds Act continues to be formalized at the state level, and as the associated systems of accountability emerge, it will be important to monitor how elements of these systems are designed to promote coherence or an acknowledgement of context. The mixed evidence on the effects of SIGs, RTTT, and ESEA Waiver turnaround strategies suggests that state-level decisions about accountability systems under ESSA and their effects on school performance should be a growing area of focus and concern.

This research provides evidence of the short-term impact of these systems and suggests that they may not be incurring the desired results. Understanding the impact of this new accountability system can inform the development and updating of similar systems in other states as well as the requirements established in the eventual NCLB reauthorization. Based on our findings here, it suggests a need to examine and consider some of the undergirding assumptions about reform and the degree to which policy demands align with effective practice. Specifically, it may behoove policymakers to consider whether asking those with less to do more when more is a constrained set of untested choices is likely to be an effective means of enhancing performance. Such considerations may serve to enhance both the quality of discussion happening at all levels of the system and with it, the daily lives of teachers and students impacted by the resultant decisions.

Running Head: Rhode to Turnaround

If not, we may be destined to continue as Duke (2012) has said to "tinker" towards turnaround

rather than create the change so desperately needed in many of the nation's lowest performing

schools.

References

Argyris,C ., & Schon, D. A. (1996). *Organizational learning I: Theory, method, and practice.* Reading, MA: Addison-Wesley Publishing Company

Author 2011

Author 2016

Bottoms, G., & Schmidt-Davis, J. (2010). The three essentials: Improving schools requires district vision, district and state support, and principal leadership. *Southern Regional Education Board (SREB)*.

Carnoy, M., & Loeb, S. (2002). Does external accountability affect student outcomes? A Cross-State Analysis. *Educational Evaluation & Policy Analysis*, *24*(4), 305-331.

Chakrabarti, R. (2010). Program design, incentives, and response: Evidence from educational interventions. *Economic Policy Review (19320426)*, *16*(2), 1-22.

Chiang, H. (2009). How accountability pressure on failing schools affects student achievement. *Journal of Public Economics*, *93*(9/10), 1045-1057.

De la Torre, M., Allensworth, E., Jagesic, S., Sebastian, J., Salmonowicz, M., Meyers, C., & Gerdeman, R. D. (2013). *Turning around low-performing schools in Chicago: Research report*. Consortium on Chicago School Research.

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

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

Duke, D. L. (2012). Tinkering and turnarounds: Understanding the contemporary campaign to improve low-performing schools. *Journal of Education for Students Placed at Risk (JESPAR)*, *17*(1-2), 9-24.

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

Finn C. E., Winkler A. M. (2010). Foreword. In Stuit D. A. (Ed.), *Are bad schools immortal? The scarcity of turnarounds and shutdowns in both charters and district sectors* (pp. 4-9). Washington DC: Thomas B. Fordham Institute.

Gelman, A., & Imbens, G. (2014). *Why high-order polynomials should not be used in regression discontinuity designs* (No. w20405). National Bureau of Economic Research.

Hanushek, E. A., & Raymond, M. F. (2005). Does School Accountability Lead to Improved Student Performance?. *Journal of Policy Analysis & Management*, *24*(2), 297-327.

Hargreaves, A., & Fink, D. (2004). The seven principles of sustainable leadership. *Educational leadership*, *61*(7), 8-13.

Harris, A., Day, C., Hopkins, D., Hadfield, M., Hargreaves, A., & Chapman, C. (2013). *Effective leadership for school improvement*. Routledge.

Heissel, J., & Ladd, H. (2016). The effects of school turnaround in North Carolina: A regression-discontinuity approach. CALDER Working Paper No.156.

Hemmer, L. M., Madsen, J., & Torres, M. S. (2013). Critical analysis of accountability policy in alternative schools: implications for school leaders. *Journal of Educational Administration*, *51*(5), 655-679.

Hofer, C. W. (1980). Turnaround strategies. *Journal of Business Strategy*, *1*(1), 19-31.

Honig, M. I. (2012). District central office leadership as teaching: How central office administrators support principals' development as instructional leaders. *Educational Administration Quarterly, 48*, 733–774.

Honig, M. I., Copland, M. A., Rainey, L., Lorton, J. A., & Newton, M. (2010). Central office transformation for district-wide teaching and learning improvement. *Center for the Study of Teaching and Policy, University of Washington*.

Honig, M. I., & Hatch, T. C. (2004). Crafting coherence: How schools strategically manage multiple, external demands. *Educational Researcher*, *33*(8), 16-30.

Hopkins, M., & Spillane, J. P. (2015). Conceptualizing relations between instructional guidance infrastructure (IGI) and teachers' beliefs about mathematics instruction: Regulative, normative, and cultural-cognitive considerations. *Journal of Educational Change*, *16*(4), 421-450.

Hopkins, M., Spillane, J. P., Jakopovic, P., & Heaton, R. M. (2013). Infrastructure redesign and instructional reform in mathematics. *The Elementary School Journal*, *114*(2), 200-224.

Imbens, G. & Kalyanaraman, K. (2012). Optimal bandwidth choice for the regression discontinuity estimator. *The Review of Economic Studies*, *79*(3), 933-959.

Imbens, G. & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, *142*(2), 615-635.

Klar, H. W., Huggins, K. S., Hammonds, H. L., & Buskey, F. C. (2015). Fostering the capacity for distributed leadership: a post-heroic approach to leading school improvement. *International Journal of Leadership in Education*, (ahead-of-print), 1-27.

Lee, D. S. & Card, D. (2008). Regression discontinuity inference with specification error. *Journal of Econometrics 142(2).* p. 655-674.

Lee, M., Seashore Louis, K., & Anderson, S. (2012). Local education authorities and student learning: the effects of policies and practices. *School Effectiveness and School Improvement*, *23*(2), 133-158.

McGuinn, P. (2012). Stimulating Reform Race to the Top, Competitive Grants and the Obama Education Agenda. *Educational Policy*, *26*(1), 136-159.

McQuillan, P. J., & Salomon-Fernandez, Y. (2008). The impact of state intervention on" underperforming" schools in Massachusetts: Implications for policy and practice. *Education Policy Analysis Archives*, *5*, 18.

Mendels, P., & Mitgang, L. D. (2013). Creating strong principals. *Educational leadership*. 70 (7), 22-29.

Newmann, F. M., Smith, B., Allensworth, E., & Bryk, A. S. (2001). Instructional program coherence: What it is and why it should guide school improvement policy. *Educational evaluation and policy analysis*, *23*(4), 297-321.

Ovenden-Hope, T. (2014). Leading schools in challenging circumstances. *Journal of Education for Teaching*, (ahead-of-print), 1-3.

Papay, J. (2015). The Effects of School Turnaround Strategies in Massachusetts, Association for Public Policy Analysis and Management Fall Conference.

Papay, Murnane & Willett (2011). Extending the Regression-Discontinuity Approach to Multiple Assignment Variables. Journal of Econometrics, 161(2): 203-207.

Peck, C., & Reitzug, U. C. (2012). How existing business management concepts become school leadership fashions. *Educational administration quarterly*, *48*(2), 347-381.

Peck, C., & Reitzug, U. C. (2014). School turnaround fever: the paradoxes of a historical practice promoted as a new reform. *Urban Education*, *49*(1), 8-38.

Peurach, D. J., & Neumerski, C. M. (2015). Mixing metaphors: Building infrastructure for large scale school turnaround. *Journal of Educational Change*, *16*(4), 379-420.

Rockoff, J., & Turner, L. (2010). "Short-Run Impacts of Accountability on School Quality." *American Economic Journal: Economic Policy*, 2(4): 119-47.

Rouse, C. E., Hannaway, J., Goldhaber, D., & Figlio, D. (2013). Feeling the Florida Heat? How low-performing schools respond to voucher and accountability pressure. *American Economic Journal: Economic Policy*, *5*(2), 251-281.

Ruble, W. (2015). The Effect of Contracting Out Low Performing Schools on Student Performance, Association for Public Policy Analysis and Management Fall Conference.

Schochet, Peter, Thomas Cook, Jonathan Deke, Guido Imbens, J.R. Lockwood, Jack Porter, and Jeffrey Smith. 2010. *Standards for regression-discontinuity designs*. Available http://ies.ed.gov/ncee/wwc/pdf/wwc_rd.pdf. Accessed 4 December 2013.

Schueler, E., Goodman, J., Deming, D. (2015)Can States Take Over and Turn Around School Districts? Evidence from Lawrence, Massachusetts, Association for Public Policy Analysis and Management Fall Conference.

Smarick, Andy. 2015. States v. districts in the every student succeeds act. *Education Next*. http://educationnext.org/states-v-districts-in-the-every-student-succeeds-act/

Strunk, K.O., Marsh, J. A., Hashim, A. K., Mecenas-Bush, S. C. & Weinsten, T. (forthcoming). The impact of turnaround reform on student achievement: Evidence and insights from the Los Angeles Unified School District. *Education Finance and Policy*.

Stuit, D. A. (2010). Are bad schools immortal?: The scarcity of turnarounds and shutdowns in both charter and district sectors. *Thomas B. Fordham Institute*.

Stuit, D. (2012). Turnaround and closure rates in the charter and district sectors. *Journal of Education for Students Placed at Risk (JESPAR)*, *17*(1-2), 40-54.

Trujillo, T. (2012). The paradoxical logic of school turnarounds: A Catch-22.*Teachers College Record*.

Thoonen, E. E., Sleegers, P. J., Oort, F. J., & Peetsma, T. T. (2012). Building school-wide capacity for improvement: the role of leadership, school organizational conditions, and teacher factors. *School effectiveness and school improvement*, *23*(4), 441-460.

Winters, M. A., & Cowen, J. (2012). Grading New York: Accountability and student proficiency in America's largest school district. *Educational Evaluation and Policy Analysis*, *34*(3), 313-327.

Zimmer, R., Henry, G., Kho, A. (2016). The Role of Governance in School Turnaround Policies: The Case Tennessee's Achievement School District, Association for Education Finance and Policy Annual Conference.

Figures & Tables

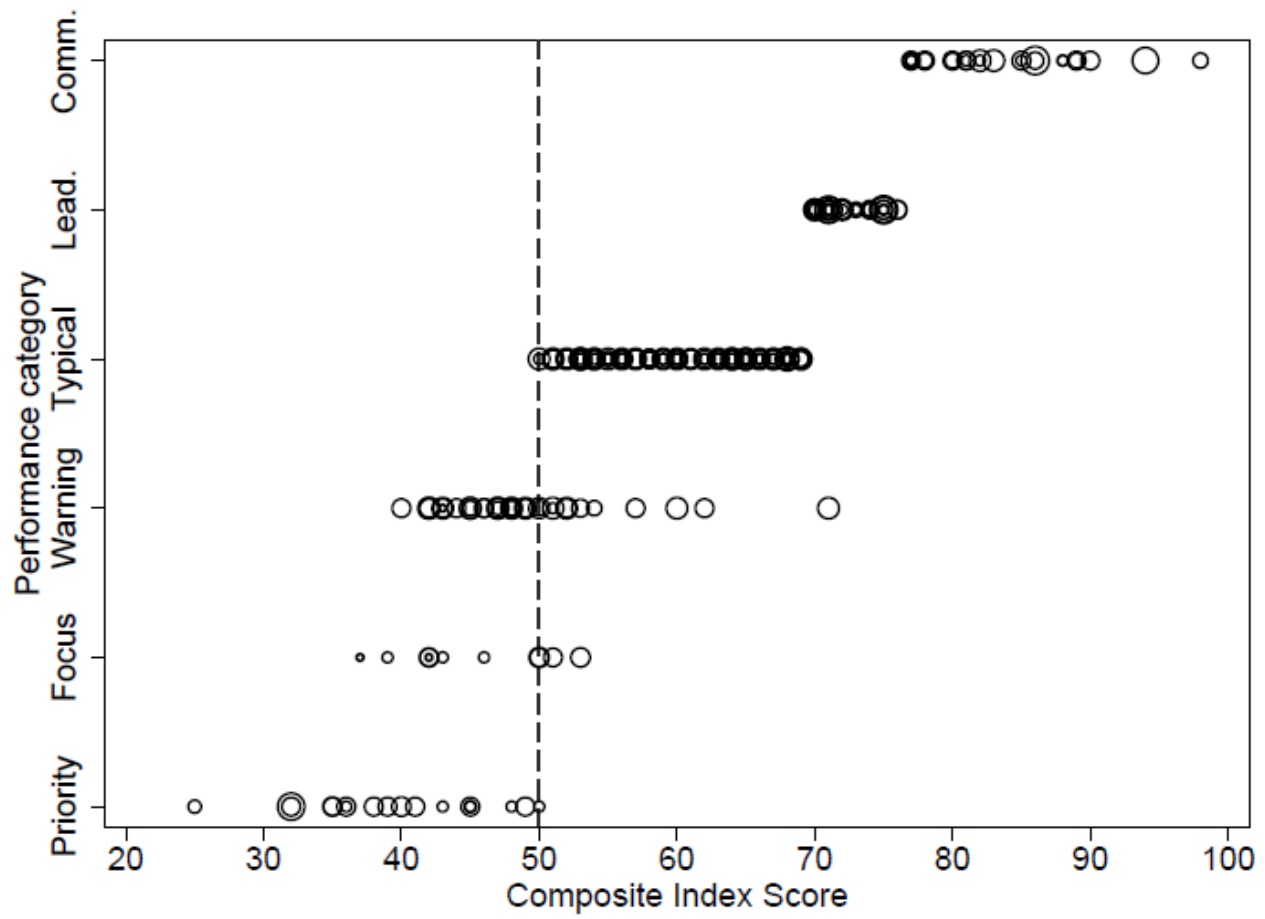Figure 1: Assignment to Performance Category as a Function of Composite Index Score

Running Head: Rhode to Turnaround

Figure 2: Distribution of Forcing Variable around Warning and Focus Cutoffs

(A) Warning

(B) Focus

Running Head: Rhode to Turnaround
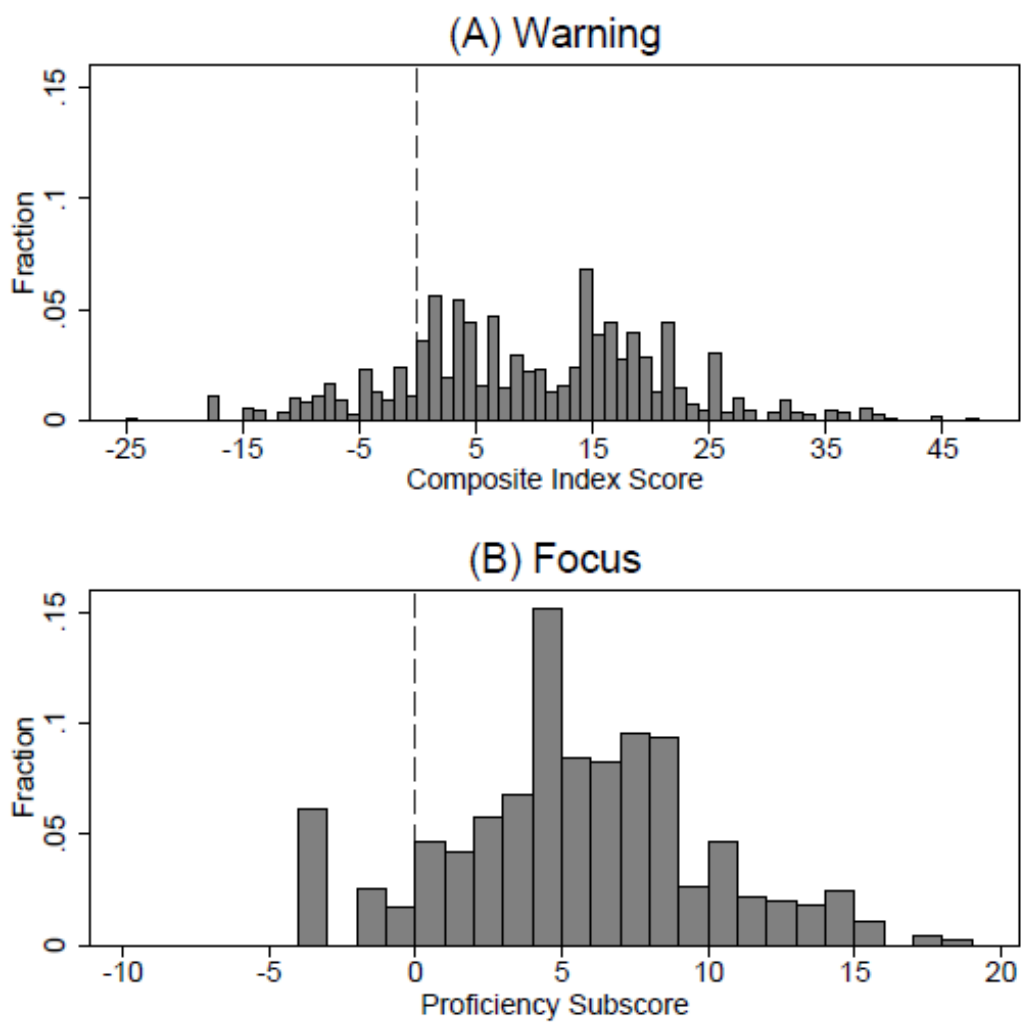
Figure 3: Fuzzy Eligibility for Interventions at Warning and Focus Levels

Running Head: Rhode to Turnaround
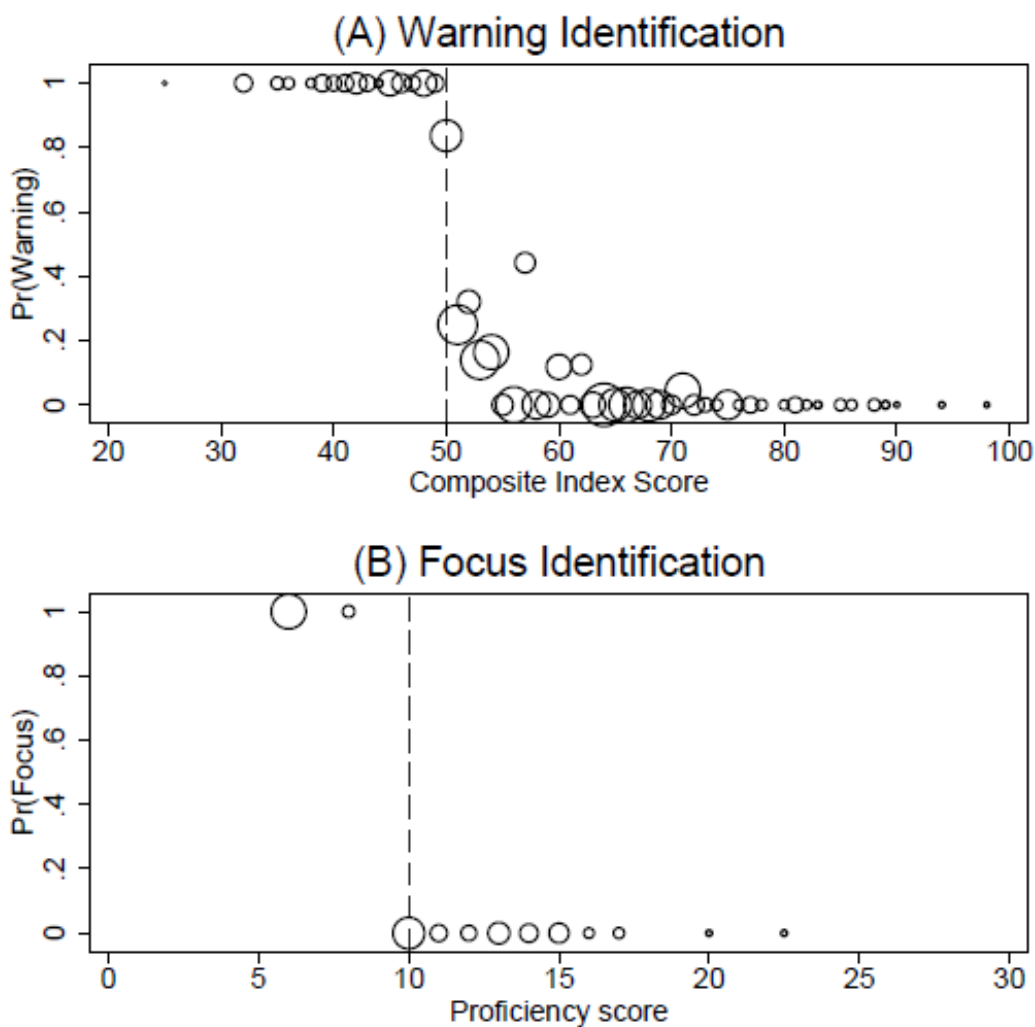
Figure 4: Reduced Form Impact of Warning Status on Mean Scores and Mobility

Running Head: Rhode to Turnaround
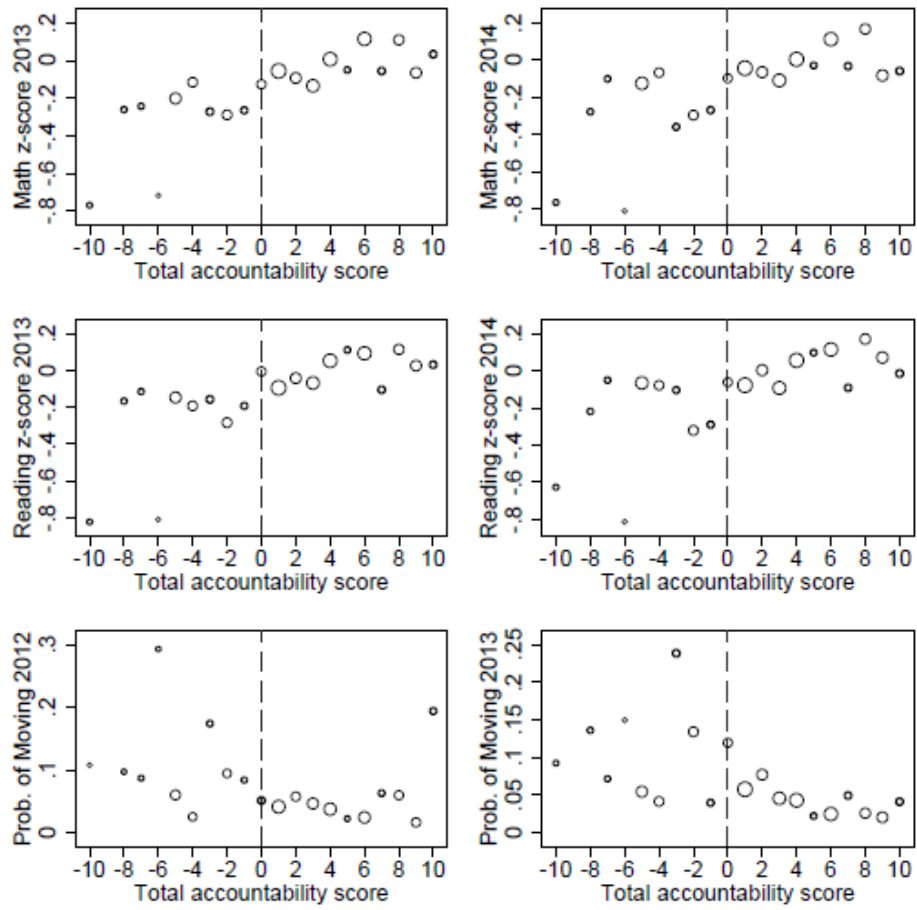
Figure 5: Reduced Form Impact of Focus Status on Mean Scores and Mobility

Running Head: Rhode to Turnaround
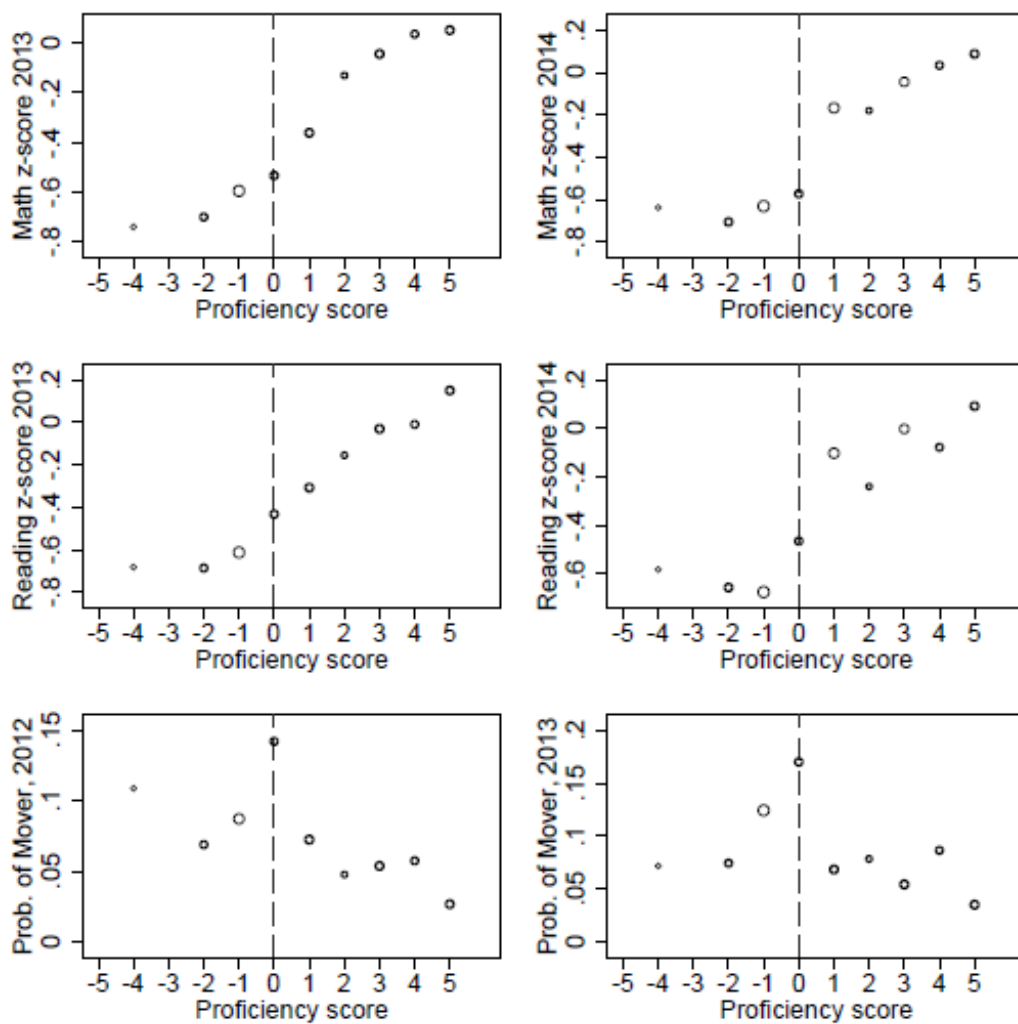
Table 1: Summary Statistics

	(1)	(2)	(3)	(4)	(5)	(6)
			School Classification			
	Priority	Focus	Warning	Typical	Leading	Commended
Number of Schools	18	9	40	136	33	23
Accountability Points 2012	38.5	46.9	49.1	60.6	73	84.6
Male	0.517	0.545	0.512	0.518	0.5	0.5
Asian	0.038	0.041	0.034	0.027	0.023	0.023
Black	0.164	0.178	0.106	0.063	0.096	0.021
Latino/a	0.591	0.617	0.263	0.147	0.164	0.067
Native American	0.007	0.01	0.007	0.006	0.007	0.007
Low Income	0.862	0.876	0.611	0.425	0.388	0.209
Special Education	0.297	0.189	0.153	0.148	0.117	0.125
ELL	0.225	0.257	0.112	0.047	0.045	0.015
Total Days Enrolled 2012	166.8	164.7	173.2	174.9	175.8	176.4
Baseline Math	-0.964	-0.716	-0.151	0.06	0.197	0.458
Baseline Reading	-0.939	-0.704	-0.118	0.048	0.167	0.411
Baseline Writing	-0.739	-0.37	-0.171	0.021	0.188	0.33

Notes: Mean values of key variables are shown by school performance category in 2011-2012.

Table 2: Covariate Balance Test

	(1) Male	(2) White	(3) Black	(4) Latino	(5) Prior	(6) Prior Math	(7) Poor Reading	(8) Sp. ed.	(9) LEP
Warning	0.012	-0.027	-0.017	0.051	-0.040	-0.166	0.054	0.038	0.080
	(0.016)	(0.125)	(0.054)	(0.086)	(0.130)	(0.147)	(0.103)	(0.023)	(0.092)
μ	0.570	0.567	0.130	0.298	-0.163	-0.153	0.650	0.207	0.159
N	298	298	298	298	296	296	298	298	298
Focus	0.037	0.006	-0.065*	0.061	-0.121	-0.159	-0.062	0.083**	-0.038
	(0.025)	(0.112)	(0.032)	(0.136)	(0.125)	(0.114)	(0.060)	(0.032)	(0.095)
$\mu2$	0.576	0.295	0.214	0.504	-0.528	-0.513	0.860	0.236	0.226
N	129	129	129	129	126	126	129	129	129

Notes: Heteroskedasticity robust standard errors clustered by school are in parentheses (* $p<.10$ ** $p<.05$ *** $p<.01$). The coefficients shown are generated by OLS using a linear specification of the forcing variable and a bandwidth of 10 CIS points for schools near the Warning threshold and 5 CIS points for those near the Focus threshold. Each regression also includes grade fixed effects. Also listed is the mean of the covariate for students in schools within 3 CIS points of the respective threshold.

Running Head: Rhode to Turnaround

Table 3: Reduced Form Effects in First Two Years of Policy and One Year Prior, for Warning and Focus Schools

	(1) Math Z-score	(2) Read Z-score	(3) Mover	(4)
Panel A - Effects of Warning Status Eligibility				
Warning Effect, 2012	-0.050	-0.060	-0.002	0.010
	(0.050)	(0.047)	(0.021)	(0.013)
N	18,178	18,178	18,178	18,178
Warning Effect, 2013	0.062	0.041	-0.051	-0.002
	(0.056)	(0.058)	(0.061)	(0.007)
N	17,275	17,275	17,275	17,275
Warning Effect, 2014	0.026	0.038	0.000	0.000
	(0.067)	(0.062)	(.)	(.)
μ	-0.22	-0.21	0.07	0.04
N	17,174	17,174	17,174	17,174
Panel B - Effects of Focus Status Eligibility				
Focus Eligibility 2012	0.011	-0.119	0.032	-0.024
	(0.067)	(0.088)	(0.065)	(0.016)
N	7,199	7,199	7,199	7,199
Focus Eligibility 2013	-0.088	-0.166**	0.073	-0.018
	(0.081)	(0.065)	(0.083)	(0.012)
N	6,372	6,372	6,372	6,372
Focus Eligibility 2014	-0.058	-0.245***	0.000	0.000
	(0.043)	(0.029)	(.)	(.)
μ	-0.62	-0.65	0.00	0.00
N	6,074	6,074	6,074	6,074

Notes: Heteroskedasticity robust standard errors clustered by school are in parentheses (* $p<.10$ ** $p<.05$ *** $p<.01$). Estimates show the impact of being in a school assigned to receive interventions in response to low performance on various measures of student test scores. The coefficients shown are generated by OLS with local linear regression and a triangular bandwidth of 5 CIS points. Covariates are included in all models as are grade fixed effects. In the final row is the mean of the outcome variable among students just below the threshold.

Table 4: Robustness Checks at Focus Cutoff

	(1) Math Z-score	(2) Read Z-score	(3) Mover	(4) Attrite
Warning Margin				
Linear, BW=IK	0.111	0.093	-0.315***	-0.019**
	(0.100)	(0.111)	(0.081)	(0.009)
Quadratic, BW=IK	0.136	0.100	-0.405***	-0.031*
	(0.157)	(0.179)	(0.078)	(0.016)
Linear, BW=10	0.062	0.041	-0.051	-0.002
	(0.056)	(0.058)	(0.061)	(0.007)
Quadratic, BW=10	0.089	0.096	-0.178**	-0.009
	(0.092)	(0.103)	(0.080)	(0.009)
Linear, BW=15	0.062	0.041	-0.019	0.001
	(0.051)	(0.052)	(0.056)	(0.007)
Quadratic, BW=15	0.071	0.045	-0.108	-0.012
	(0.082)	(0.091)	(0.072)	(0.009)
Focus Margin				
Linear, BW=IK	-0.078	-0.150**	0.069	-0.020*
	(0.081)	(0.066)	(0.086)	(0.011)
Quadratic, BW=IK	-0.009	-0.156*	-0.085	-0.015
	(0.080)	(0.082)	(0.102)	(0.011)
Linear, BW=10	-0.025	-0.138**	0.084	-0.020*
	(0.097)	(0.062)	(0.082)	(0.010)
Quadratic, BW=10	-0.080	-0.146**	-0.041	-0.017*
	(0.065)	(0.066)	(0.090)	(0.009)
Linear, BW=15	-0.037	-0.138**	0.091	-0.019*
	(0.093)	(0.060)	(0.081)	(0.010)
Quadratic, BW=15	-0.033	-0.140**	-0.033	-0.018**
	(0.064)	(0.063)	(0.085)	(0.008)

Notes: Heteroskedasticity robust standard errors clustered by school are in parentheses (* $p<.10$ ** $p<.05$ *** $p<.01$). Estimates are of the impact of mandated interventions on student outcomes for students in schools near the cutoff mandating interventions. The coefficients in each row are generated using a rectangular kernel, OLS with the listed functional form for the forcing variable, and specified bandwidth. Specifications include fixed effects for grade but do not include controls unless noted.

Running Head: Rhode to Turnaround

Appendix:

Figure A.1: Covariate Continuity Between Typical and Warning Schools
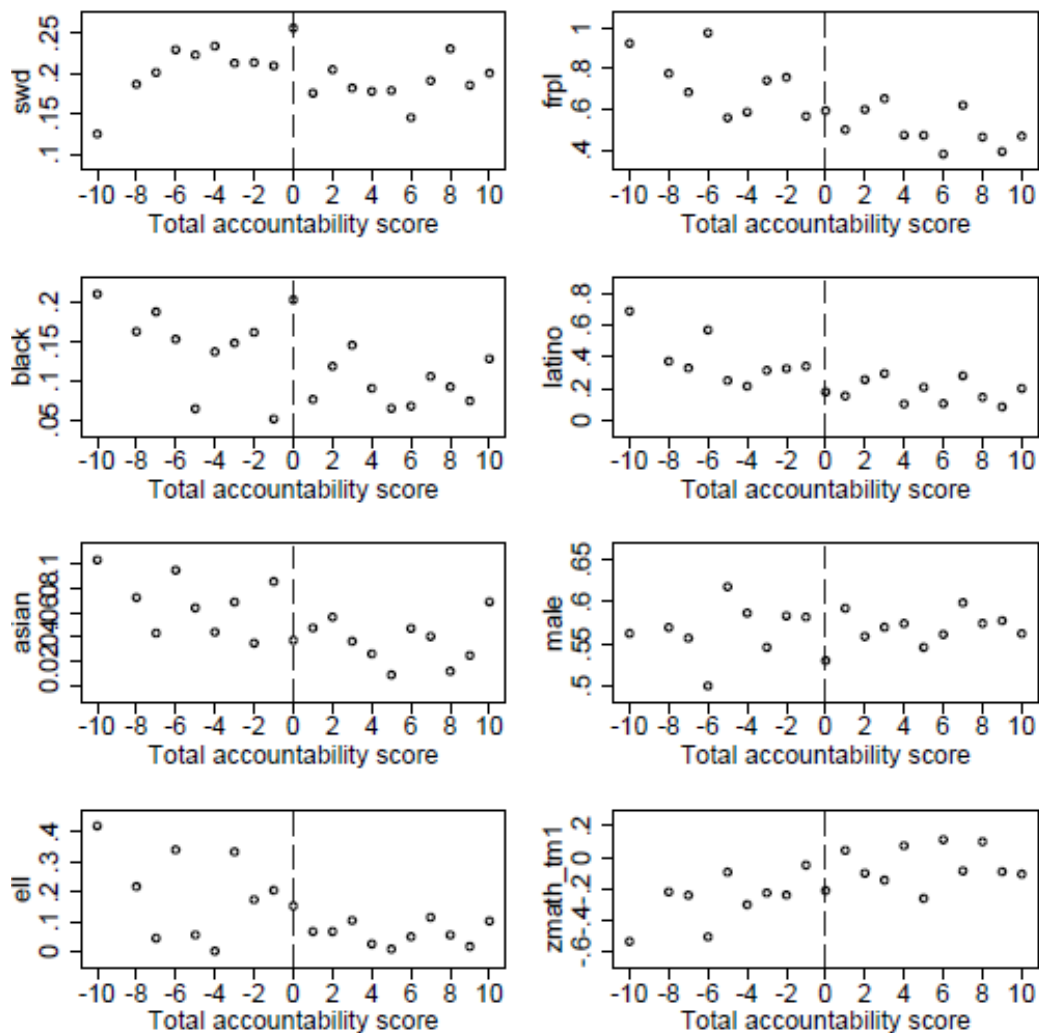
Running Head: Rhode to Turnaround

Figure A.2: Covariate Continuity Between Warning and Focus Schools
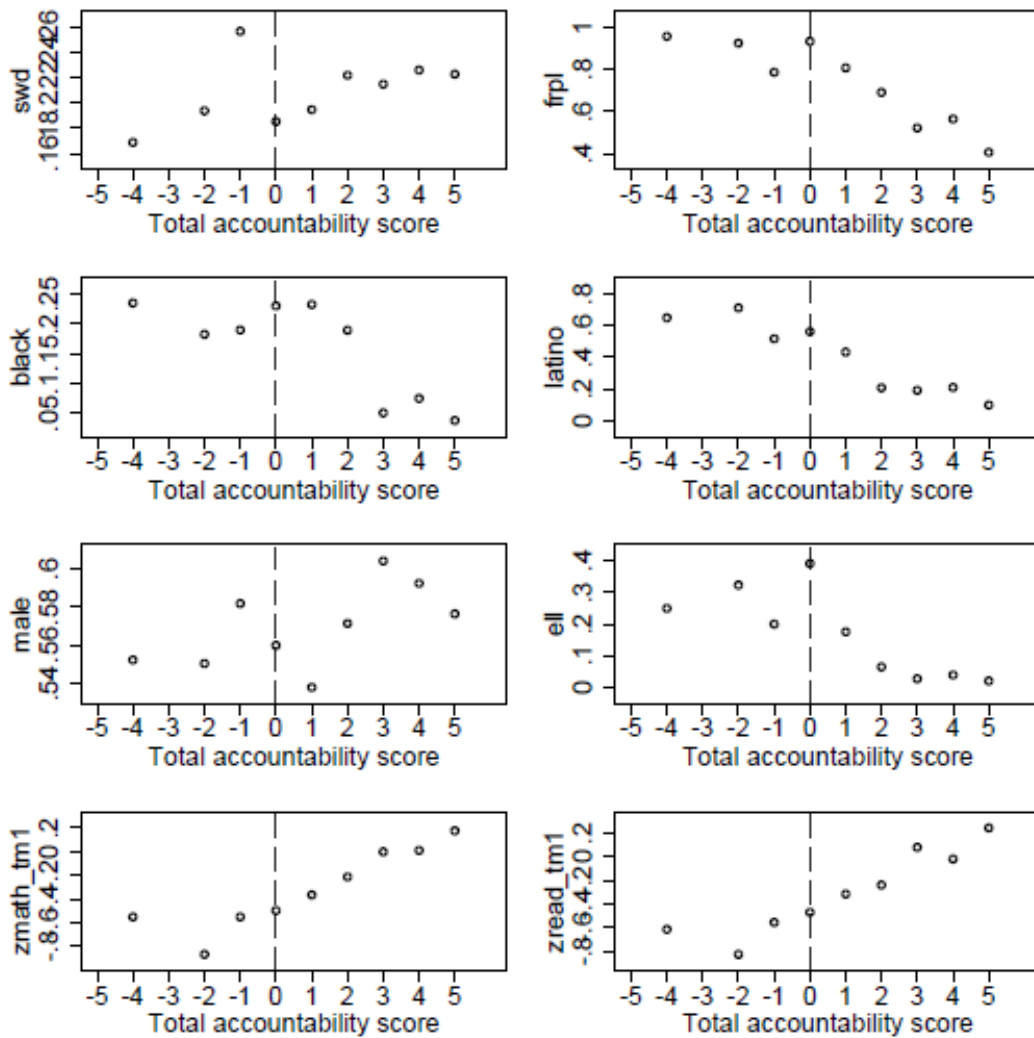
Table A.1: Reduced Form Effects in First Two Years of Policy and One Year Prior, for Warning and Focus Schools

	(1) Math Z-score	(2) Read Z-score	(3) Mover
Panel A - Effects of Warning Status Eligibility			
Warning Effect, 2012	-0.026	-0.030	-0.003
	(0.146)	(0.149)	(0.025)
N	30,858	30,858	30,858
Warning Effect, 2013	0.030	0.033	-0.021
	(0.176)	(0.169)	(0.048)
N	30,537	30,300	30,574
Warning Effect, 2014	-0.008	-0.008	0.000
	(0.178)	(0.162)	(.)
μ	-0.23	-0.21	0.06
N	31,416	31,108	31,912
Panel B - Effects of Focus Status Eligibility			
Focus Eligibility 2012	-0.016	-0.097	0.012
	(0.159)	(0.125)	(0.042)
N	11,307	11,307	11,307
Focus Eligibility 2013	-0.128	-0.196	0.045
	(0.213)	(0.178)	(0.056)
N	11,189	11,002	11,196
Focus Eligibility 2014	-0.175	-0.356***	0.000
	(0.171)	(0.132)	(.)
μ	-0.67	-0.67	0.00
N	11,549	11,375	11,720

Notes: Heteroskedasticity robust standard errors clustered by school are in parentheses (* $p<.10$ ** $p<.05$ *** $p<.01$).Estimates show the impact of being in a school assigned to receive interventions in response to low performance on various measures of student test scores. The coefficients shown are generated by OLS with local linear regression and a triangular bandwidth of 5 CIS points. Covariates are not included unless indicated, but school-by-grade fixed effects are included in all models. In the final row is the mean of the outcome variable among students just below the threshold.

Running Head: Rhode to Turnaround

Appendix A: Menu of Interventions

Flex Model Intervention Strategy Option			
Leadership	*Support*	*Infrastructure*	*Content*
Intervention III Strategies: Priority schools select one from each area. Focus schools select two strategies from areas of their choice.			
L-III.1: Removal of building principal and replacement with a leader with experience and/or training in turnaround environments	S-III.1: Require at least 30 hours of focused professional development with a focus on instructional strategies to support students with disabilities and English learners	I-III.1: Implement staff recommitment process to substantially different working conditions, including definition of school hours, job assignment, and job duties	C-III.1: Implement comprehensive improvement of instructional approaches for struggling students including focused professional development and a system for student progress monitoring
L-III.2: Restructure building leadership team to dramatically increase time available for instructional leadership	S-III.2: Hire building-level instructional specialists to support educators to serve English learners, students with disabilities, and other students at risk for failure	I-III.2: Dramatically increase common planning time and implement a system for its effective utilization, both horizontally and vertically	C-III.2: Review student course-taking patterns and make substantial changes to school schedule and student placement to ensure access to rigorous academic core
L- III.3: Provide building administrators the authority and autonomy to hire, manage teacher placement, budget, and school schedule	S-III.3: Implement a system of peer support and assistance to support the needs of educators	I-III.3: Review and change student enrollment and placement processes to increase family engagement & improve student outcomes	C-III.3: Implement a culturally competent support system to improve safety, reduce suspensions, increase attendance, and support all students
Intervention II Strategies: Priority and Focus Schools select two strategies from areas of their choice.			
L-II.1. Evaluate the principal and connect him or her with a mentor or appropriate resources to ensure ability to lead the school reform work	S- II.1: Implement a comprehensive drop-out prevention and reentry program	I-II.1: Complete an external audit of the use of school funds to guide staffing decisions and implement findings	C-II.1: Increase advanced coursework opportunities for students
L-II.2: Evaluate, assess, and diagnose the performance of the existing school leadership team and take appropriate job action	S-II.2: Implement a comprehensive ramp-up program for students at-risk of failure or subpopulations with the largest achievement gaps	I-II.2: Reallocate resources to increase support for direct instruction of students at risk for failure	C-II.2: Assign additional instructional coaches or other core content focused, job-embedded support for teachers
L-II.3: Contract with a vendor or partner with a track record of success to support the leadership team in school turnaround	S-II.3: Implement culturally competent family and community engagement program focused on	I-II.3: Develop and implement support systems for student transition into kindergarten and/or across break grades	C-II.3: Offer virtual education options for both at-risk and advanced students

	instruction and academic performance		
L-II.4: Identify one leader to routinely monitor the implementation and effectiveness of the core curriculum/instruction and services to traditionally underserved students	S-II.4. Hire full time parent/community engagement specialist to implement family and community engagement that is systemic, sustained, and integrated with school improvement	I-II.4: Establish a comprehensive system to support struggling teachers with content and pedagogy, especially teachers of students with disabilities and English learners	C-II.4: Implement an instructional monitoring system to ensure that the curriculum is being fully implemented and traditionally underserved students have access to academic core
L-II.5: Assign family/community outreach to member of leadership team and hold him/her accountable	S-II.5: Establish flexible or expanded learning opportunities with a focus on students at risk for failure	I-II.5: Implement a culturally competent tiered system of support focused on student psycho-social health	C-II.5: Increase student access to career, technical, or credentialing programs