Journal of Research on Educational Effectiveness

ISSN: 1934-5747 (Print) 1934-5739 (Online) Journal homepage: http://www.tandfonline.com/loi/uree20

Intervening with Struggling Readers in Seventh Grade: Impact Evidence from Six School Districts

Scott K. Baker, Nancy J. Nelson, Mike Stoolmiller, Patrick Kennedy Paine, Jessica Turtura, Deanne Crone & Hank Fien

To cite this article: Scott K. Baker, Nancy J. Nelson, Mike Stoolmiller, Patrick Kennedy Paine, Jessica Turtura, Deanne Crone & Hank Fien (2018): Intervening with Struggling Readers in Seventh Grade: Impact Evidence from Six School Districts, Journal of Research on Educational Effectiveness, DOI: 10.1080/19345747.2018.1472357

To link to this article: https://doi.org/10.1080/19345747.2018.1472357

Accepted author version posted online: 18 Jun 2018.

Intervening with Struggling Readers in Seventh Grade:

Impact Evidence from Six School Districts

Authors

Affiliation

Abstract

This study uses a regression discontinuity design to evaluate the impact of seventh grade interventions delivered by 25 schools on the reading and school engagement outcomes of struggling readers. Students in participating schools were assigned to intervention (n = 1,495) or comparison (n = 4,397) conditions, based on their state reading performance and reading fluency scores in sixth grade. On the state reading test, intervention impact favoring the treatment group approached, but did not reach statistical significance (p = .056, g = .15). On a measure of reading fluency, there was a significant interaction between the cut point schools chose for condition assignment and intervention impact (p = .015). In schools that chose a lower cut point, intervention students outperformed comparison students, while in schools that close a higher cut point, comparison students outperformed intervention students. On a student self-report measure of school engagement, no significant differences between the intervention and comparison groups were observed (p = .83, g = .02). Implications for rigorous evaluation of interventions implemented by districts and schools under naturalistic conditions are discussed.

Intervening with Struggling Readers in Seventh Grade:

Impact Evidence from Six School Districts

A substantial proportion of U.S. students lack the skills to meet school reading demands—a problem that appears to be worsening among adolescents. According to the National Center for Education Statistics (NCES: Nation's Report Card, 2015), the percentage of eighth grade students who read proficiently has decreased since 2013 (from 36% to 34%), and the percentage of eighth grade students who read below basic (i.e., lack even "partial mastery" of essential knowledge and skills, Snyder & Dillow, 2013, p. 218) has increased from 22% to 24%.

Poor readers in the middle grades are far more likely than other students to disengage from school (Guthrie & Davis, 2003) and eventually drop out. For example, eighth grade students in one longitudinal study in Philadelphia had a 75% chance of dropping out of school if their attendance was problematic *and* they failed a course in English or mathematics (Neild & Balfanz, 2006). Other studies have demonstrated that poor readers beyond third grade are four times less likely to graduate from high school on time compared to their grade level peers (e.g., Hernandez, 2011). After leaving school, dropouts are far more likely to experience diminished life opportunities (Dynarski et al., 2008; McLaughlin, Speirs, & Shenassa, 2014).

Encouragingly, however, high quality reading interventions in middle school have been shown to increase the likelihood that poor readers understand challenging academic content and meet rigorous academic standards (Torgesen, Houston Miller, & Rissman, 2007), suggesting that interventions to increase reading achievement could reduce school dropout (Dynarski et al., 2008). Based on this evidence, middle schools across the country are implementing interventions targeting reading and school engagement, and using data to guide implementation efforts, to improve academic achievement and prevent school dropout.

The purpose of this study is to evaluate the impact of a multi-component intervention on the reading and school engagement outcomes of seventh grade struggling readers in six school districts, under naturalistic conditions using a prospective regression discontinuity (RD) design. The Middle School Intervention Project (MSIP) was funded by the Institute of Education Sciences (IES), *Evaluation of State and Local Education Programs and Policies* competition. School districts implemented interventions consisting of three components with struggling readers: (a) reading intervention, (b) school engagement support, and (c) the use of data by school teams to make ongoing decisions about the reading intervention and engagement supports provided to students.

The rationale for evaluating these three intervention components together was based on school district partner priorities and areas of emphasis for addressing the needs of struggling readers. District leaders were concerned about the high dropout rate of students with poor reading skills and wanted their middle schools to address this problem in a comprehensive way. Thus, middle schools provided targeted reading interventions to struggling readers and addressed the elevated rates of school disengagement among poor readers by providing various types of school engagement supports for these students. In addition, school data teams monitored the implementation and impact of the reading and engagement interventions through regular meetings and analysis of relevant data. In the next three sections, we briefly describe the theoretical and evidentiary support for each of these intervention components to improve reading achievement and prevent high school dropout.

Using Reading Interventions to Improve Achievement in Middle School

In the past decade, several findings regarding reading intervention research with adolescents have emerged. One finding is that these interventions can have a positive impact on reading achievement. For example, Edmonds et al. (2009) analyzed 29 published reading intervention studies involving adolescent students and concluded the evidence was strong that interventions provided to students with reading difficulties could improve reading comprehension. Thirteen of these studies were included in a meta-analysis in which the overall effect size for reading comprehension was 0.89, a strong effect (Cohen, 1988). Other published reviews also support the use of reading interventions with struggling adolescents (e.g., Scammacca et al., 2007; Slavin, Cheung, Groff, & Lake, 2008; Wanzek et al., 2013). The IES Practice Guide on improving adolescent literacy recommends providing "intensive and individualized interventions for struggling readers" (Kamil et al., 2008, p. 7), with the authors concluding the level of scientific evidence for this practice is strong.

A second important finding of reading intervention research with adolescents is that the magnitude of the impact varies depending on who implements the intervention. In the Edmonds et al. (2009) study, although effects were significant in both cases, the impact was larger when researchers implemented interventions as opposed to teachers. Scammacca et al. (2007) observed a similar pattern among reading intervention studies in late elementary and middle and high school. In the eight studies that used a standardized reading comprehension outcome measure, the mean effect size was 1.08 (95% CI = .57, 1.59) when research staff delivered the intervention and .21 (95% CI = -.09, .50) when teachers delivered the intervention.

Scammacca et al. (2007) hypothesized that quality and consistency of implementation and stronger implementation fidelity (e.g., the notion that researchers may be more familiar with reading interventions and better trained to implement interventions for struggling readers), may help explain why decreased effects are observed when teachers, as opposed to researchers, deliver reading interventions. These findings are important, because in the vast majority of cases, teachers and other school staff provide reading interventions, and struggling adolescent readers need highly effective interventions to catch up to their peers academically.

The Striving Readers project provides the strongest evidence to date of the diminished effects of adolescent reading interventions when implemented by school staff (Boulay, Goodson, Frye, Blocklin, & Price, 2015). The authors describe 17 randomized controlled trials (RCTs) of 10 published programs. Of the 17 studies, 12 met What Works Clearinghouse (WWC) standards without reservations, three met WWC standards with reservations, and two did not meet the standards. Across the 15 studies that met WWC standards with or without reservations, six had a positive effect on reading achievement and nine had no observed effect. Of the 10 programs, only two included one or more studies with positive effects and zero studies with neutral or negative effects. For six of the 10 programs, there were no studies with positive effects.

A review of research by Slavin et al. (2008) may help explain why so many programs in the Striving Readers evaluation showed non-significant effects when implemented by teachers. Slavin et al. categorized studies into one of four approaches: (a) reading curricula (i.e., teacher-led implementation of a specific curriculum), (b) mixed-method models (i.e., implementation of a specific program combined with other components), (c) computer-assistive instruction (i.e., technology-delivered reading programs), and (d) instructional–process programs (i.e., professional development for teaching reading using cooperative learning, strategy instruction, or comprehensive school reform approaches). The main finding was that "programs designed to change daily teacher practices [i.e., instructional–process programs] had substantially greater research support than those focused on curriculum or technology alone" (Slavin et al., 2008, p. 290). In other words, implementation of targeted instructional practices supported by strong professional development was more important than the specific intervention program used.

Use of external resources to support implementation. Whether interventions were delivered by researchers or teachers, virtually all the studies in these reviews relied on resources external to the schools in which they were implemented to enhance implementation quality. For example, Striving Readers staff trained teachers in program implementation and provided ongoing coaching to improve fidelity during the study. In contrast, Dougherty (2015) investigated reading achievement in one school district that used only resources at its disposal to develop and implement a supplementary middle school reading program focused on the seven reading strategies described by Dole, Valencia, Greer, and Wardrop (1991). Dougherty used a RD design to determine whether students scoring just below the cut point who received the supplemental program made greater progress than students scoring just above the cut point who did not receive it. The cut point was set at the 60^{th} percentile, as the program was not intended to be remedial in nature. Dougherty found no overall treatment effect associated with the program. The current study also used RD to study intervention impact when implementation relied only on internal district and school resources, but did so with students who had pronounced reading difficulties. To our knowledge, the current study, the Dougherty study, and Baker et al. (2015), which examined the impact of a Tier 2 reading intervention on first grade reading achievement, are three of only a handful of studies that have used a prospective RD design to evaluate the impact of reading interventions implemented under naturalistic conditions.

School Engagement Interventions in Middle School

In addition to providing reading interventions, the districts participating in MSIP were interested in increasing student engagement in school as a mechanism for preventing school dropout. A growing body of research has investigated the impact of dropout prevention programs and other engagement interventions on the academic and social outcomes of middle school students. Dynarski, Gleason, Rangarajan, and Wood (1998) evaluated eight promising programs in the School Dropout Demonstration Assistance Program. Typical program components included counseling and support services, attendance monitoring, challenging curricula, accelerated learning approaches, parent outreach, and career-awareness activities (Dynarski et al., 1998). Dynarski et al. found that only one of the eight programs (Seattle's *Middle College High School*, a selective alternative program hosted on a community college campus) reduced the dropout rate in high school. There was no clear trend on improvements in grades, test scores, or attendance. In addition, across 42 treatment–control group comparisons, only one difference in students' self-reported school engagement was statistically significant.

A more recent review of dropout prevention studies conducted by the WWC (Dynarski, et al., 2008) identified three middle school programs with potentially positive effects. The *Twelve Together* program in grade 8 focused on peer support and mentoring in weekly after-school discussion groups, homework assistance, and connections to postsecondary education in the form of local college tours (Dynarski et al., 1998). It had a significant impact on students staying in school into grade 11. The *ALAS* intervention, studied with Latino students in grade 7 (Larson & Rumberger, 1995), was designed to increase school engagement by addressing student, school, family, and community factors that contribute to dropping out. At the end of ninth grade, *ALAS* students were significantly more likely to be enrolled in school and on track to graduate on time than control students. The *Check and Connect* program involves mentoring, case management, and regular monitoring of school performance. In the study reviewed, students received Check and Connect in seventh and eighth grades, and 94 students were randomly assigned to continue receiving the intervention in ninth grade (Sinclair, Christenson, Evelo, & Hurley, 1998). Students who continued to receive the intervention in ninth grade earned significantly more credits toward graduation and were significantly less likely to drop out of school at the end of grades 9 and 10 (Sinclair et al., 1998).

Overall, the evidence is modest for the potential of interventions to improve school engagement and/or reduce school dropout. It is important to note that most interventions studied that had positive impacts included academic support as part of the intervention. For example, in the IES Practice Guide on dropout prevention (Dynarski et al., 2008), eight of the nine programs that showed positive or potentially positive effects included an academic support component.

School Data Teams

A third part of the multi-component intervention districts implemented was a focus on regular review of student data to support adjustments to reading interventions and engagement supports. The use of school-based data teams to improve decision making has been prominent for decades (Deno & Mirkin, 1977). No Child Left Behind (2002) and the Individuals with Disabilities Education Act (2004) substantially increased the pressure on schools to use data to make decisions. There has been sufficient emphasis on the use of data to inform instructional decisions that an entire IES Practice Guide has been devoted to the topic (Hamilton et al., 2009).

Although relatively little research has been conducted demonstrating whether and how data use improves student outcomes, several themes have emerged from the available studies of data use (Crone et al., 2016). According to Mandinach (2012), data use is an iterative process that involves collection, analysis, decision making, and follow-up. Multiple sources of data are typically used to evaluate student progress, the reliability and validity of data for specific uses is emphasized, and data implications should be linked to professional development as needed (Mandinach, 2012). Similarly, Hamilton et al. (2009) recommended that data be used as part of an ongoing cycle of improvement: students are taught to examine their own data and set learning goals, schools establish a clear vision for building-wide use, a data-driven culture is fostered within the school, and a districtwide data system is developed and maintained. In MSIP, districts were committed to implementing recommended data-based decision making practices to support the effectiveness of their selected reading interventions and engagement supports.

The Current Study

The primary purpose of this study was to evaluate the impact of a multi-component intervention targeting seventh graders with reading difficulties. The study took place during the second year of a five-year project (MSIP) funded by IES to evaluate the impact of interventions selected and implemented by schools in authentic settings. Schools in the six participating districts had been implementing reading interventions with struggling readers for multiple years and districts wanted to know if the interventions they were using were effective. In addition, because many struggling readers were disengaging from school, the schools also implemented engagement supports with these students. School data teams used ongoing implementation and student data to monitor and adjust interventions and supports for struggling readers.

Each participating school, in consultation and coordination with the district, had established, was managing, and made ongoing decisions about the multi-component interventions. In other words, the school was the unit for precisely specifying the topography of the three intervention components. The common objective among all districts and schools was to improve reading and engagement outcomes among struggling readers through the implementation of reading and engagement interventions, and the use of school data teams for ongoing decision making.

A second but equally important purpose of this study was to conduct a rigorous evaluation using a prospective RD design, vetted by methodological experts. To date, most RD studies in education have been retroactive. Although retroactive designs are useful, we wanted to demonstrate the compatibility of using prospective RD in researcher–practitioner partnerships to evaluate the effectiveness of approaches that cannot be easily evaluated in an RCT. In this study, an RCT was not possible because the interventions with struggling readers were already being implemented in all participating schools. It was not feasible to discontinue these interventions and randomly assign students to intervention or comparison conditions. In addition, schools had strong beliefs about the positive impact of the interventions they were already providing. The use of a RD design according to the parameters identified for the evaluation was manageable to schools because it allowed their interventions to continue using an assignment procedure that closely mirrored methods schools already used to assign students to receive the intervention.

To permit a rigorous evaluation of intervention impact, participating schools agreed to follow a set of shared implementation parameters. First and foremost, they agreed to use a formal cut point as the criterion for determining which students received the reading intervention, and agreed that only students scoring below the cut point would receive the reading intervention. Schools also agreed to use evidence-based strategies to teach reading; ensure that teachers were trained to deliver interventions; deliver interventions for a substantive amount of time; deliver interventions that were substantively different from the reading instruction students in the comparison group received as part of regular, English Language Arts (ELA) coursework; and regularly monitor student progress. They also agreed to implement an engagement intervention for all students assigned to receive reading intervention, and to discuss all reading intervention students in data team meetings. The full set of features schools committed to implementing for each component of the intervention is described in Table 1. Because schools were committed to using these research-informed practices to implement interventions, we had reason to expect that

(a) the multi-component intervention would have a significant impact on one or both of the reading measures used to evaluate outcomes, and (b) we could draw conclusions about the effects of using a research-based, comprehensive approach to significantly improve adolescent reading achievement.

A third purpose of the study was to examine intervention effects when implemented by school staff under typical conditions, without external resources. Thus, participating schools implemented all intervention components without training or guidance from research staff. This was intentional, as the focus of the evaluation was on providing districts with data and information about whether their selected interventions were working—not on evaluating the effects of a particular intervention with a specified level of implementation support as one might expect to occur in a tightly controlled efficacy trial. The fact that this evaluation was conducted to study practices that schools elected to implement based on their interpretation of research evidence helps address an unanswered question regarding the intervention impact districts and schools can achieve on their own, when no research support or other atypical external resources are available to support implementation. The answer to this question is vital for students receiving reading interventions in naturalistic settings (i.e., students in most schools across the country), where districts and schools must rely on internal resources and capacity to implement and sustain academic programs, practices, and interventions.

Method

Participants

Participants were teachers and students from six school districts from one Pacific Northwest state. Two districts were located adjacent to the second largest city in the state, and four were suburban districts located within the state's largest metropolitan area. Districts ranged in size from 6,072 to 39,736 students (NCES, 2011), and were recruited to participate because (a) we had worked with them previously on research and outreach projects, including the first year of the MSIP project; and (b) they were implementing interventions with struggling readers to improve reading and school engagement outcomes.

Twenty-five comprehensive middle schools in these districts provided intervention in seventh grade, representing 63% of the districts' schools that served seventh grade students. Schools did not participate if they (a) did not offer a traditional middle school curriculum (n = 9), or (b) had less than 10% of seventh grade students eligible for the intervention (n = 6). On average, participating schools were larger and served a more academically diverse student population than non-participating schools in the six districts. Participating schools ranged in size from 373 to 1,096 students.

All students who attended a participating school for any portion of seventh grade were included in the study, as long as they had a valid score on at least one of the two reading assessments used for condition assignment (n = 5,892). The assessments were administered the prior year, when students were in sixth grade. Students' condition assignment was based on whether they scored above or below a cut point. Students below the cut point were assigned to intervention and students above the cut point were assigned to comparison. The sample was similar to both state and national averages on a range of demographic characteristics. For example, 22% of students in the study were Hispanic, and 17% represented other minority groups, compared to 22% and 25% Hispanic, and 13% and 25% other minority at the state and national levels. Similarly, 59% of the sample was eligible for free or reduced price lunch, compared to 54% and 52% at the state and national levels.

Description of the Intervention

The three intervention components (i.e., reading intervention, school engagement supports, and school data teams) were in place in the participating schools prior to the study. This underscores a central purpose of the project, which was to evaluate the impact of school districts' existing intervention practices on outcomes. For the evaluation, schools committed to (a) provide reading interventions only to students who scored below the cut point, with one categorical exception, described below; (b) provide school engagement interventions to all intervention students; and (c) monitor all intervention students and use data teams to make ongoing decisions about the interventions, complying with a set of critical intervention features for each component of the intervention (see Table 1). It was permissible for schools to also provide school engagement interventions and use data teams with other students[1] (see the Results section for data on how frequently this occurred).

The attributes of the intervention components varied among schools and included type of program (e.g., curricula used); frequency, duration, and length of interventions; staff qualifications; and ratio of teachers to students (Author et al., 2015). To document intervention features, we conducted direct observations of reading intervention classes, English Language Arts (ELA) classes, and data team meetings. We also documented school engagement interventions using intervention logs completed by school staff. The observations and logs are not a primary focus of this study and are not described in-depth. For more information about the observations and logs see Crone, Baker, & Fien (2014), Kennedy, Schwartz, Baker, Crone, & Fien (2013), and Nelson-Walker et al. (2012).

[1] This was necessary because schools felt strongly that some students needed engagement interventions even though they were not experiencing reading difficulties. Similarly, not all the students discussed at data team meetings were students with reading problems.

Schools made all intervention decisions without assistance from the research team. Although the research team intentionally refrained from providing direct support to improve the quality of intervention implementation, the team did offer indirect support in several ways. First, we summarized and presented to districts and schools student performance, reading intervention implementation, and engagement data prior to and during the intervention, and we worked with them in data summit meetings at least once per year to identify ways in which they could use these data to improve their interventions. Second, we brought experts in adolescent reading interventions to the data summit meetings to present research on reading instruction and interventions for struggling adolescent readers. In some cases, districts elected to use their own resources to further engage with these experts following data summits. Third, we required schools to provide detailed plans on the reading interventions they were providing to intervention students. This helped ensure that criteria related to dosage were met, and in a number of cases helped schools determine how to provide more frequent and intense reading interventions.

Reading intervention. The purpose of the reading interventions was to provide targeted support to improve reading achievement for struggling readers. All students assigned to the intervention received a reading intervention; most of these students also received ELA instruction. Students in the comparison condition received ELA instruction only. We observed each reading intervention class and each ELA instructor three times during the study. We tracked instruction in ELA classes because, outside of the reading intervention, ELA classes were the most likely setting in which struggling readers would receive reading support. In addition, all comparison students were enrolled in ELA classes, and this served as our best estimate of the reading support students received in the absence of a reading intervention.

Engagement supports. Engagement supports were designed to strengthen students' behavioral and psychological engagement in school (Finn, 1989). Schools determined the engagement supports students received, documented plans, and varied the intensity as needed based on academic and behavioral data. Engagement supports included (a) check-in/check-out programs (cf. Crone, Hawken, & Horner, 2010), (b) student interest groups such as computer clubs and book groups, (c) academic tutoring, (d) mentoring and counseling, (e) homework club or study hall, (f) social skills groups, and (g) after school extracurricular activities. Schools modified supports during the year based on engagement, attendance, and academic performance.

School data teams. The purpose of school data teams was threefold: (a) summarize and report ongoing academic achievement and student engagement data, (b) use these data to evaluate the effectiveness of the interventions, and (c) modify interventions as necessary to improve outcomes (Crone et al., 2016). School data team meetings were observed by trained research staff at least once per academic term.

Study Measures

We administered two reading assessments at the end of sixth grade to assign students to condition, and again at the end of seventh grade to measure achievement. We also collected a student self-report of school engagement in the fall and spring of seventh grade.

Oregon Assessment of Knowledge and Skills Reading/Literature (OAKS–R). The OAKS–R (Oregon Department of Education, 2012) is a criterion referenced test aligned with grade-level content standards. It is an untimed, computer-based, multiple-choice test; 79% of the items address reading comprehension and 21% address vocabulary knowledge (ODE, 2012). School personnel oversaw administration of the OAKS–R. Students were allowed to take the test up to three times during the year (ODE, 2012), and the highest score was used for analysis.

easyCBM Passage Reading Fluency (PRF). easyCBM PRF (Alonzo, Tindal, Ulmer, & Glasgow, 2006) is a standardized, individually administered measure of oral reading fluency. Students read aloud from a 250–350 word passage for one minute. Districts used a combination of classroom teachers and trained data collectors to administer the measure and calculate the number of words read correctly, which was used for analysis. The average correlation between a reference PRF passage and 19 other seventh grade passages was .89 (Alonzo & Tindal, 2008). easyCBM PRF also predicts performance on the OAKS-R. Based on a convenience sample from three districts, the correlation between seventh grade fall PRF scores and seventh grade OAKS-R scores was .68, accounting for 15% of OAKS–R variance (Anderson, Alonzo, & Tindal, 2010).

Student Engagement Instrument (SEI). The SEI is a measure of students' self-perception of their engagement with school, family, and peers (Appleton, Christenson, Kim, & Reschly, 2006). The SEI consists of 35 belief statements (e.g., *school is important for achieving my future goals*). For each, students indicate the degree to which they believe the statement describes them (1 = *strongly agree*, 2 = *agree*, 3 = *disagree*, and 4 = *strongly disagree*). Items are grouped into one of six factors: teacher–student relationships, peer support for learning, family support for learning, control and relevance of schoolwork, future aspirations and goals, and extrinsic motivation. Coefficient alphas for each factor, based on a sample of 1,931 ninth grade students, ranged from .72 to .88 (Appleton et al., 2006). Validity evidence for the SEI is limited, although correlations between the SEI and relevant academic and behavioral outcomes provide some evidence of convergent and discriminant validity (Appleton et al., 2006).

Assignment of Students to Condition

Normative information was used to standardize the OAKS–R and PRF scores into a single z-score distribution. The z-score for the OAKS-R was based on the statewide standard deviations. The z-score for PRF was based on the standard deviation of the study sample, as no statewide normative information was available. Students' standardized *z*-scores for each measure were then averaged to create a composite cut score for each student. The cut score for students who took only one of the two assessments was the z-score for that assessment.

Prior to the intervention, a rank ordered list of cut scores was distributed to each school. Key staff (e.g., principals, specialists) chose the cut point their school would use for assignment to condition, typically based on the number of students they had the capacity to serve in the intervention group. Students with cut scores below this point were assigned to the intervention group, and students with scores above it were assigned to the comparison group. Schools followed project guidelines so that a minimum of 20% and a maximum of 80% of students received the intervention, helping maintain statistical power for the RD analysis. Allowing schools to select the cut point that they felt best met the needs of their students was essential to ensuring schools' participation in the evaluation, but also presents a potential threat to the RD design. To empirically evaluate this threat, we performed the McCrary test (McCrary, 2008) to test for discontinuities in the cut score distribution at the cut point that would suggest deliberate manipulation of intervention assignment status.

Strategies to Increase Study Compliance

Multiple cut points for condition assignment. The use of multiple cut points for condition assignment in RD studies is not typical in education research (e.g., Bloom, 2012; Nomi & Allensworth, 2009), but was critical to maintaining school participation. A single cut point for all schools would have resulted in some schools serving a very different proportion of students in the reading intervention than they typically served, placing disproportionate demands on district and school resources, and compromising a fundamental premise of the study, which was to investigate effects as schools and districts typically applied interventions. To account for different cut points in analyses that pooled students across schools, we centered the cut scores within each school by subtracting from each student's cut score the individual school's cut point so that, for analysis and interpretation purposes, all schools had the same cut point (i.e., zero). Comparable to using school mean centering in a multilevel model (Raudenbush & Bryk, 2002), we included the school cut point value in all models as an additional school level predictor.

Use of *wild cards* in the assignment process. In addition to selecting their cut point, schools were allowed to exempt up to 5% of their sixth-grade students from placement into the condition indicated by the student's cut score. These *wild card* exemptions were necessary to ensure schools' participation because some school staff felt strongly that certain students should or should not be placed into intervention. These wild card students are included in the analysis, but the effect size estimates pertain only to the students who complied with their assignment.

Final Student Sample

Using the wild card procedure, 120 students (8.0%) below the cut point were assigned to the comparison condition, and 103 students (2.3%) above the cut point were assigned to the intervention condition. Overall, 223 of 5,892 students (3.8%) were in the wild card group, below our a priori limit of no more than 5% of all students. At the end of the year, we checked actual participation against fall condition assignment. The vast majority of students participated in the condition they were assigned to, either by cut score or by wild card. Most students who did not were the result of additional (unplanned) non-compliance, but in a few cases, students originally designated as planned non-compliers (i.e., wild card slots) reverted to compliers. In all, 6.5% of the sample did not comply with their assigned condition. Below the cut point, 12.7% of students assigned to intervention did not receive it. Above the cut point, 4.3% of students assigned to the comparison condition received the intervention. A total of 1,495 students (25.4%) received intervention, 1,304 (87.3%) of whom scored below the cut point. The other 4,397 students (74.6%) represent the comparison condition, 4,207 (95.7%) of whom scored above the cut point.

Fuzzy RD Design and Analysis

The fuzzy RD (Bloom, 2012) accommodates imperfect compliance with group assignment (i.e., non-compliers are included in the analyses), under the same background assumptions as the standard RD, with one important addition: there must be a significant gap in the probability of receiving intervention at the cut point, despite the non-compliance (Bloom, 2012). An important qualification is that the effect size estimate only pertains to compliers. The treatment effect in a fuzzy RD is typically estimated using two regression equations, one for the outcome and one for the treatment probability. The outcome analysis is the same as in the sharp RD design, and consists of estimating the difference in mean outcomes between the intervention and comparison groups at the cut point. The treatment probability analysis is similar, but estimates the difference in mean probability of receiving the intervention above but right at the cut point, versus below but right at the cut point. The fuzzy RD effect is estimated by dividing the outcome gap by the treatment probability gap (Bloom, 2012). An approximate standard error can be computed for the fuzzy RD effect from two separate regression equations (Bloom, 2012; Schochet, 2008) and a slightly more accurate standard error can be computed using an SEM program that estimates the two equations simultaneously.

For a standard RD analysis, it is important to accurately model the relation between the cut score and the outcome on both sides of the cut point. For a fuzzy RD analysis, the same concern applies to the relation between the treatment probability and the cut score. Consequently, non-parametric or local regression procedures that do not impose a functional form on the regression are popular (Berk, 2011; Bloom, 2012). A disadvantage is that if the true regressions are relatively simple (e.g., linear), there can be a substantial loss of power. Because our cut score was reading proficiency from the prior year, we expected the regressions for reading outcomes to be relatively simple. To check this assumption, we used two different non-parametric RD procedures. One was the multilevel GAM procedure (GAMM4) as implemented in R (R Core Team, 2016) by Wood (2006), where the data determines the degree of smoothing using generalized cross-validation procedures. We used separate smoothing procedures on each side of the cut point and a random intercept to capture variation in mean levels of the outcome on the right (high) side of the cut point and the random RD effect to capture variation in differences between mean outcomes on the left and right side of the cut point. A complete description of the model equations is provided in the Appendix.

The second procedure was local linear regression. With this method, the analysis is restricted to participants with cut scores that are *close* to the cut point and within a certain bandwidth. A standard linear regression model is used, with separate intercepts and slopes on each side of the cut point. The procedure uses weighting to progressively minimize the influence of participants as they get farther from the cut point. The size of the bandwidth is typically varied to assess the sensitivity of the RD effect to the choice of bandwidth (Bloom, 2012), but recently criteria have been proposed for identifying a bandwidth that is supposedly optimal with respect to bias and efficiency (IK bandwidth, Imbens & Kalyanaraman, 2012). We used this approach as implemented in R (Dimmery, 2013) across a wide range of bandwidths, with robust standard errors corrected for students clustered in schools.

All models included school cut point as an additional covariate, and because of the importance of the goodness of fit of the regressions, we include detailed scatter plots of the raw data and the fitted GAMs (see Figures 1 & 2). The GAM models also included the cross-level interaction term between school cut point and intervention assignment variable to check for variation in RD intervention effects that might be linearly related to the level of the school cut point. This tests for a specific form of heterogeneity in RD effects across schools in addition to the more general school level random effect in the GAM that tests for any form of heterogeneity in RD effects.

Results

Missing Data

The rate of missing outcome data was minimal, 2–4% overall, 2–4% for the comparison group and 4–6% for the intervention group. The rate for the SEI was slightly higher (3.6% and 5.8% in comparison and intervention, respectively) than the rate for PRF (2.1% and 4.3%) and OAKS–R (1.7% and 3.6%).

Initial Checks of Assignment Integrity

As described above, schools were allowed to select the cut point that they felt best met the needs of their students. Therefore, to check for discontinuities in the cut score distribution at the cut point that would suggest deliberate manipulation of intervention assignment status, we performed the McCrary test (McCrary, 2008) using a GAM, and as implemented in the rdd package in R (Dimmery, 2013). We found no evidence of a discontinuity using either approach. The cut score took on 5,102 (86.6%) unique values among the 5,892 students in the study, meaning it was relatively continuously distributed, even close to the cut point. This remained true even as we progressively shrank the window about the cut point to within .5 points of the cut point. Normal quantile plots of these scores within the windows (not shown) appeared smooth and continuous and showed very little clumping of points at identical values.

We also performed fuzzy multilevel RD analyses on observed pseudo-covariates to check that assignment was based solely on the cut score (Imbens & Lemieux, 2008). We tested gender, special education status (SPED), limited English proficiency status (LEP), and self-reported engagement in school in the fall of seventh grade, before the start of the intervention. We found no significant evidence of fuzzy RD effects on any of these covariates. That is, it appears schools did not rely on any of these variables in addition to the cut score to assign students to condition.

Descriptive Statistics

In Table 2, we present descriptive statistics for the intervention and comparison groups across the full study sample. The large mean difference in the cut score between intervention and comparison groups was by design, because students were assigned to condition based on their cut score. Mean values for reading outcomes at the end of grade 7 indicate that the comparison group continued to score higher than the intervention group. On the SEI, the mean values at the end of seventh grade were virtually the same for both groups. However, neither of these characteristics necessarily preclude a significant discontinuity at the cut point. Table 2 also reports descriptive statistics for students within .50 standard deviations of the cut point. For brevity, we refer to this group as *students close to the cut*.

Implementation data by condition for students close to the cut. Descriptive data for students close to the cut on each of the three intervention components are presented in Tables 3 and 4. For the intervention group, we present data for two subgroups of students: those whose reading instruction included *both* a reading intervention and ELA instruction (n = 550), and those whose reading instruction included just the reading intervention and no ELA instruction (n = 86: i.e., those for whom the reading intervention replaced their ELA class rather than supplementing it). Across Tables 3 and 4, the key point is that for all three components of the intervention,

intervention students close to the cut had very different, more instructionally intense experiences than did comparable comparison students.

For example, comparison students close to the cut received, on average, approximately 161 hours of literacy instruction over the course of the year, whereas both groups of intervention students close to the cut received an average of about 235 hours of literacy instruction, almost 50% more than what comparable comparison students received (see Table 3). Another important difference between conditions illustrated in Table 3 is the difference in class sizes. Reading intervention classes were, on average, only slightly more than half as large as ELA classes.

With respect to content coverage, observations indicated that ELA classes spent about half of their time on writing and reading comprehension activities (25% and 23%, respectively). Reading intervention classes spent about half of their time on reading comprehension (32%) and other literacy activities (19%), which included spelling and grammar instruction. At 7%, reading intervention classes spent much less time on writing than ELA classes. Only about 2% of instructional time in reading intervention classes was devoted to word-level decoding activities, while 19% of time was devoted to reading connected text. ELA classes spent very little time (1%) on decoding activities and about 11% of time on reading connected text.

Table 3 also presents data on the use of published programs by condition. The most commonly used published programs in reading intervention classes were Language! (47 classrooms; Greene, 2004), Read 180 (25 classrooms; Hasselbring, 2006), Read Naturally (19 classrooms; Ihnot, Mastoff, Gavin, & Hendrickson, 1992), Corrective Reading (18 classrooms; Meyer, Carnine, Becker, Eisele, & Johnson, 1989), and Step Up to Writing (14 classrooms; Auman, 2011). The bottom portion of Table 3 presents data on the third component of the intervention, the use of data teams to review data and adjust interventions. The table shows the average number of students close to the cut discussed at these meetings and the average discussion length per student. These data indicate that at data team meetings, participants discussed intervention students close to the cut almost four times as often as comparison students close to the cut, and the discussions of intervention students lasted slightly longer on average than those for comparison students.

Table 4 presents descriptive data on the engagement supports students close to the cut received. Only about 20% of comparison students close to the cut received some form of engagement support, and for those who did, most of what they received was categorized as student interest support (e.g., art, music, and sports—activities students typically self-select into). In contrast, 100% of intervention students close to the cut received some form of engagement supports, and many participated in multiple types of activities, with more than half receiving mentoring or systematic check-in support.

Fuzzy RD Results

Preliminary analyses included the cross-level interaction between school cut point and the intervention assignment variable to check for variation in RD intervention effects. No significant effects were detected for OAKS or SEI, so we dropped the cross-level interaction from each of these models. However, the cross-level interaction effect was significant in the PRF model, so it was retained. To test for a discontinuity in the treatment probability, we fitted smooth functions from the multilevel logistic GAM. The mean treatment probability above but right at the cut point was .196, and the corresponding probability below but right at the cut point was .837, resulting in a significant gap of .641 (z = 13.91, p < .001), as required by the fuzzy RD assumptions (Marmer, Feir, & Lemieux, 2014). Results of the RD analyses for all outcomes are shown in Table 5. Details regarding the parameters in Table 5 are provided in the Appendix.

OAKS–R. As shown in Figure 1, a small positive gap (i.e., intervention higher than comparison) in the mean OAKS–R score occurs at the cut point. In the multilevel Gaussian GAM, the intent-to-treat (ITT) RD effect was 0.87, which was not significant ($z = 1.93$, $p = .054$, $g = .10$). The corresponding treatment-on-the-treated (TOT) effect using the approximate standard error from Schochet (2008) was 1.36, which was also not significant ($z = 1.91$, $p = .056$, $g = .15$). Results were essentially the same using the bandwidth approach. The fuzzy RD effect ranged from 0 to a maximum of 1 across a wide range of bandwidths, was 0.84 at the IK optimal bandwidth, and was never close to significant for any bandwidth. A meta-analysis treating the 25 schools as 25 individual studies, each with its own separate bandwidth analysis, led to the same results. Consistent with the GAM approach, the overall fuzzy RD effect was not significant, although the Q test for heterogeneity in fuzzy RD effects across schools was ($Q = 40.10$, $df = 24$, $p = .021$). Specifically, one of the 25 schools had a statistically significant positive effect ($g = .34$), and one school had a statistically significant negative effect ($g = -.38$).

Passage Reading Fluency (PRF). The overall intervention effect on PRF was not significant ($p > .05$). The two panels in Figure 2 illustrate the significant cross-level interaction effect (-7.26, $z = -2.44$, $p = .015$) between the school cut point and the ITT RD effect. The negative effect indicates that as schools chose higher cut points, the ITT RD effect went down. For example, the left and right panels correspond to schools that chose cut points close to 0 and -1 respectively, which roughly correspond to the maximum and minimum school cut points used for student assignment to condition. Across schools, a cut point of 0 would correspond to an average of 13% of students being assigned to intervention (range = 4% to 26%), whereas a cut point of -1 would correspond to an average of 49% of students being assigned to intervention (range = 36% to 65%).The left panel (i.e., schools with cut points near 0) shows a small negative gap (comparison higher than intervention) of -4.45 points (z = -2.15, p = .032) at the centered cut point (i.e., zero) and the right panel (i.e., schools with cut points near -1) shows a small positive gap (intervention higher than the comparison) of 2.81 points (z = 1.37, p = .170) at the centered cut point (i.e., zero). The corresponding TOT effects using the approximate standard error from Schochet (2008) were -6.94 (z = -2.12, p = .034) and 4.39 (z = 1.37, p = .172), respectively. The standard deviation for PRF in our sample was approximately 40, so the respective standardized fuzzy RD effect sizes were -.17 and .11, both small effects. The variance component for school level variation in the PRF gap was not significant (p = .567), so it was dropped from the model.

School engagement. No gap in the mean SEI score was observed at the cut point. The ITT outcome gap from the GAM model was .17, which was not significant (z = 0.21, p = .83). The corresponding TOT effect using the approximate standard error from Schochet (2008) was .27, also not significant (z = 0.21, p = .83). Dividing the fuzzy RD effect by the standard deviation of SEI in our sample, 13.8, results in a standardized effect size estimate of 0.02, a very small effect. Results for the overall effect were similar using the bandwidth approach, and the Q test for heterogeneity in fuzzy RD effects across schools was not significant. For the bandwidth approach, we skipped the overall test due to the significant cross-level interaction, and instead checked for any kind of heterogeneity in fuzzy RD effects across the 25 schools using the meta-analytic Q test. The test was not significant (Q = 32.73, df = 24, p = 0.11), suggesting that the specific form of heterogeneity captured by the significant cross-level interaction was not strong enough to reach significance.

Discussion

We used a prospective regression discontinuity (RD) design to evaluate the impact of comprehensive interventions implemented by schools in seventh grade to improve reading and school engagement outcomes for struggling readers. The study included 5,892 seventh grade students, 1,495 who received the intervention and 4,397 who were in the comparison group. The overall effect of the multi-component intervention on student reading and engagement outcomes was non-significant. Nevertheless, the study makes several contributions to the knowledge base on reading interventions with middle school students.

First, although the intervention effect on the primary reading measure (the OAKS–R) was not statistically significant, the effect size (g = .15) was consistent with previous findings, especially those conducted in settings with limited or no external support for implementation. This is important, because it reinforces that reading interventions implemented in naturalistic settings typically result in much smaller effects than the effects observed in tightly controlled studies under ideal conditions, and that stronger interventions, professional development, and fidelity monitoring approaches are needed for struggling readers to benefit. In addition, the interaction effect on the reading fluency measure—in which schools that selected a lower cut point had positive effects and schools that selected a higher cut point had negative effects—offers a potentially useful way of thinking about how to better support struggling readers in middle school.

Second, although schools had to make some small changes in how interventions were delivered to adhere to the RD design specifications, they were able to do so without deviating much from their typical protocol for intervention provision. This provides evidence that research institutions and schools can work together to conduct rigorous evaluations of intervention practices implemented in naturalistic settings. Third, few studies have been conducted in which interventions are implemented by district and school staff using only the resources normally at their disposal, as was the case in the current study. Each of these contributions are discussed in more detail below.

Implications for School-Based Evaluations

This study contributes to the growing number of studies in education in which interventions to improve student outcomes are implemented at scale in typical school settings by school staff. Many of the studies published to date, however, have included external resources to identify interventions and support implementation (e.g., Striving Readers). The current study is one of only a few studies to rigorously evaluate intervention impact at scale in which no resources from the project or other external sources were used to select or implement the intervention (Dougherty, 2015). This is important because SEAs, districts, and schools most often make intervention decisions using internal resources to determine how to support struggling students. Determining the impact of these efforts is critical to understanding how well struggling readers are being supported in the vast majority of school settings.

The use of an RD design enabled an unbiased estimate of impact, and the design was compatible with the approach participating districts and schools were already using to deliver interventions. Districts were not required to randomly assign students to interventions to assess effectiveness, which is often a problematic approach when districts have already adopted practices they believe are effective. The RD approach required schools to make only minor adjustments to their interventions so that valid conclusions could be reached about impact. For example, schools had to adhere more strictly to RD cut score procedures than they would have in the absence of the experiment. However, schools insisted on having leeway to use professional judgment to override *some* assignments based on a cut score. Consequently, a compromise was reached whereby schools could place up to 5% of students in either condition regardless of the cut score. In total, 3.7% of the sample was assigned to condition using this wild card strategy.

Many RD studies have relied on a cut point based on local educational policy and have experienced variability in cut score adherence (e.g., Dougherty, 2015; Wong, Cook, Barnett, & Jung, 2008). In the present study, compliance was sufficient to permit use of a fuzzy RD to estimate a causal association between the intervention and student reading and engagement outcomes. This compliance finding is valuable to researchers and practitioners who work in partnership. Researchers can benefit by applying the strategies employed in this study to increase RD compliance, while districts benefit from understanding that compliance enhances certainty that it is the interventions producing the outcomes observed, rather than other factors.

Implications for Intervention Research and Practice

Reading outcomes and interventions in middle school. The main feature that distinguished the intervention and comparison groups in this study was the reading intervention. Two measures were used to assess intervention impact—the state assessment (OAKS-R) and an oral reading fluency measure. Impact on the state reading test was not statistically significant, although the difference approached significance and favored the treatment group (p = .054, g = .15). Figure 1 depicts a small break in the regression line illustrating this effect. The magnitude of the treatment effect equates to an increase of approximately one and a half points on the state assessment for students in the intervention group at the cut point.

There was no indication that the treatment effect varied as a function of the cut point schools chose to determine condition assignment. However, one school had a significant positive intervention effect (p = .008, g = .34), and one school had a significant negative effect (p = .005, g = -.38) on OAKS-R. Although this is inconsistent with the overall pattern of no school-level effects, our follow-up research studies are exploring this issue. For example, we are exploring effects related to specific implementation and context factors including types of reading intervention programs used, the association between intervention delivery in the classroom and student outcomes, and whether student characteristics are associated with treatment effects.

A significant interaction occurred between the cut point schools chose for condition assignment and the PRF score in the spring of seventh grade (p = .015), even though the overall intervention effect on PRF was not significant. In schools that set a low cut point, and thereby provided the intervention to students with more pronounced reading difficulties, students in the intervention condition gained more than students in the comparison condition. In schools that set a higher cut point, comparison students outperformed intervention students. Figure 2 illustrates this pattern, using two cut points to represent low and high cut score locations. The estimated effect sizes at these two cut point locations are small, .11 and -.17, respectively.

One possible reason for this cut point by PRF interaction effect is that intervention students in schools that used a relatively low cut point had lower reading fluency skills on average than students in higher cut point schools, and consequently had more room to improve on this type of measure. This growth could occur if reading interventions emphasized activities designed to build oral reading fluency, such as reading connected text. In the current study, intervention classes spent roughly one-fifth of their time on reading connected text, compared to one-tenth of the time in comparison ELA classrooms. This difference may have contributed to the greater growth of students with more pronounced reading fluency difficulties. In high cut point schools, time spent reading connected text at the expense of comprehension or other reading activities may have had an unintended negative impact on oral reading fluency.

Research on reading interventions in middle school demonstrates that in studies in which implementation quality is a driving concern, reading interventions can have practically and statistically significant impacts on reading comprehension (e.g., Edmonds et al., 2009;

Scammacca et al., 2007). In contrast, studies that implement interventions at scale by school staff under more typical conditions usually do not achieve these impacts (Boulay et al., 2015; Dougherty, 2015). Similarly, effects are stronger when research staff deliver the interventions compared to practicing teachers (e.g., Edmonds et al., 2009; Scammacca et al., 2007).

The Striving Readers studies provide the clearest evidence for estimates of impact when middle school reading interventions are delivered in naturalistic contexts by practicing teachers (Boulay et al., 2015). In six of the 10 programs investigated, not a single study produced a positive effect. One program did stand out from the others: *Read 180* resulted in positive effects across all three studies that used the program. The effect sizes, however, were small (.18, .14, and .21). In the current study, although the effect size on the OAKS–R was not statistically significant, the magnitude of the effect was consistent with other intervention studies in which the interventions were delivered by teachers in naturalistic settings (Boulay et al., 2015).

A crucial question remains how likely it is that reading interventions in middle school, when delivered by practicing teachers in typical settings, can achieve substantial, positive effects on student outcomes, like those observed in tightly controlled studies. This may be a sizable undertaking. For instance, a reading intervention program such as *Read 180*, in which multiple studies report positive findings, shows effects that are significant but small (ES = .14 to .21). Yet, *Read 180* appears to be precisely the type of intervention called for in the IES practice guide (Kamil et al., 2008). It includes daily, 90-minute lessons throughout the year, and lessons target multiple areas of reading (e.g., decoding, fluency, vocabulary, comprehension). A variety of lesson delivery methods are used, including whole class, teacher-led, or small group instruction, and individual computer work and in-program assessments are used to track student progress.

It may be that small effects over the course of a single year are what middle schools should expect when they implement reading interventions for struggling readers. For students who are well below grade level in reading, multiple years of intervention may be necessary for them to reach grade level expectations (cf., Dougherty, 2015). If small, yearly effects across multiple years are necessary for students to reach grade-level expectations, then surely the evaluation and monitoring components are as critical as the interventions themselves. In the absence of credible evidence that small to moderate improvements are being realized each year, the value of the intervention for students is questionable.

Engagement outcomes in middle school. The multi-component intervention did not have an impact on students' self-reported engagement in school. Based on meta-analyses suggesting that comprehensive interventions delivered in middle school can improve students' connection to school (e.g., Dynarski et al., 2008), our partner districts reasoned that together, school engagement supports and reading intervention could impact students' perceived school engagement. We offer some potential reasons why the two components did not realize that hypothesis. First, we suspect the level of intensity, on average, of the engagement supports schools implemented may not have been sufficient to result in students developing stronger connections to school (Rumberger & Larson, 1998; Sinclair et al., 1998). Although all students in the treatment condition received targeted engagement support, schools determined supports based on interventions available; few schools committed resources to consistently provide the kind of intense engagement interventions research suggests is effective. For example, few schools widely implemented evidence-based engagement interventions, such as check-in/check-out programs. Moreover, schools could provide engagement supports to comparison students, and many of these students received them. The intensity of the engagement supports and how widely those supports were used may help explain the limited impact on student engagement.

Several other factors may have contributed to the non-significant engagement effects. One important issue is the way engagement was measured. We used the SEI, partly because there was reasonable evidence for its use as a measure of perceived engagement, and partly because we believed the constructs being measured (cognitive, behavioral, and psychological engagement) would allow us to assess important facets of school engagement in middle school. The 35 items and corresponding factors appear to address engagement in a clear, direct way, and the internal psychometrics of the instrument are sound. Further, there is evidence the measure is related to academic achievement, including grades and reading proficiency (Appleton et al., 2006). However, in this study, there was little variability in scores among students, a constraint that limited our ability to detect differences in student engagement due to the intervention. It may be that a more sensitive instrument is needed to detect such differences. The SEI also does not specifically address perceived levels of academic achievement or engagement in reading. Second, reading interventions may need to demonstrate a stronger overall impact on reading achievement than occurred in this study to impact a more distal outcome, such as student reported engagement in school, or similarly, the impact of the reading intervention on student engagement could be indirect. It is possible more effective reading interventions would positively impact reading outcomes, and students' awareness of this positive impact (e.g., that their reading skills were improving), would result, in turn, in a positive impact on perceived engagement in school.

Limitations

There are several limitations to note regarding this study. First, because there was considerable variability among the types of intervention programs that were used across the 25 participating schools, it makes specifying the precise cause of the nonsignificant finding on

OAKS and the positive interaction effect on the PRF measure difficult. It is not possible to conclude that a specific program(s) caused the overall outcomes observed. Interpretation requires considering what was implemented, which, in terms of the reading intervention, was largely measured through direct observation. In this context, the challenge to interpretation is the number of implementation factors that can be considered across schools. However, it is clear that the intervention was substantively different from the instruction provided in the comparison condition. In addition, the data indicate intervention students received more reading instruction than comparison students. Thus, we can say that intervention efforts targeting issues related to dosage and frequency are not likely to be sufficient in and of themselves to improve reading outcomes for struggling students. We are conducting additional analyses to determine the extent to which issues related to program use, student characteristics, and intervention delivery (e.g., the quantity and quality of student-teacher interactions) are associated with student outcomes.

A second limitation relates to statistical power. In our analyses, we tested for variability in RD effects due to variation in school cut points, in addition to testing for residual school level variability from all other sources. Low power is an issue with RD analyses compared to RCTs, and although we had sufficient power to detect small overall effects, power to detect significant RD effects within individual schools or predictors of school level variation in RD effects is more difficult. Thus, we limited the scope of our current efforts to identifying significant residual school level variability if it existed. We took this first step because we knew that schools operationalized reading and engagement interventions very differently from one another (within and across districts). Future research will further explore estimating RD effects in individual schools and exploring potential school level predictors of the residual school level variability.

A third limitation was our need to use a fuzzy, rather than sharp, RD design. We planned for and observed a certain amount of non-compliance by partner districts, and our primary objective was to evaluate effects when interventions were implemented in authentic school settings without external resources for implementation. This required the use of a fuzzy RD design, which nonetheless decreased our power to detect differences between groups. Use of a fuzzy design, coupled with the study's relatively small sample of schools, means we may have been underpowered to detect small (< .20), but significant effects of the intervention on student outcomes. In fact, we observed a non-significant effect on OAKS–R (effect size = .15) that was comparable to the effect sizes observed in other, similar studies (e.g., Boulay et al., 2015).

References

Alonzo, J., & Tindal, G. (2008). *Examining the technical adequacy of fifth-grade reading comprehension measures in a progress monitoring assessment system* (Technical Report No. 0807). Eugene, OR: Behavioral Research and Teaching, University of Oregon

Alonzo, J., Tindal, G., Ulmer, K., & Glasgow, A. (2006). *easyCBM online progress monitoring assessment system*. Eugene, OR: University of Oregon. Retrieved from http://easycbm.com

Anderson, D., Alonzo, J., & Tindal, G. (2010). *easyCBM® mathematics criterion related validity evidence: Oregon state test* (Technical Report No. 1011). Eugene, OR: Behavioral Research and Teaching, University of Oregon. Retrieved from http://www.brtprojects.org/publications/dl/1014

Appleton, J. J., Christenson, S. L., Kim, D., & Reschly, A. L. (2006). Measuring cognitive and psychological engagement: Validation of the student engagement instrument. *Journal of School Psychology*, *44*, 427–445. doi: 10.1016/j.jsp.2006.04.002

Auman, M. E. (2011). *Step up to writing*. Longmont, CO: Sopris West.

Baker, S. K., Smolkowski, K., Chaparro, E. A., Smith, J. L. M., & Fien, H. (2015). Using regression discontinuity to test the impact of a Tier 2 reading intervention in first grade. *Journal of Research on Educational Effectiveness*, *8*(2), 218-244.

Berk, R. (2011). Asymmetric loss functions for forecasting in criminal justice settings. *Journal of Quantitative Criminology, 27*, 107–123. doi: 10.1007/s10940-010-9098-2

Bloom, H. S. (2012). Modern regression discontinuity analysis. *Journal of Research on Educational Effectiveness*, *5,* 43–82. doi: 10.1080/19345747.2011.578707

Boulay, B., Goodson, B. D., Frye, M., Blocklin, M., & Price, C. (2015). *Summary of research generated by striving readers on the effectiveness of interventions for struggling adolescent readers* (NCEE No. 2016-4001). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/pubs/20164001/pdf/20164001.pdf

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York: Academic Press.

Crone, D. A., Baker, S. K., & Fien, H. (2014). *Middle School Intervention Project (MSIP): Technical report for data team tool for observation (dtto 2.0) from 2011-2012*. Eugene, OR: Center on Teaching and Learning, University of Oregon

Crone, D. A., Carlson, S. E., Haack, M. K., Kennedy, P. C., Baker, S. K., & Fien, H. (2016). Data-based decision-making teams in middle school: Observations and implications from the Middle School Intervention Project. *Assessment for Effective Intervention, 41*, 79–93. doi: 10.1177/1534508415610322

Crone, D. A., Hawken, L. S., & Horner, R. H. (2010). *Responding to problem behavior in schools, second edition: The behavior education program* (2nd ed.). New York, NY: Guilford Publications. Retrieved from https://books.google.com/books?id=wWnNsr2L808C

Deno, S. L. & Mirkin, P. K. (1977) Data-based Program Modification: A Manual. Reston VA: Council for Exceptional Children.

Dimmery, D. (2016). rdd: Regression Discontinuity Estimation. R package version 0.57. Retrieved from https://CRAN.R-project.org/package=rdd

Dole, J. A., Valencia, S. W., Greer, E. A., & Wardrop, J. L. (1991). Effects of two types of prereading instruction on the comprehension of narrative and expository text. *Reading Research Quarterly, 26*, 142–159. doi: 10.2307/747979

Dougherty, S. M. (2015). Bridging the discontinuity in adolescent literacy? Mixed evidence from a middle grades intervention. *Education Finance and Policy, 10*, 157–192. doi: 10.1162/EDFP_a_00157

Dynarski, M., Clarke, L., Cobb, B., Finn, J., Rumberger, R. W., & Smink, J. (2008). *Dropout prevention: A practice guide* (Technical Report No. NCES 2008-4025). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, US Department of Education

Dynarski, M., Gleason, P. M., Rangarajan, A., & Wood, R. (1998). *Impacts of dropout prevention programs* (Final Report No. 8014) (A Research Report from the School Dropout Demonstration Assistance Program Evaluation). Princeton, NJ: Mathematica Policy Research.

Edmonds, M. S., Vaughn, S., Wexler, J., Reutebuch, C., Cable, A., Tackett, K. K., & Schnakenberg, J. W. (2009). A synthesis of reading interventions and effects on reading comprehension outcomes for older struggling readers. *Review of Educational Research, 79*, 262–300. doi: 10.3102/0034654308325998

Finn, J. D. (1989). Withdrawing from school. *Review of Educational Research, 59*, 117–142.

Greene, J. F. (2004). *Language! A literacy intervention curriculum* (3rd ed.). Logmont, CO: Sopris West.

Guthrie, J. T., & Davis, M. H. (2003). Motivating struggling readers in middle school through an engagement model of classroom practice. *Reading & Writing Quarterly, 19*, 59–85. doi: 10.1080/10573560308203

Hamilton, L., Halverson, R., Jackson, S., Mandinach, E., Supovitz, J., & Wayman, J. (2009). *Using student achievement data to support instructional decision making* (NCEE 2009-4067).

Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Hasselbring, T. (2006). *Read 180*. Woodland Hills, CA: Scholastic.

Hernandez, D. J. (2011). *Double jeopardy: How third-grade reading skills and poverty influence high school graduation*: Department of Sociology Hunter College and the Graduate Center, City University of New York, Annie E. Casey Foundation

Ihnot, C., Mastoff, J., Gavin, J., & Hendrickson, L. (1992). *Read Naturally®*. St. Paul, MN: Read Naturally, Inc.

Imbens, G.W., & Kalyanaraman, K. (2012). Optimal bandwidth choice for the regression discontinuity estimator. *The Review of Economic Studies, 79*, 933–959. doi: 10.1093/restud/rdr043

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

Individuals with Disabilities Education Act, 20 U.S.C. § 1400 (2004).

Kamil, M. L., Borman, G., Dole, J., Kral, C. C., Salinger, T., & Torgesen, J. (2008). *Improving adolescent literacy: Effective classroom and intervention practices: A practice guide* No. NCEE #2008-4027). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, US Department of Education. Retrieved from http://ies.ed.gov/ncee/wwc

Kennedy, P. C., Schwartz, M. C., Baker, S. K., Crone, D. A., & Fien, H. (2013). *Middle School Intervention Project (MSIP) engagement intervention tracking: Project year 2* (Technical Report No. 1304). Eugene, OR: Center on Teaching and Learning, University of Oregon

Mandinach, E.B. (2012). A perfect time for data use: Using data-driven decision making to inform practice, *Educational Psychologist, 47*(2), 71-85.

Marmer, V., Feir, D., & Lemieux, T. (2014). *Weak identification in fuzzy regression discontinuity designs*. Vancouver, BC, Canada.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics, 142*, 698–714. doi: 10.1016/j.jeconom.2007.05.005

McLaughlin, M. J., Speirs, K. E., & Shenassa, E. D. (2014). Reading disability and adult attained education and income: Evidence from a 30-year longitudinal study of a population-based sample. *Journal of Learning Disabilities, 47*, 374–386. doi: 10.1177/0022219412458323

Meyer, L. A., Carnine, L., Becker, W. C., Eisele, J., & Johnson, G. (1989). *Corrective reading*. Chicago, IL: SRA.

National Center for Education Statistics. (2011). *The nation's report card: Trial urban district assessment reading 2011* No. NCES 2012-455). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education

National Center for Education Statistics. (2015). *The nation's report card: 2015 mathematics & reading assessements. National achievement level results, 8th grade*. Washington, DC: National Center for Education Statistics. Retrieved from http://www.nationsreportcard.gov/reading_math_2015/#reading/acl?grade=8

Neild, R. C., & Balfanz, R. (2006). *Unfulfilled promise: The dimensions and characteristic's of Philadelphia's dropout crisis, 2000-2005*. Philadelphia, PA: Philadelphia Youth Network; University of Pennsylvania; Johns Hopkins University. Retrieved from http://files.eric.ed.gov/fulltext/ED538341.pdf

Nelson, N. J., Baker, S. K., Kennedy, P. C., & Fien, H. (2015, February). *Exploratory analyses of data patterns in a large-scale rd evaluation*. Paper presented at the National Association of School Psychologists Annual Convention, Orlando, FL.

Nelson-Walker, N. J., Turtura, J., Cohen, J., Kennedy, P. C., Crone, D. A., Baker, S. K., & Fien, H. (2012). *Middle School Intervention Project: Classroom observation tool (MSIP-cot) 2011-2012 technical report*. Eugene, OR: Center on Teaching and Learning, University of Oregon

No Child Left Behind Act of 2001, Pub. L. No. 107-110, § 115, Stat. 1425 (2002).

Nomi, T., & Allensworth, E. (2009). “Double-dose” algebra as an alternative strategy to remediation: Effects on students' academic outcomes. *Journal of Research on Educational Effectiveness, 2*, 111–148. doi: 10.1080/19345740802676739

Oregon Department of Education. (2012). *Oregon's statewide assessment system. Test administration* (Technical Report Volume No. 5). Salem, OR: Author.

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org/

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks: Sage Publications.

Rumberger, R. W., & Larson, K. A. (1998). Student mobility and the increased risk of high school dropout. *American Journal of Education*, *107,* 1–35. doi: 10.1086/444201

Scammacca, N., Roberts, G., Vaughn, S., Edmonds, M., Wexler, J., Reutebuch, C. K., & Torgesen, J. K. (2007). *Interventions for adolescent struggling readers: A meta-analysis with implications for practice*. Portsmouth, NH: RMC Research Corporation, Center on Instruction.

Schochet, P. Z. (2008). *Technical methods report: Statistical power for regression discontinuity designs in education evaluations* No. NCEE 2008-4026). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, US Department of Education.

Sinclair, M. F., Christenson, S. L., Evelo, D. L., & Hurley, C. M. (1998). Dropout prevention for youth with disabilities: Efficacy of a sustained school engagement procedure. *Exceptional Children*, *65*, 7–21.

Slavin, R. E., Cheung, A., Groff, C., & Lake, C. (2008). Effective reading programs for middle and high schools: A best-evidence synthesis. *Reading Research Quarterly*, *43*, 290–322.

Snyder, T. D., & Dillow, S. A. (2013). *Digest of education statistics 2012* (NCES No. 2014-015). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://nces.ed.gov/pubs2014/2014015.pdf

Torgesen, J. K., Houston Miller, D., & Rissman, L. M. (2007). *Improving literacy instruction in middle and high schools: A guide for principals*. Portsmouth, NH: RMC Research Corporation, Center on Instruction.

Wanzek, J., Vaughn, S., Scammacca, N. K., Metz, K., Murray, C. S., Roberts, G., & Danielson, L. (2013). Extensive reading interventions for students with reading difficulties after grade 3. *Review of Educational Research, 83*, 163–195. doi: 10.3102/0034654313477212

Wong, V. C., Cook, T. D., Barnett, W. S., & Jung, K. (2008). An effectiveness-based evaluation of five state pre-kindergarten programs. *Journal of Policy Analysis and Management, 27*, 122–154. doi: 10.1002/pam.20310

Wood, D. E. (2006). Modeling the relationship between Oral Reading Fluency and performance on a statewide reading test. *Educational Assessment*, *11*, 85–104.

Appendix

Fuzzy RD Model Equations

To establish notation and variable names, estimated coefficients are italicized, variables are bolded and i and j subscripts indicate individual schools and students respectively. Parametric terms are indicated by *B* and smooth functions are indicated by *S*. L is a binary indicator variable that is 1 if the student is lower than or equal to the school cut point, P, and 0 otherwise, and C is the cut score. The equation for the multilevel GAM for the RD outcome gap is:

$$\eta_{i,j} = B_{0,i} + B_{1,i}\mathrm{L}_{i,j} + B_2\mathrm{P}_i + B_3\mathrm{P}_i\mathrm{L}_{i,j} + S_1\left(\mathrm{C}_{i,j} - \mathrm{P}_i\right)\mathrm{L}_{i,j} + S_2\left(\mathrm{C}_{i,j} - \mathrm{P}_i\right)\left(1 - \mathrm{L}_{i,j}\right) + \varepsilon_{i,j}$$

for the j^th student within the i^th school, where eta is the linear predictor and is related to the outcome, $y_{i,j}$, through the inverse of the link function.

For all outcomes, the link function was simply the identity function, a standard choice for a continuous outcome. The random school effects for school intercepts and RD gaps are respectively $B_{0,i}$ and $B_{1,i}$ and are assumed to be bivariately normally distributed. The within school residual is epsilon, which is assumed to be normally distributed and uncorrelated with the school effects. The model includes different smooth functions for above (i.e., S_2 only applies to students above the cut point because of the (1-$\mathrm{L}_{i,j}$) term) and below (i.e., S_1 only applies to students below or on the cut point because of the $\mathrm{L}_{i,j}$ term) the cut point with separate sets of smoothing parameters. The equation for the multilevel GAM for the treatment probability is similar but does not include terms for the school cut point:

$$\eta_{i,j} = B_{0,i} + B_{1,i}\mathrm{L}_{i,j} + S_1\left(\mathrm{C}_{i,j} - \mathrm{P}_i\right)\mathrm{L}_{i,j} + S_2\left(\mathrm{C}_{i,j} - \mathrm{P}_i\right)\left(1 - \mathrm{L}_{i,j}\right)$$

for the jth student within the ith school. The treatment probability model used the logistic link function, a standard choice for binary outcomes, which eliminates the need for a within school residual effect.

Table 1

Critical Features of Each Component of the Intervention

Reading Intervention	School Engagement Support	School Data Teams
1. Procedures are in place to identify students who need reading intervention. 2. Instructors are trained to deliver the intervention. 3. Intervention is delivered for the equivalent of 20–30 minutes daily, five days per week, for at least one full academic term. 4. Procedures are in place to monitor student progress on targeted skills based on need. 5. Reading intervention is distinct and different from the reading instruction and support comparison students receive.	1. Students check-in with a supportive adult on a regular basis; frequency depends on the intensity of the student's needs. 2. Students receive regular, constructive feedback on behavioral and/or academic performance. 3. Students are recognized and reinforced for demonstrating improvement. 4. Data are collected on each student's behavioral and/or psychological engagement in school. 5. Intervention is determined based on available school options and matched to student need.	1. Systematic process is in place for reviewing data and making decisions about students at risk for reading or school engagement problems. 2. School-based teams exist and include key stakeholders. 3. Data team meets at least once per academic term. 4. Data reviewed is relevant, formative, and current.

Table 2

Descriptive Statistics by Group Status for the Full Sample and For Students Near the Cut Point

Variable	*n*	*M*	*SD*	*Min*	Q25	Q50	Q75	*Max*
Full sample								
Cut Score								
Comparison	4397	0.95	0.70	-1.98	0.43	0.88	1.40	3.54
Intervention	1495	-0.48	0.57	-2.98	-0.76	-0.40	-0.11	2.39
OAKS Reading								
Comparison	4324	237.53	8.23	212.00	232.00	237.00	242.00	277.00
Intervention	1441	225.36	7.21	202.00	221.00	226.00	230.00	248.00
Passage Reading Fluency								
Comparison	4304	178.42	34.56	28.00	153.00	177.00	200.25	303.00
Intervention	1431	127.48	31.34	19.00	105.00	128.00	149.00	247.00
Student Engagement Instrument								
Comparison	4239	104.79	13.54	6.00	97.00	106.00	114.00	132.00
Intervention	1408	104.39	14.42	33.00	97.00	106.00	114.00	132.00
Students within .50 SDs of the cut point								
Cut Score								
Comparison	1194	-0.31	0.29	-1.45	-0.47	-0.28	-0.11	0.52
Intervention	817	-0.66	-0.32	-1.53	-0.85	-0.67	-0.46	0.50
OAKS Reading								
Comparison	1174	231.77	5.27	214.00	229.00	232.00	235.00	248.00
Intervention	791	227.84	5.89	208.00	224.00	228.00	231.00	248.00
Passage Reading Fluency								
Comparison	1170	150.51	23.13	72.00	136.00	151.00	162.00	239.00
Intervention	789	139.66	24.43	54.00	124.00	140.00	155.00	229.00
Student Engagement Instrument								
Comparison	1143	104.84	13.94	38.00	98.00	107.00	114.00	132.00
Intervention	772	104.17	14.30	33.00	96.75	106.00	114.00	132.00

Note. Q25 = 25th Quartile; Q50 = 50th Quartile; Q75 = 75th Quartile; OAKS = Oregon Assessment of Knowledge and Skills. The cut score represents the average of standardized z-scores based on a student's highest OAKS score from 6th grade and a single Passage Reading

Fluency passage administered in the spring of 6th grade and serves as a composite measure of baseline reading proficiency. All other measures represent 7th grade outcomes.

Table 3

Reading Implementation and Inclusion in Discussions of School Data Teams by Condition for Students Near the Cut Point

Variable	Comparison (*n* = 1039)				Intervention plus ELA (*n* = 550)				Intervention only (*n* = 86)			
	M	*SD*	*Min*	*Max*	*M*	*SD*	*Min*	*Max*	*M*	*SD*	*Min*	*Max*
Reading implementation												
Dosage—hours of reading instruction												
English Language Arts	160.63	48.63	31.29	317.08	140.44	47.22	13.69	256.69	–	–	–	–
Reading Intervention	–	–	–	–	92.83	38.40	8.32	239.76	237.00	67.26	16.47	381.17
Dosage—weeks of reading instruction												
English Language Arts	39.77	1.84	23.29	40.29	38.16	6.11	8.29	40.29	–	–	–	–
Reading Intervention	–	–	–	–	32.62	10.40	3.29	40.29	38.96	4.51	12.29	40.29
Number of published programs used												
English Language Arts	0.90	1.07	0	6	0.66	0.81	0	3	–	–	–	–
Reading Intervention	–	–	–	–	0.95	0.79	0	4	1.13	0.57	0	3
Proportion of programs used that were published												
English Language Arts	0.48	0.50	0	1	0.47	0.49	0	1	–	–	–	–
Reading Intervention	–	–	–	–	0.68	0.46	0	1	0.89	0.30	0	1
Average class size												
English Language Arts	30.89	4.09	14	40	30.10	5.52	6	40	–	–	–	–
Reading Intervention	–	–	–	–	17.05	5.70	5	29	21.48	6.32	7	29
School data teams												
Number of students discussed	1.17	1.43	0	6	4.37	5.14	0	24	0.30	1.20	0	10
Time per student (in minutes)	1.79	2.48	0.02	20.40	1.96	4.81	0.02	51.22	0.38	0.33	0.03	1.22

Table 4

Engagement Supports by Condition for Students Near the Cut Point

Variable	Percent of students in each category		
	Comparison (*n* = 1039)	Intervention plus ELA (*n* = 550)	Intervention only (*n* = 86)
Support type			
Academic	0.87%	26.00%	27.91%
Mentoring/Check-in	2.86%	71.09%	86.05%
Social skills	0.20%	9.46%	40.70%
Student interest	16.36%	39.64%	67.44%
Meeting frequency			
Daily	17.61%	54.36%	86.05%
More than 1/week	1.06%	29.46%	33.72%
1/week	1.25%	32.55%	27.91%
More than 1/month	0.29%	16.36%	3.49%
1/month	0.09%	7.27%	0.00%
Less than 1/month	0.00%	0.00%	34.88%

Note. Students could choose to participate in more than one engagement activity, and were included in the percentages for all activities they participated in. Academic engagement supports included study hall, tutoring, and activities oriented around math, reading and writing. Mentoring engagement supports included formal check in programs, counseling, and student mentoring. Social skill engagement supports included social skills training, volunteering, and service activities. Student interest engagement supports included arts, music, technology, and sports.

Table 5

Parameter Estimates (Est.) Standard Errors (SE), and Tests of Significance (p) for RD Models Tested

			OAKS			PRF			SEI			Treatment Prob.		
	Parameter	Term	Est.	SE	*p*	Est.	SE	*p*	Est.	SE	*p*	Est.	SE	*p*
Bandwidth Approach														
Optimal IK			0.38	1.00	.70	-2.60	3.48	.45	1.52	1.95	.44			
Half-Optimal IK			-0.46	1.35	.74	-3.18	4.32	.46	3.40	2.86	.24			
Double-Optimal IK			0.84	0.75	.26	-0.84	2.26	.71	0.61	1.58	.70			
Meta-Analytic w/Optimal IK			-0.13	0.83	.88	-2.80	2.73	.31	0.41	1.54	.79			
GAM Fixed Effects														
Control Mean at Cut Point	E(B0)	Constant	232.98	0.55	.00	161.57	2.45	.00	105.51	1.33	.00	-1.41	0.22	.00
Below Cut Point (Raw RD Gap)	E(B1)	L	0.76	0.46	.10	-4.45	2.07	.03	0.17	0.81	.83	3.05	0.31	.00
School Cut Point	B2	P	6.76	1.02	.00	33.42	3.88	.00	2.42	2.17	.27			
School Cut Point by Below Cut Point	B3	L * P				-7.26	2.98	.01						
Fuzzy RD Gap			1.18	0.73	.10	-6.94	3.27	.03	0.27	1.26	.83	0.64	0.05	.00
GAM School Level (Co)Variance Parameters														
Control Mean at Cut Point	Var(B0)		1.28			17.45			8.32			0.77		
On or Below Cut Point	Var(B1)		1.97						4.63			1.56		
Control Mean with Below Cut Point	Cor(B0, B1)		-0.11						-0.65			-0.73		
Within School Residual	Var(epsilon)		3.14			488.70			181.80					
			EDF	Chi	*p*	EDF	Chi	*p*	EDF	Chi	*p*	EDF	Chi	*p*
GAM Smooth Functions														
On or Below Cut Point	S1	C * L	2.29	331.90	.00	3.35	1439.40	.00	1.00	0.61	.44	1.00	30.54	.00
Above Cut Point	S2	C * (1 - L)	1.00	4141.30	.00	1.00	831.20	.00	1.00	0.29	.59	1.00	141.77	.00

Note. OAKS = Oregon Assessment of Knowledge and Skills; PRF = Passage Reading Fluency; SEI = Student Engagement Instrument.

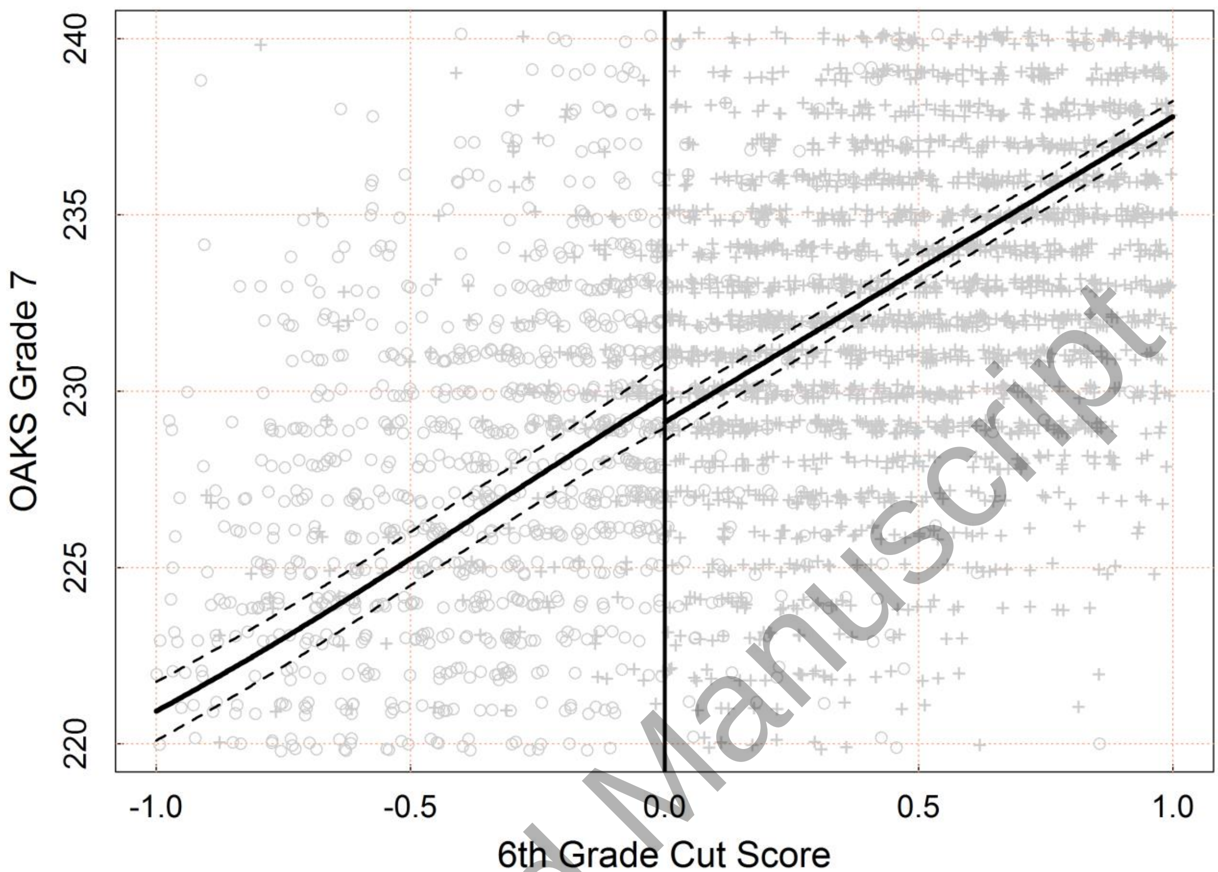

Figure 1. Fitted GAM (solid lines) with 95% pointwise confidence intervals (dashed lines) for the grade 7 OAKS reading score as a function of the cut score. Students who received the intervention are shown as circles and students in the comparison condition are shown as pluses. The vertical line at zero is the cut point. Non-compliers are circles to the right of the cut point and pluses to the left of the cut point.

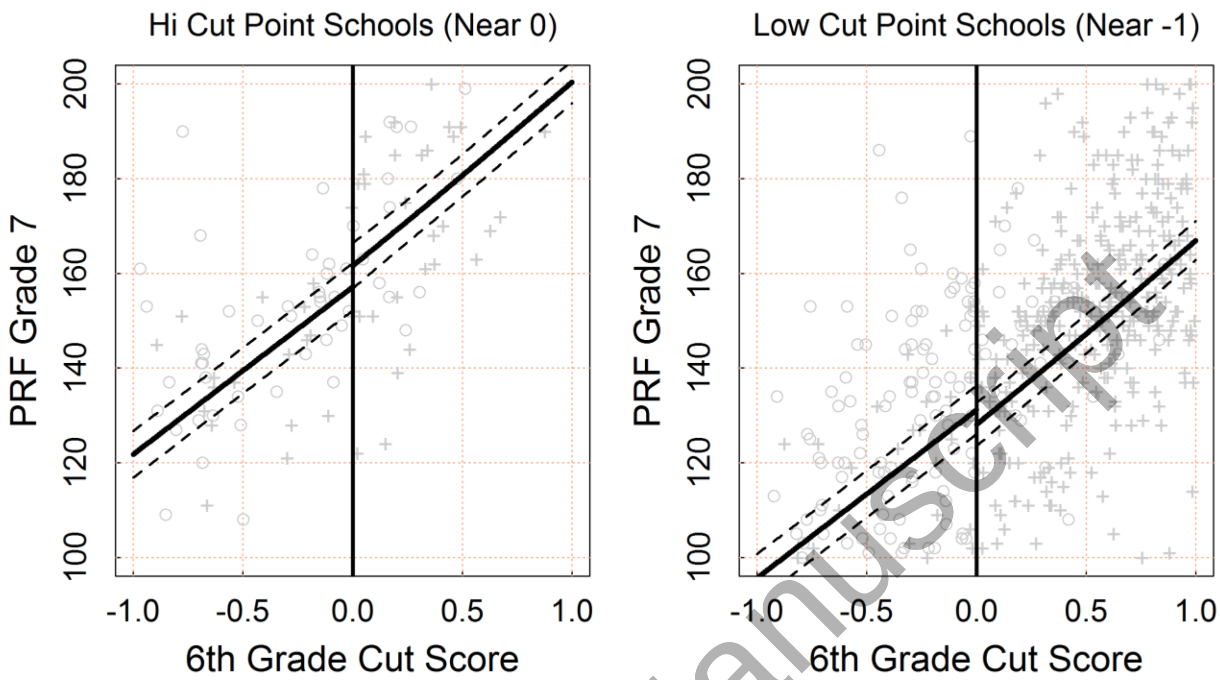

Figure 2. Fitted GAM (solid lines) with 95% pointwise confidence intervals (dashed lines) for the grade 7 PRF reading score as a function of the cut score. The interaction between school cut point and the outcome gap is illustrated by the difference between the fitted curves in the left panel, which are for schools with a cut point near 0, and the fitted curves in the right panel, which are for schools with a cut point near -1. Students who received the intervention are shown as circles and students in the comparison condition are shown as pluses. The vertical line at zero in both panels represents the cut point. Non-compliers are circles to the right of the cut point and pluses to the left of the cut point.