NATIONAL CENTER for ANALYSIS of LONGITUDINAL DATA in EDUCATION RESEARCH

TRACKING EVERY STUDENT'S LEARNING EVERY YEAR

A program of research by the American Institutes for Research with Duke University, Northwestern University, Stanford University, University of Missouri-Columbia, University of Texas at Dallas, and University of Washington

AIR

Do Students Benefit from Longer School Days? Regression Discontinuity Evidence from Florida's Additional Hour of Literacy Instruction

David Figlio

Kristian L. Holden

Umut Ozek

WORKING PAPER No. 201-0818-1 • August 2018

Do Students Benefit from Longer School Days? Regression Discontinuity Evidence from Florida's Additional Hour of Literacy Instruction

David N. Figlio
Northwestern University and NBER

Kristian L. Holden
American Institutes for Research

Umut Ozek
American Institutes for Research

Contents

Acknowledgements

This research was supported by the Smith Richardson Foundation and the Institute for Education Sciences. We are grateful to seminar and conference participants and discussants at AEFP and APPAM for helpful comments and suggestions. We would also like to thank the Florida Departments of Education and Health for providing the matched data used in this analysis. All opinions expressed in this paper are those of the authors and do not reflect the views of the Florida Departments of Education and Health or our funders.

CALDER • American Institutes for Research
1000 Thomas Jefferson Street N.W., Washington, D.C. 20007
202-403-5796 • www.caldercenter.org

Do Students Benefit from Longer School Days? Regression Discontinuity Evidence from Florida's Additional Hour of Literacy Instruction

David N. Figlio, Kristian L. Holden, Umut Ozek

CALDER Working Paper No. 201-0818-1
August 2018

Abstract

Instructional time is a fundamental educational input, yet we have little causal evidence about the effect of longer school days on student achievement. This paper uses a sharp regression discontinuity design to estimate the effects of lengthening the school day for low-performing schools in Florida by exploiting an administrative cutoff for eligibility. Our results indicate significant positive effects of additional literacy instruction on student reading achievement. In particular, we find effects of 0.05 standard deviations of improvement in reading test scores for program assignment in the first year, though long-run effects are difficult to assess.

Keywords: instruction time; extended school days; inequality

1. Introduction

Education is one of the most important determinants of improved economic and social outcomes.[1] As such, education policy in the United States over the last few decades has focused on raising the academic performance of low-performing schools to improve the outcomes of disadvantaged students. Extending the school day has been a popular, yet controversial, policy to improve achievement of low-performing students. Currently, several states (e.g., Florida, Colorado, Connecticut, and Massachusetts) and many large school districts (e.g., New York City, Chicago, Boston, and the District of Columbia) have longer school day policies targeting low-performing schools. This policy is potentially appealing because instructional time is a fundamental resource in education, and not surprisingly, researchers have long described the theoretical importance of instructional time for student outcomes.[2] Additional instructional time for low-achieving students certainly has intuitive appeal - struggling students may simply need more time to learn. On the other hand, such policies are expensive, and the costs may outweigh the potential gains in achievement if the benefits of additional instructional time are mitigated by low teacher quality, disadvantaged school settings, or failure to efficiently use the additional time.

Empirically, we know little about the effectiveness of longer school days. Many studies have explored the correlation between longer school days and achievement, but schools that implement these policies are likely very different from others in terms of observable and unobservable characteristics. For example, high-performing schools may have more resources, and these resources allow high-performing schools to provide more instructional time, so that simple comparisons would overstate the effect of additional instructional time on achievement. Similarly, extended school day

[1] See Card (1999) for a thorough review of the literature on education and earnings.
[2] Carroll (1963) introduced the instructional model, which focused on time needed for learning and time devoted to learning. This lead to the "mastery model" of learning (Bloom, 1968; Block and Burns, 1976), and subsequent work by Levin and Tsang (1987), which introduced complications in the form of the student's capacity for learning, effort, total available time, and school resources.

policies are typically tied to other school-level interventions, making it difficult to isolate the causal effect of additional instruction time on student outcomes.

In this study, we contribute to a growing literature that attempts to address these potential sources of bias by exploiting school-level administrative cutoffs that determine whether students receive additional instructional time in a regression discontinuity (RD) framework. In 2012, Florida introduced one of the most ambitious instructional time programs in the United States by increasing the length of its school day in the lowest-performing elementary schools by an hour in order to provide additional literacy instruction. For the first year of the extended school day (ESD) program, the lowest 100 elementary schools (out of roughly 1,800 elementary schools in Florida) were selected according to an index of school-level reading accountability measures.[3] Our RD framework leverages this index to compare the academic outcomes of students who participate in the program to students who have similar observable characteristics around the administrative cutoff.

We find statistically and educationally significant benefits of the extended school day on reading test scores. In particular, RD estimates suggest that students enrolled in schools whose reading accountability scores fell immediately below the ESD cutoff (and hence were required to provide additional instruction) score roughly 0.05σ better in reading compared to students enrolled in schools just above the cutoff. Elementary students in Florida typically gain about 40 percent of a standard deviation per year on the current-year reading test[4], so these effect sizes represent gains comparable to about a month of instruction, which is almost equivalent to the additional instruction time introduced by the ESD program in these schools. That said, we find evidence that these benefits are lower for

[3] The ESD program also selected 100 schools in 2013-2014, and 300 schools in subsequent years. Due to a change in the state's assessment regime, we do not currently have access to student-level data from school years beyond 2012-13, so we focus on the first year of the program, and provide a more limited school-level analysis of later years.

[4] Based on authors' calculations using vertically-aligned developmental reading scores for students in Florida public elementary schools in 2010-11 and 2011-12 school years. Specifically, we calculated the average year-to-year change (difference between year $t+1$ and year t) in developmental reading scores for 3rd and 4th graders in terms of the standard deviation of the current year (year t) reading test scores.

students with the lowest reading skills (students who scored in the lowest achievement level in the prior year reading test), and the program was more beneficial for kids with basic, yet limited, skills in reading.

2. Prior Literature

What we know about extended school day policies comes mostly from descriptive research, and a relatively smaller number of studies focused on the causal effects of extending the school day on student outcomes.[5] For instance, Bellei (2009) uses a difference-in-differences approach to study a policy in Chile that selects high schools to transition from part-time to full-time. Bellei's results suggest an improvement in language but provide less evidence on math achievement. Jensen (2013) uses a school fixed-effects model to estimate the effects of a Danish policy that narrowed gaps in classroom hours, and finds positive effects for math but no significant effect for literacy. Lastly, Battistin and Meroni (2016) use a difference-in-differences approach to study a reform in southern Italy that extended the school day at selected low-performing schools. These authors find positive effects for math but not for literacy.

Estimating the causal effects of extending the school day on student outcomes is difficult due to two major empirical challenges. First, schools with longer days are likely to differ from others along unobserved dimensions, and the observed differences in student outcomes will therefore under- or overstate the true effects of additional instructional time. An example especially relevant for disadvantaged students is that school districts providing additional instructional time may also have greater financial resources and likely have better student outcomes even in the absence of additional instructional time.[6] It is also possible that parents with high-performing children may select districts with greater resources, again, introducing bias into estimates of additional instructional time.

[5] See Patall, Cooper, and Allen (2010) for a review of older work.
[6] Other forms of bias are possible as well, including nonrandom selection of students or schools into education programs that provide additional instructional time.

Second, extended school day policies are usually inseparably tied to other school policies so that even compelling identification strategies cannot separate the improvement due to additional instructional time from other policy elements. For example, many School Improvement Grant (SIG) schools increase instructional hours as part of school turnaround models, but this effect cannot be separated from other features of turnaround models, such as additional funding, professional development, and changes in leadership. Similarly, effective charter schools tend to have longer class days, but their practices also differ from traditional public schools in other ways, such as teacher retention and school leadership policies, and these schools may also draw more proficient students.[7]

This paper attempts to address both of these challenges by focusing on the ESD program in Florida. First, the features of the Florida ESD program allow us to use a RD framework in which we compare schools with reading accountability scores immediately below the cutoff (and hence were required to extend the school day) with observationally similar schools with scores immediately above the cutoff. Second, the ESD program focuses specifically on additional instructional time and does not include additional interventions the effects of which would be impossible to separate.

While there is little research on longer school days with which to compare our estimates, there is more causal work in several related streams of research. For instance, substantially more research focuses on extending the school year. Early studies use cross-state variation in the length of the school year, and cross-country variation in overall instruction time.[8] Recent research on extended school years attempts to exploit potentially exogenous variation in education policies. Pischke (2007) examines the impact of German short years and finds that shorter school years leads to more grade repetition and

[7] Dobbie and Fryer (2011) examine charter schools in New York City and find that instructional time, along with teacher feedback, high expectations, data-guided instruction, and high expectations, explain about 50 percent of the variation in school effectiveness. Farbman and Kaplan (2005) studied the Massachusetts 2020 project and found suggestive evidence that schools with additional instructional time had better student achievement.

[8] See Patall, Cooper, and Allen (2010) for a detailed review of the literature on extended school time between 1985 and 2009. Notably, they conclude that "the research designs are generally weak for making causal inferences". See Lavy (2015) for a discussion of cross-state and cross-country studies as well as a discussion about difficulties with identification, and Rivkin and Schiman (2015) for recent research using PISA data.

fewer students attending higher secondary school tracks, but finds no effect on earnings or employment. Parinduri (2014) uses a RD framework to estimate the effect of longer school years in Indonesia and finds positive effects on educational attainment and earnings. Aucejo and Romano (2016) examine a policy in North Carolina that provides variation in the number of days prior to testing, and find positive effects that are small relative to reducing absences. Similarly, Sims (2008) uses variation in the timing of standardized tests to study the effects of additional instruction days before examination, and finds positive effects on math achievement but not reading.

A separate thread of research has examined the effects of unexpected school closures. Marcotte (2007) and Marcotte and Hemelt (2008) find some evidence that weather-related closures cause decreases in achievement; in contrast, results from Goodman (2014) may cast some doubt on the effects of weather-related school closures. That said, it is important to study the effects of extended school days because it is reasonable to expect that the effects of these policies on student outcomes will differ from the effects of extended school year policies. For instance, while studies that make use of exogenous weather-related shocks have high internal validity, their results are less likely to be applicable to extended day policies under which the change in instruction time is known and therefore teachers can plan their instruction ahead of time.

Extended school day policies are also closely related to several other threads of research. Taylor (2014) uses a RD framework to examine the effects of requiring an additional hour of remedial math instruction without extending the school day. He finds large effects on achievement in math that decay over time, and finds that math instruction crowds out instruction in other subjects. Cortes, Goodman, and Nomi (2015) use a RD framework to study a policy that assigned 9th graders to an additional algebra course without extending the school day, and find positive impacts on test scores.[9] Another related

[9] Moreover, Cortes, Goodman, and Nomi (2015) note that test scores tend to understate attainment effects, and suggest the importance of long-run studies.

question is studied by Anderson and Walker (2015) who estimate effects for schools that change to a four-day school week and increase the length of the school day, and they find generally positive effects from this policy.[10] Finally, Machin and McNally (2008) examine the effects of a program in the United Kingdom called "the Literacy Hour" that changed the way literacy skills are taught (rather than the time devoted to English), and found significant benefits on the reading and English skills of primary school children.[11]

Three studies specifically focus on the effectiveness of the ESD program and suggest mixed or negative findings. West and Vickers (2014) were commissioned by the Florida Department of Education (FDOE) to examine student outcomes as well as the implementation of the ESD program. West and Vickers (2014) demonstrate that student outcomes improve for many ESD schools relative to student's previous performance, but mixed results when compared to similar non-participating schools.[12] Research by Folsom et al. (2016, 2017) note that mean reversion is an important concern when evaluating the ESD program. These schools are chosen based on test score measures, and given that test scores are noisy measures of student ability, we would expect some gains in test score performance for ESD schools independent of the program's effectiveness. They seek to address this concern by estimating the amount of expected mean reversion for ESD schools, and suggest that the ESD program did not improve student achievement.

[10] Extended instructional time is somewhat related to research on early childhood education and half-day vs full-day kindergarten. See Currie (2001), DeCicca (2007), and DeCicca and Smith (2013) for a discussion of the research on full-day kindergarten.

[11] A few of the studies mentioned above have concisely examined the heterogeneous effects of extended instruction time. Bellei (2009) found that rural students had larger positive effects, and that high-achieving students benefited more from extended school days. Similarly, Battishtin and Meroni (2016) find that the high-achieving students benefit more from extended school days. Parinduri (2014) also finds that rural students benefit more from extended school years, and additionally that female students benefit more than male students. Lastly, Cortes, Goodman, and Nomi (2015) find that students with below average reading skills benefit the most from an additional math course within the existing school day.

[12] The results of this study were provided as a presentation to the Senate Appropriations Subcommittee on Education. As such, there is relatively little detail about the methods used to estimate effects. The presentations states that students were compared to "students with the same FCAT scores in the year prior" and comparison schools were "Title 1 schools with the same school grade".

There are two potential concerns with the approaches used by West and Vickers (2014) and Folsom et al. (2016, 2017) that we attempt to address in this study. First, and as discussed above, our use of a RD framework allows us to demonstrate that schools just above the cutoff can likely serve as a valid counterfactual because they have similar observable characteristics and preprogram test performance. Second, the RD framework also provides a compelling way to address mean reversion that does not rely on assumptions about the distributions of school test scores (see Chay et al., 2005). Because schools near the cutoff have nearly identical preprogram test scores, they have similar differences in performance relative to the average Florida school and would have the same degree of mean reversion on average in the absence of the ESD program.

3. ESD Program

Florida's ESD program is one of the most ambitious instructional time interventions in the United States, both in terms of the number of schools affected and the amount of additional time provided in each school. In 2012, the state of Florida passed legislation requiring the lowest-performing elementary schools to extend the school day by an hour.[13] The set of ESD schools changes each year, and the size of the group has expanded in recent years: the lowest 100 elementary schools (out of roughly 1,800 elementary schools in Florida) were selected in 2012–2013 and 2013–2014; starting with the 2014–2015 school year, the ESD program has selected the lowest 300 elementary schools. In our analysis, we focus primarily on the effects of the program during the first year, 2012-2013, for reasons discussed in detail in Section 7.

The ESD program selects schools by using school-level reading accountability measures. All elementary schools are ranked according to the sum of points for "reading performance" and "annual

[13] The ESD program is ambitious in the amount of instructional time added. It greatly increases the length of the school day relative to other U.S. states. In 2008, Florida public elementary schools averaged 6.5 hours per day; the ESD program requires an additional hour, which is beyond the 99th percentile of the length of school day for elementary schools nationwide (7.2 hours).

learning gains in reading." Reading performance is determined by the percentage of students in the school who scored a "satisfactory" (also known as Level-3) on the Florida Comprehensive Assessment Tests (FCAT 2.0), and annual learning gains are determined by the percentage of students that make adequate gains in reading achievement levels. Schools with the lowest sum of these components are selected for the ESD program to provide an additional hour of literacy instruction each day. This discontinuity in ESD program assignment is the key element of our identification strategy.

Figure 1 presents the percentage of schools selected by the ESD program by their relative reading accountability score, and shows that selection for the program is almost perfect; all schools with scores greater than zero are not selected, while virtually all schools with zero or below are selected.[14] [15] This suggests that small differences in the relative reading accountability score produce potentially exogenous differences in exposure to the ESD program.[16] Most importantly, this variation in assignment to ESD in 2012-13 does not appear to be related to nonrandom selection on the part of schools, districts, or the state. As we report below, we carefully consider the balance of covariates and density tests, and verify that selection on observable characteristics does not appear to be causing bias in our results.

As noted above, this program is costly: district superintendents have estimated the annual cost of the program to be approximately $300,000–$400,000 per school, or about $800 per student. The

[14] Previous versions of this paper relied on a fuzzy RDD, as the Florida Department of Education revised accountability data for 2011-12, which included changes to the student gains variable used to rank schools. This analysis uses the initial student gains data used by Florida Department of Education to rank schools for the program, resulting in a sharp RDD.

[15] Only one school was ranked below the cutoff for 2012-13 and was not selected for treatment in 2012-13. A potential source of fuzziness in treatment is that high-performing students enrolled in ESD schools have the option of not participating in the additional hour. In particular, the ESD program provides this option to students who scored a Level-5 on the FCAT Reading test in the previous year, whereas students performing below this level must participate in the additional hour. However, only 2 percent of students enrolled in ESD schools in 2012-13 scored in the highest achievement level in reading in the previous year, and our results are not sensitive to excluding these high-performing students from our analysis.

[16] Several schools that were selected for the first year of the ESD program closed during our sample period. Two schools closed before the end of the 2012-13 school year, four closed in 2013-14, and three closed in 2014-15 and 2015-16. Similarly, four schools from the second year of the ESD program closed between 2014-15 and 2016-17. One possible concern is that school closures cause nonrandom attrition from the sample and cause bias; however, this accounts for about 0.6 percent of our sample in our primary specification, and both our falsification tests and primary outcomes are not sensitive to the exclusion of schools that close.

state has increased two sources of funding to help districts meet the cost of the program: most districts fund the ESD program through the state's Supplemental Academic Instruction (SAI) fund and the Research-Based Reading Instruction allocation. To support the program, Florida has increased the Research-Based Reading Instruction allocation to $75 million and the SAI fund to $642 million, and set aside a total of $30 million for the ESD program.[17] If the SAI and Research-Based Reading Instruction funds do not sufficiently cover all of the schools in a district, the district must use discretionary funds to cover the remaining deficit. While a formal cost analysis of this program would be very useful, the method of funding described above likely precludes such an analysis because schools are not directly apportioned a predetermined allocation, and districts may spend more SAI and Research-Based Reading Instruction funding on the program than what was set aside by the state.

Schools that participate in the ESD program provide an additional hour of instruction specifically for literacy instruction. The ESD program requires that the following components be included in this instruction: practice must be based on research; instruction must be adapted for student ability; instruction should include phonemic awareness, phonics, fluency, vocabulary, and comprehension; students must have guided practice; and students must read material from social studies, science, and math classes.

West and Vickers (2014) and Folsom et al. (2016, 2017) also investigate the ESD program's implementation using school/district surveys and principal interviews conducted by FDOE in the first three years of implementation (2012-13 through 2014-15). West and Vickers (2014) find broad compliance among schools with the state's requirements in the first year of implementation, and that the ESD schools generally provided instruction that was research based and differentiated based on student reading proficiency; the instruction tended to focus on integrated phonemic awareness, phonics, fluency, vocabulary, and comprehension; teachers used guided practice, error correction, and

[17] This allocation was increased to $90 million in later years of the ESD program when 300 schools were selected.

feedback; and teachers included social studies and mathematics text, reading, discussion, and writing in response to reading. Sixty-seven percent of schools reported providing training for teachers who provided the additional instruction, yet less than half of teachers were provided additional planning time.

Schools varied in how they incorporated the additional hour of literacy instruction into the school day. Folsom et al. (2016) use surveys from 2013-14 ESD schools and find that the majority schools chose to extend the end of the school day, in addition to rearranging of the instructional day to provide additional instruction at a different time (40 percent), many schools chose to extend the end of the school day without rearranging (37.6 percent), and substantially fewer schools chose to rearrange the instructional day without starting earlier or ending later (14.1 percent).[18] On the surface, this appears to contrast with previous investigation by West and Vickers (2014), though the studies focus on distinct questions. West and Vickers (2014) surveyed on the *exact timing of the hour* and report that 32 percent of schools implemented the hour "during the school day". This is likely consistent with Folsom et al. (2016) if schools implement the hour of special reading instruction "during the school day" and either start earlier or end later.[19]

Schools also varied in how students were grouped during the additional hour of instruction and how the additional hour was staffed. Folsom (2016) finds that in 69 percent of ESD schools in 2013-14 used a combination of small and large group instruction and 28 percent used small group instruction exclusively. More than half of the schools (56 percent) reported grouping students by ability during the additional hour and 37 percent used a combination of grouping students by ability and grouping students with mixed abilities. 30 percent of schools reported using the students' regular classroom

[18] Folsom et al. (2016) report that the remaining schools chose earlier start times (4.7 percent), earlier start times and rearranging the day (2.4 percent), and a combination of earlier start times, later end times, and rearranging the day (1.2 percent).
[19] For example, a school may report providing the special instruction at 10:00am (and be identified as "during the school day"), and end the school day an hour later to provide the displaced instruction time.

teachers to provide the additional hour of instruction and 58 percent reported using a combination of classroom teachers and other staff such as reading coaches. Finally, virtually all schools that were identified as an ESD school for the first time reported hiring additional staff (e.g., reading coaches, teachers, paraprofessionals, or volunteers) for the additional hour.

Another important finding from these implementation studies is that ESD schools differ significantly from other schools with lower enrollment, greater proportion of minority students, and greater proportion of students eligible for free and reduced price lunch (Folsom et al., 2016, 2017). As such, simple comparisons between students at ESD and non-ESD schools will likely be biased. We address this concern directly by demonstrating that schools *near the cut point* tend to have similar school demographics.

4. Data

In our analysis, we rely on student-level administrative data covering school years between 2005-06 and 2012-13. These data contain reading scores for all students between grades 3 and 10, and also include a wealth of student characteristics including student demographics (e.g., race/ethnicity, gender), free-or-reduced-priced lunch (FRPL) eligibility, limited English proficiency status, and special education status. We are also able to link these student-level administrative data with student birth records for every child born in Florida between 1992 and 2002 and who subsequently attended a public school in Florida. Birth records offer background information not typically seen in school records such as maternal education, mother's age and marital status at the time of birth, and mother's country of origin, providing additional measures of student socioeconomic status beyond FRPL eligibility.

Table 1 presents the descriptive statistics for students enrolled in schools chosen for the ESD program (ESD schools) and non-ESD schools in 2012-13 school year. Comparisons between elementary school students enrolled in ESD and non-ESD schools reveal that the former group of students are significantly more likely to come from academically and socioeconomically disadvantaged backgrounds. In particular, students in ESD schools in the first year of the program scored roughly 66 percent of the standard deviation worse in reading and 57 percent worse in math in the prior year, were 30 percentage points more likely to be classified as FRPL eligible in 2012-13, 45 percentage points less likely to be persistently classified as FRPL eligible since they entered the school system, and more than three times more likely to be African American. Further, mothers of students in non-ESD schools were more than three times more likely have some college education or higher, and half as likely to be married and twice as likely to be a teenager (younger than 20 years old) at the time of birth compared to the mothers of students enrolled in ESD schools. These differences illustrate the fundamental challenge in revealing the causal effect of longer school days on student achievement – unobserved differences between students receiving additional instruction and other students might lead to biased estimates. In this study, we utilize the non-linearity created by the ESD program and compare students whose schools fell immediately below and immediately above the ESD cutoff in a regression discontinuity framework. In what follows, we detail this empirical approach.

5. Empirical Strategy

Let ESD_{st} denote an indicator variable that equals one if school *s* is selected for the ESD program in year *t* (which equals 2012-13 in our main analysis), and r_{st} denote the difference between the reading accountability score of school *s* and the cutoff that determines the highest-ranked school that will participate in ESD in year *t*, with negative values indicating scores below cutoff. In this case, the regression model representation of this evaluation problem becomes:

$$Y_{ist} = \alpha + \beta ESD_{st} + \varepsilon_{ist} \tag{1}$$

where Y_{ist} is the test score of student *i* enrolled in school *s* in year *t*, standardized to zero mean and unit variance at the grade level. Because ESD program is a reading intervention, we use reading test scores as our primary outcome of interest.[20] Provided that $E[\varepsilon|r]$ is continuous at the treatment cutoff, the causal effect of additional instruction on test scores is given as:

$$\beta = \lim_{r \uparrow 0} E[Y|r] - \lim_{r \downarrow 0} E[Y|r] \tag{2}$$

There are several ways to estimate β in this context. First is to estimate equation (2) non-parametrically using kernel-weighted local polynomial smoothing (Hahn et al, 2001; Porter 2003). This method reduces the possibility of misspecification bias in parametric models and achieves the optimal rate of convergence.

When the selection variable is discrete, as in this case, non-parametric estimator might lead to biased estimates as it is not feasible to compare averages within arbitrarily small neighborhoods around the cutoff (Lee and Card (2008)). Therefore, following Lee and Card (2008), we estimate equation (2) parametrically using the following framework:

$$Y_{ist} = \alpha + \beta B_{st} + k(r_{st}) + k(r_{st}) * B_{st} + \varepsilon_{ist} \tag{3}$$

where $k(r_{st})$ is a polynomial function of the relative accountability score, and B_{st} is an indicator for schools below the cutoff. Given that the source of variation in assignment is at the school level, we cluster our standard errors at the school level. We also consider clustering our standard errors at the reading accountability score level; this approach actually leads to smaller standard errors, suggesting that there is not correlation across schools within the same value of the reading accountability score (see Kolesar and Rothe, 2017). In the preferred specification, we limit the analysis to students in schools

[20] We also investigated the potential spillovers in math, either positive or negative, and found no statistically significant estimated discontinuities.

within a bandwidth of 25 accountability points, and use a linear specification. We check the robustness of this specification using different bandwidths.

6. Results

In this section, we present our results for effects on student achievement along with a series of falsification tests. The first panel of Figure 2 presents the local linear smoothing of prior year reading test scores of students on the relative reading accountability score to test baseline equivalency, calculated separately for each side of the cutoff using the triangle Kernel and a bandwidth of 25 points, with the solid circles representing the average standardized test scores for each relative accountability score. The second panel repeats the same analysis with reading test scores in reading following the introduction of the ESD program. These figures show that while students enrolled in schools immediately below the cutoff had similar prior year test scores compared to students enrolled in schools on the other side of the cutoff, they performed significantly better in reading after the introduction of extended school day. In particular, students enrolled in schools immediately below the ESD cutoff performed roughly 0.05σ better in reading.

We present our empirical estimates using equation (3) above in Table 2. The first panel presents the estimated effects of extended school day on reading scores in 2012-13. The estimates reported in column labeled as (I) use all tested students in elementary schools; column (II) restricts the analysis to students with prior year test scores; column (III) introduces student characteristics (including prior year test scores) to column (II); column (IV) restricts the analysis in column (III) to students with maternal characteristics; and column (V) introduces maternal characteristics as controls. The estimated effects remain virtually unchanged regardless of the sample or the covariates included, with students whose schools fell immediately below the ESD cutoff scoring roughly 0.05σ better in reading compared to students who schools fell on the other side of the cutoff. Figure 3 checks the robustness of these findings, and presents the estimated effects using the model in column (III) and bandwidths of 10, 15, 20, 25, and 30. In all specifications, the treatment effect sizes are almost identical, and statistically significant at the 5 percent level in all but one bandwidth.

An important concern in this context is that our estimated effects of extended school day are capturing the effects of other educational interventions in Florida targeting low performing schools. For example, Florida's school accountability system (named Florida's A+ plan) has imposed sanctions and assistance on low-performing schools that receive a school grade of D or F based on the performance of their students on FCAT since 1999. Given that almost 60 percent of all ESD schools in 2012-13 received school grades of D or F, there is the potential for observed positive effects of extended school day to be driven by the accountability program. We explore this possibility in the bottom panel of Table 2, where we present a falsification test investigating the discontinuity in reading scores at 'pseudo' cutoffs calculated using the reading accountability scores in 2011-12, the year before the policy took effect. If our estimated effects of extended school day are indeed driven by previously existing programs/policies in Florida targeting low-performing schools, we would expect to see positive "effects" of extended school day in 2011-12 as well. We find no significant discontinuities in reading scores at these pseudo cutoffs (and estimated coefficients ranging between -0.001σ to 0.012σ, much smaller than the real effects in 2012-13), providing evidence that the observed differences in 2012-13 indeed represent the causal effects of extended school day.[21]

6.1. Identification Checks

The main identification assumption in our empirical framework is that the schools whose reading accountability scores fell immediately below the cutoff are comparable to those above the cutoff. The fact that the estimated discontinuities in reading scores reported in the top panel of Table 2 remain unchanged with the inclusion of observed student and maternal characteristics provide evidence

[21] Moreover, we are not aware of any other Florida program that uses the ESD index.

supporting the validity of this assumption. Table 3 further investigates the potential for discontinuities in observable characteristics by replacing current year test scores with baseline student characteristics listed in Table 1. Each estimate represents a separate regression where the dependent variable is the indicated characteristic. In agreement with the graphical inspection in Figure 2, we find no significant discontinuity in prior year student achievement in reading, or in math. We also find that students enrolled in ESD and non-ESD schools around the cutoff are comparable along other student characteristics (e.g., FRPL eligibility, limited English proficiency, immigrant status etc.) and maternal characteristics including maternal education, marital status, mother's age at birth, and mother's nativity. Appendix Figure A1 presents a graphical depiction of these student and maternal attributes around the ESD cutoff and reach the same conclusion.

Another potential concern regarding identification in the RD design in this context, as noted in McCrary (2008), is the possibility of running variable manipulation (i.e. the reading accountability scores in this case) by teachers and/or school administrators. Under this scenario, one might expect to see an unusual discontinuity in the reading accountability score distribution around the ESD cutoff. It is important to note that this is very unlikely in our setting for two reasons. First, FCAT scores are assessed without any school involvement. Second, the ESD threshold changes every year depending on the accountability score of the marginal school that was selected for the program, making it harder for schools to manipulate their scores. Regardless, we present graphical evidence to dismiss this possibility.

Figure 4 shows the reading accountability score distributions for elementary school students (upper panel) and elementary schools (lower panel) around the ESD cutoff. The number of schools in each bin seems to be increasing as the retention cutoff falls on the left tail of the normally distributed accountability scores, and there seems to be a drop in the number of schools and the number of students immediately above the cutoff. That said, these distributions are not smooth for almost any range of accountability scores, and changes in density are roughly similar in magnitude across the entire distribution, which does not seem consistent with nonrandom selection.

A third concern in this context is differential student attrition from ESD schools after the school is selected for the program. For instance, more educationally motivated parents might leave ESD schools due to the stigma of their school being labeled as low-performing. Table 3 already presents some evidence that students who remained in ESD schools are comparable to students who remained in non-ESD schools in 2012-13 along observables around the cutoff. We also consider a more direct approach by looking for evidence of higher rates of student attrition from ESD schools that could hint at differences in unobserved student attributes and lead to biased estimates. Table 4 examines the student mobility rates at the end of 2011-12 school year around the cutoff, and shows that students who attended ESD schools in 2011-12 were no more likely to change schools or leave the Florida public school system once the school was selected for the program.

6.2. Heterogeneous Treatment Effects

In addition to the overall effects of extended school day on student achievement, another interesting question is which student subgroups benefit the most from the ESD program. Theoretically, increased in-school instruction time will lead to improved student achievement if and only if the marginal benefit of instruction time exceeds the foregone out-of-school learning opportunities. Socioeconomic status is particularly interesting because disadvantaged students tend to have worse out-of-school learning opportunities, and therefore additional in-school instruction could be more beneficial.

Table 5 breaks down the analysis in column (III) of Table 2 by student's prior year reading achievement level (1-5, with 1 being the lowest achievement level), and measures of socioeconomic status. While the estimated treatment effects are statistically indistinguishable across student groups in many cases, there are two patterns worth highlighting. First, the benefits associated with extended school day is lower for students on the left tail of the prior year reading achievement distribution. Specifically, the estimated treatment effect for student who scored in the lowest achievement level in the previous year was 0.027σ and statistically insignificant, compared to 0.077σ for students in achievement levels 2 and 3, and 0.062σ for students in achievement levels 4 and 5, both of which are statistically significant at 1 percent level.

Second, we find mixed results when comparing the effects of the program for students from different socioeconomic backgrounds. The estimated discontinuities in reading test scores are slightly larger for persistently FRPL eligible students (0.060σ) versus others (0.030σ), yet they are almost identical for students whose mothers had less than a high school diploma and students with better educated mothers (0.044σ versus 0.049σ), and higher for students whose mothers were married at the time of birth (0.066σ) versus others (0.042σ).

7. Other Years of Implementation and Long-Term Effects

As we note above, we focus on the first year of implementation in our analysis. Figure 5 presents the first year discontinuities in school-level average reading scores at the ESD cutoff for the first (Panel A), second (Panel B), and third year (Panel C) of implementation, and suggests a decline in first-year program effects over time. ESD program led to an increase of 0.20σ in current year school-level average reading test scores in 2012-13 (which correspond to roughly 0.05σ at the student-level, perfectly matching the results from our student-level analysis), which decreased to an improvement of 0.07σ in current year school-level average reading test scores in 2013-14, and further declined to -0.004σ in 2014-15.

That said, there are several reasons why this analysis is less likely to yield unbiased estimates of the causal effects of the ESD program after the first year. Most importantly, in the second year (and beyond), schools could voluntarily continue with an extended school day even though their reading accountability score fell above the cut-point (and hence they were not required to implement the policy). In fact, as reported in West and Vickers (2014), in 2013-14 school year, 30 schools (out of the 83 surveyed schools that were identified as ESD schools in 2012-13) chose to continue with an extended school day even though they were not required. This creates a fuzziness in the RD design after the first year, for which we are unable to account because FDOE only provides information about schools below the cutoff (that are required by law to extend school day), and no information, to the best of our knowledge, is available about voluntary participants in the second year and beyond. Therefore, the results presented in Panels B and C of Figure 5 will likely underestimate the benefits of the program.

Another possible source of bias after the first year of implementation is student attrition from ESD schools. Differential student attrition is less of a concern in the first year, as we show in Table 4, because the list of ESD schools was announced only two weeks before the start of school year, and therefore parents had less time to respond to the policy. After the first year, more motivated students could leave schools that were labeled as low-performing and included in the program. However, it is not feasible to check for this differential attrition using school-level data.

How about the long-term effects of the ESD program? The top panel in Figure 6 provides the estimated discontinuity in 2013-14 school-level average reading scores at the 2012-13 ESD cutoff (solid lines) compared to the discontinuity in 2012-13 scores (dashed lines), suggesting a significant decay in treatment effect in one year. In particular, schools that fell immediately below the 2012-13 reading accountability cutoff had average reading scores in 2013-14 that were 0.01σ higher than schools immediately above the cutoff, compared to the first year treatment effect of 0.20σ.

That said, it is hard to attribute this drop in estimated discontinuity at the cutoff to a decay in program effects over time because of how treatment schools are selected under the ESD program. In particular, because ESD schools are chosen based on their "relative" reading performance, a positive treatment effect of the program in the first year would help the schools immediately below the cutoff overtake schools above the cutoff in the following year, creating a discontinuity in the likelihood of being treated in the second year around the first year cutoff. If this is the case, the estimated second year effects would underestimate the true effects, because these effects would be comparing many schools that were treated in year *t* and not treated in *t+1* to schools that were not treated in year *t* and treated in year *t+1*.[22] The bottom panel in Figure 6 presents the discontinuity in whether the school fell below the ESD cutoff in 2013-14 school year around the 2012-13 cutoff, validating this concern. This graph shows that schools that fell immediately above the cutoff in 2012-13 school year were 22 percentage points more likely to fall below the cutoff (and identified as ESD schools) in 2013-14 compared to schools immediately below the cutoff in 2012-13. Overall, due to these limitations driven by the program design and data availability, we are unable to accurately estimate the long-term effects of the program or its effects in other years of implementation, and therefore focus on the first-year effects in the first year of implementation in our main analysis.

8. Conclusions

In this study, we examine the effects of extended school day on student achievement. While instruction time is a fundamental input in the education production function, we know very little about its causal effects on student outcomes due to two major empirical issues. First, observed differences in student outcomes between schools that implement extended school day policies and other schools are likely driven by differences in student characteristics and school resources between these schools, biasing the results. Second, even when the variation across schools in instruction time is driven by exogenous policy shocks, these policies typically include educational interventions other than instruction time changes, making it difficult to attribute the observed differences in student outcomes to increased

[22] Recent work by Cellini, Ferreira, and Rothstein, 2010 consider a related issue of retreatment in an RDD framework, and propose methods to estimate dynamic RD effects over many years. Given the short number of years in our setting and relatively imprecise estimates, we do not estimate formal dynamic effects.

instruction time. We address these two issues by making use of the school assignment rule used by Florida's ESD program, which requires lowest-performing elementary schools to provide an additional hour of reading instruction. Our RD estimates suggest significant positive effects of additional instruction time on student reading scores in the first year of the program.

Is the ESD program a cost-effective method to improve outcomes for students? We answer this question by following the method described by Krueger (2003) to compare the gains in future earnings to the present costs of the ESD program. Estimates from the literature suggest that a one standard deviation increase in student achievement is associated with about 8 percent higher earnings (Murnane et al, 1995; Currie and Thomas 1999; and Neal and Johnson, 1996). The annual costs of the ESD program have been estimated by district superintendents to be around \$300,000–\$400,000 per school, corresponding to \$30 million to \$40 million per year in the first two years of the program or \$800 per student annually. The benefits of the ESD program accrue over time, so we consider the present discounted value of future increases in earnings:

$$PV\ of\ Benefits = \sum_{t=18}^{80} E_t \beta \sigma_R / (1+r)^t$$

where E_t is earnings in year t, β is the relationship between reading achievement and earnings, σ_R is the effects of the program, and r is the real discount rate. We use 2016 Census data for earnings, 8 percent for β, 0.05 for σ_R, and 3 percent for r. We find a present value of benefits equal to \$3069.26 per student, which greatly outweighs the costs of \$800, suggesting that future earnings outweigh the cost of the program.[23]

This cost-benefit comparison suggests that additional instruction time for reading provides an instructional benefit per dollar spent that is in line with or superior to large-scale class size reduction,

[23] It is important to note that estimates for the ESD program are only applicable to the marginal students around the cutoff, and not the entire treatment group; further research is necessary to understand how such a program would affect higher-performing students.

depending on the estimates employed. Therefore, while increasing the length of the school day is certainly an expensive proposition, from an instructional perspective the benefit per dollar spent may be favorable to some other popular proposals to improve student outcomes.

One caveat for our findings is that it is somewhat difficult to evaluate the long-run impacts of the ESD program. Schools that are assigned to the program in 2012-13 experience increases in student achievement that make assignment less likely in the following year; a RD method applied to the 2012-13 running variable for outcomes in the following year thus compares schools treated in the first year but generally not in the second year, to schools not treated in the first year and treated in the second. The fact that subsequent RD estimates are approximately zero are consistent with three possible outcomes: (1) second year effects are positive and similar in magnitude between both groups of schools (suggesting effects persist in the second year); (2) second year effects are negative and similar in magnitude between both groups of schools (suggesting effects do not persist in the second year); or (3) effects decline to zero for schools treated in the first year and there is no effect for schools treated in the second year (suggesting effects do not persist in the second year). Further research is needed to directly study the long-run impact of the program.

References

Anderson, D.M. & Walker, M.B. (2015). Does shortening the school week impact student performance? Evidence from the four-day school week. *Education Finance and Policy*, 10(3), 314-349.

Aucejo, E.M. & Romano, T.F., (2016). Assessing the effect of school days and absences on test score performance. *Economics of Education Review*, *55*, pp.70-87.

Battistin, E. & Meroni, E.C. (2016). Should we increase instruction time in low achieving schools? Evidence from Southern Italy. *Economics of Education Review*, *55*, pp.39-56.

Bellei, C. (2009). Does lengthening the school day increase students' academic achievement? Results from a natural experiment in Chile. *Economics of Education Review*, 28(5), 629-640.

Bitler, M., J., Gelbach, J., & Hoynes, H. What mean impacts miss: Distributional impacts of welfare reform experiments. *American Economic Review*, 96(4), 988-1012.

Block, J. H., & Burns, R. B. (1976). Mastery learning. III L. S. Shulman (Ed.), Review of research in education (vol. 4) (pp. 3-49). Itaska, IL: F. E. Peacock

Bloom, B. S. (1968). Learning for mastery. Instruction and Curriculum. Regional Education Laboratory for the Carolinas and Virginia, Topical Papers and Reprints, Number 1. Evaluation comment, 1(2), n2.

Card, D. (1999). The causal effect of education on earnings. *Handbook of labor economics*, 3, 1801-1863.

Carroll, J. (1963). A model of school learning. *The Teachers College Record*, 64(8), 723-723.

Chingos, M.M. (2013). Class size and student outcomes: Research and policy implications. *Journal of Policy Analysis and Management*, *32*(2), pp.411-438.

Chingos, M.M. & Whitehurst, G.J. (2011). Class size: what research says and what it means for state policy. *Brookings Institute. May*, *11*.

Cortes, K. E., Goodman, J. S., & Nomi, T. (2015). Intensive Math Instruction and Educational Attainment Long-Run Impacts of Double-Dose Algebra. Journal of Human Resources, 50(1), 108-158.

Currie, J. (2001). Early childhood education programs. *The Journal of Economic Perspectives*, *15*(2), pp.213-238.

DeCicca, P. & Smith, J. (2013). The long-run impacts of early childhood education: Evidence from a failed policy experiment. *Economics of Education Review*, *36*, pp.41-59.

DeCicca, P. (2007). Does full-day kindergarten matter? Evidence from the first two years of schooling. *Economics of Education Review*, *26*(1), pp.67-82.

Dobbie, W., & Fryer, Jr., R. G. (2011). Getting beneath the veil of effective schools: Evidence from New York City (No. w17632). National Bureau of Economic Research.

Farbman, D., & Kaplan, C. (2005). Time for a change: The promise of extended-time schools for promoting student achievement. Boston, MA: Massachusetts 2020.

Figlio, D., Guryan, J., Karbownik, K. & Roth, J. (2014). The Effects of Poor Neonatal Health on Children's Cognitive Development. *American Economic Review*, 104(12): 3921-3955.

Folsom, J. S., Osborne-Lampkin, L., Cooley, S., & Smith, K. (2017). Implementing the extended school day policy in Florida's 300 lowest performing elementary schools (REL 2017–253). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southeast. Retrieved from http://ies.ed.gov/ncee/edlabs.

Folsom, J. S., Petscher, Y., Osborne-Lampkin, L., Cooley, S., Herrera, S., Partridge, M., &

Smith, K. (2016). School reading performance and the extended school day policy in Florida (REL 2016–141). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southeast. Retrieved from http://ies.ed.gov/ncee/edlabs.

Goodman, J. (2014). Flaking out: Student absences and snow days as disruptions of instructional time. National Bureau of Economic Research WP No: 20221.

Hahn, J., Todd, P., & Van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, 69(1), 201-209.

Harris, D.N. (2009). Toward policy-relevant benchmarks for interpreting effect sizes: Combining effects with costs. *Educational Evaluation and Policy Analysis*, *31*(1), pp.3-29.

Jensen, V.M. (2013). Working longer makes students stronger? The effects of ninth grade classroom hours on ninth grade student performance. *Educational Research*, *55*(2), pp.180-194.

Krueger, A.B. (1999). Experimental estimates of education production functions. *The Quarterly Journal of Economics*, *114*(2), pp.497-532.

Lavy, V. (2015). Do differences in schools' instruction time explain international achievement gaps? Evidence from developed and developing countries. *The Economic Journal*, *125*(588), pp.F397-F424.

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

Levin, H. M., & Tsang, M. C. (1987). The economics of student time. Economics of education review, 6(4), 357-364.

Machin, S. & McNally, S. (2008). The literacy hour. *Journal of Public Economics*, 92, 1441-1462.

Marcotte, D. E. (2007). Schooling and test scores: A mother-natural experiment. *Economics of*

Education Review, 26(5), 629-640.

Marcotte, D. E., & Hemelt, S. W. (2008). Unscheduled school closings and student performance. *Education Finance and Policy*, 3(3), 316-338.

McCrary, J. (2008). Testing for manipulation of the running variable in the regression discontinuity design. *Journal of Econometrics*, 142(2), 698-714.

National Center for Education Statistics, Common Core of Data (CDD), 2013. National Public Education Financial Survey. *U.S. Department of Education.*

National Center for Education Statistics, School and Staffing Survey (SASS), 2012. Public Teacher Data File. *U.S. Department of Education.*

Parinduri, R.A. (2014). Do children spend too much time in schools? Evidence from a longer school year in Indonesia. *Economics of Education Review*, *41*, pp.89-104.

Patall, E. A., Cooper, H., & Allen, A. B. (2010). Extending the school day or school year: A systematic review of research (1985–2009). *Review of educational research*, 80(3), 401-436.

Pischke, J.S. (2007). The impact of length of the school year on student performance and earnings: Evidence from the German short school years. *The Economic Journal*, *117* (523), pp.1216-1242.

Porter, J. (2003). Estimation in the regression discontinuity model. Unpublished Manuscript, Department of Economics, University of Wisconsin at Madison, 5-19.

Rivkin, S.G. and Schiman, J.C. (2015). Instruction time, classroom quality, and academic achievement. *The Economic Journal*, *125*(588), pp.F425-F448.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. Econometrica, 73(2), 417-458.

Roza, M. and Ouijdani, M. (2012). The Opportunity Cost of Smaller Classes: A State-By-State Spending Analysis. Schools in Crisis: Making Ends Meet. *Center on Reinventing Public Education*.

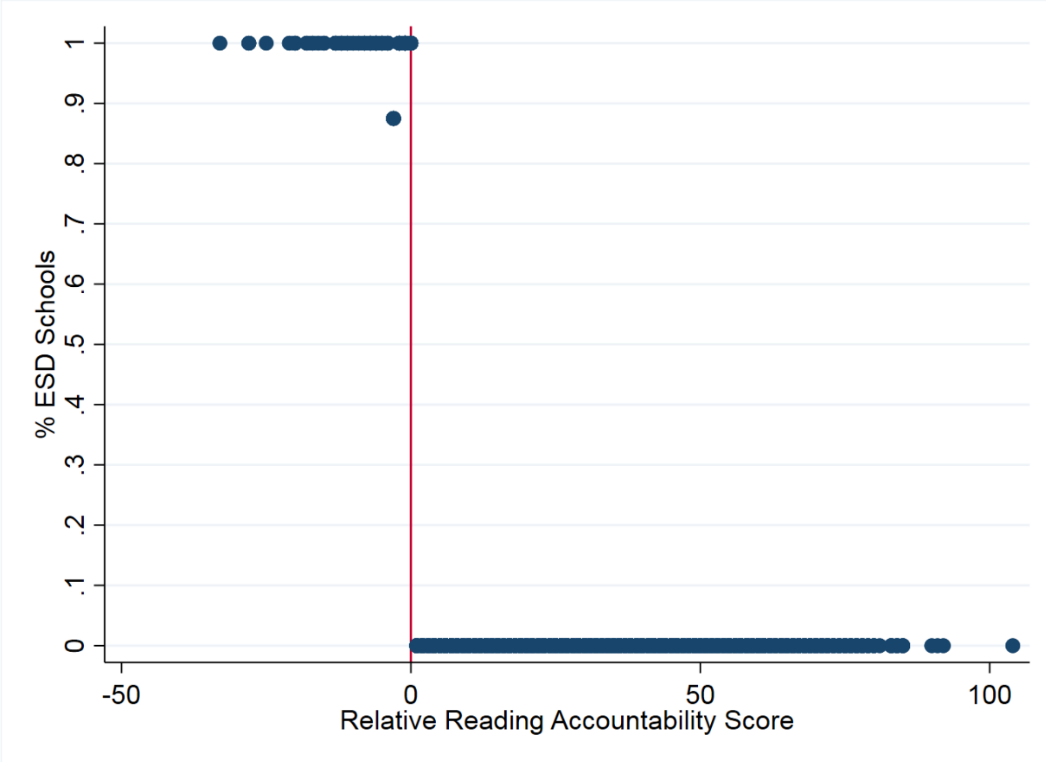
Sims, D. P. (2008). Strategic responses to school accountability measures: It's all in the timing. Economics of Education Review, 27, 58–68.

Taylor, E. (2014). Spending more of the school day in math class: Evidence from a regression discontinuity in middle school. *Journal of Public Economics*, 117, 162-181.

West, M., & Vickers, B. (2014). OPPAGA review of the Extra Hour Initiative. Tallahassee, FL: The Florida Legislature's Office of Program Policy Analysis & Government Accountability. Retrieved from http://www.edweek.org/media/18florida-extra-hour-presentation.pdf

Tables and Figures

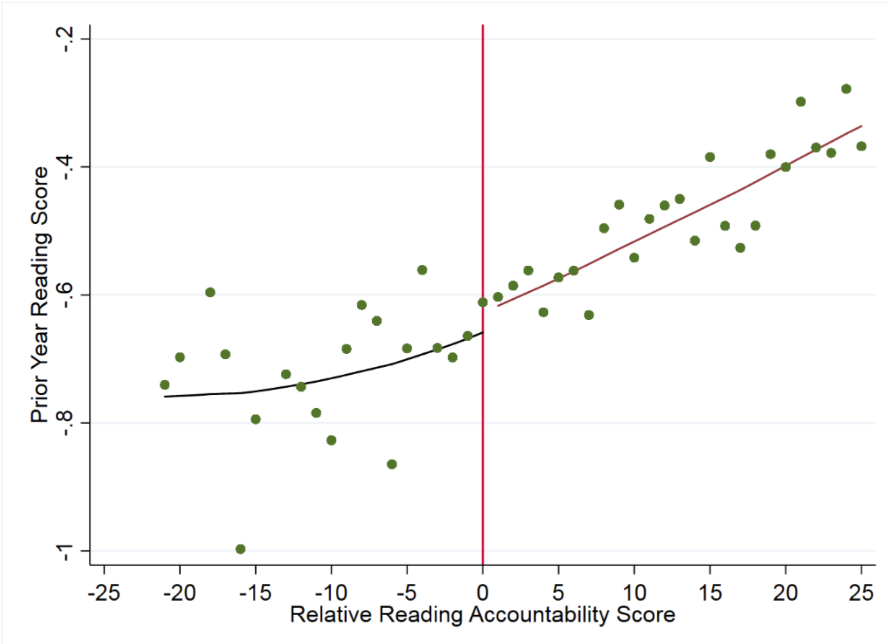
Figure 1
ESD Selection and Reading Accountability Score



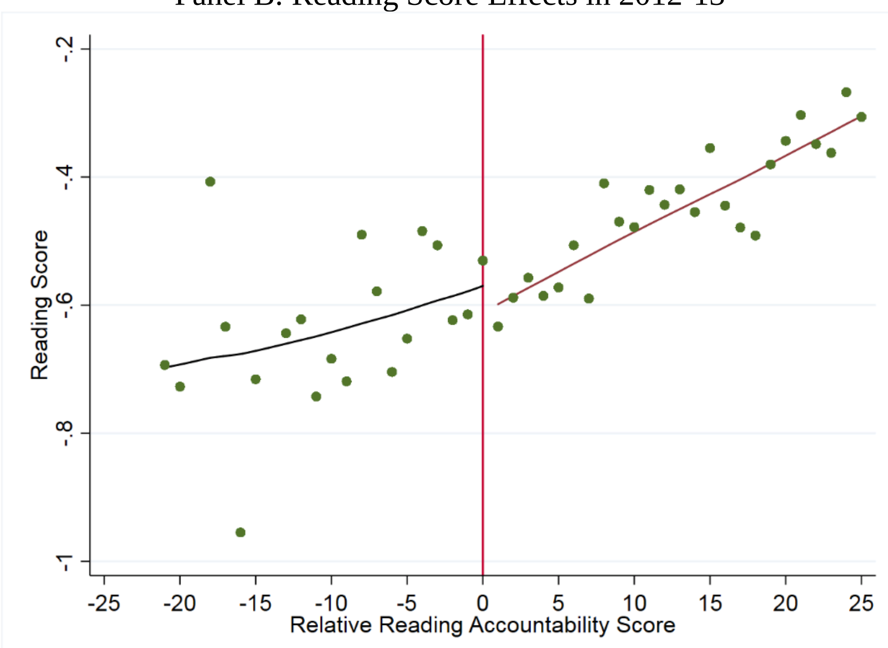
Notes: The figure plots the percentage of schools selected by the ESD program by relative reading accountability score. The solid circles represent raw cell means.

Figure 2
Extended School Day and Student Achievement

Panel A: Prior Year Reading Score in 2011-12

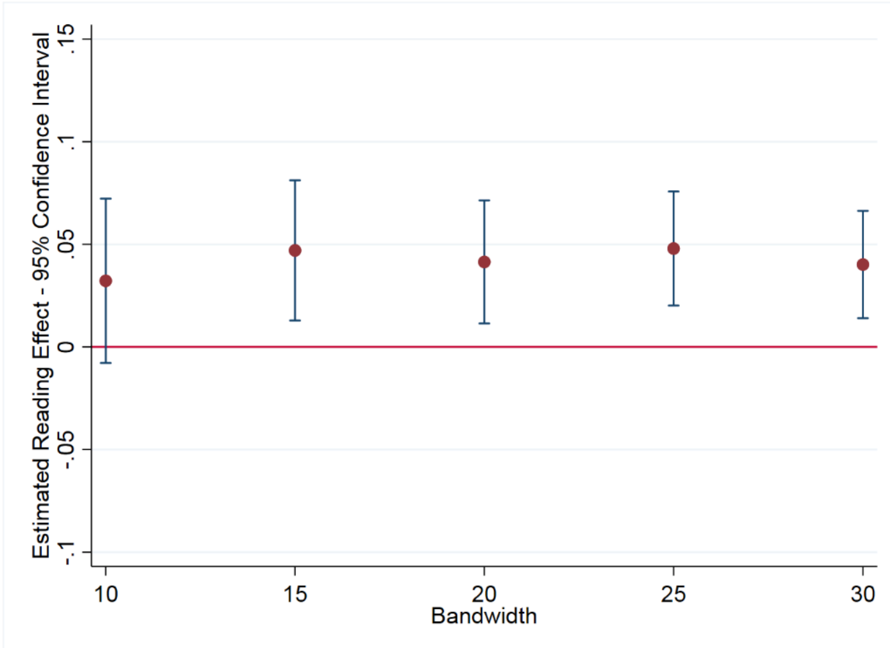


Panel B: Reading Score Effects in 2012-13



Notes: Each graph present the local linear smoothing of current reading scores (panel A), prior year reading scores (panel B), on the relative reading accountability score of the school separately for the left of the cutoff and the right. The triangle kernel and a bandwidth of 25 points are used in the estimation. The solid circles represent raw cell means.

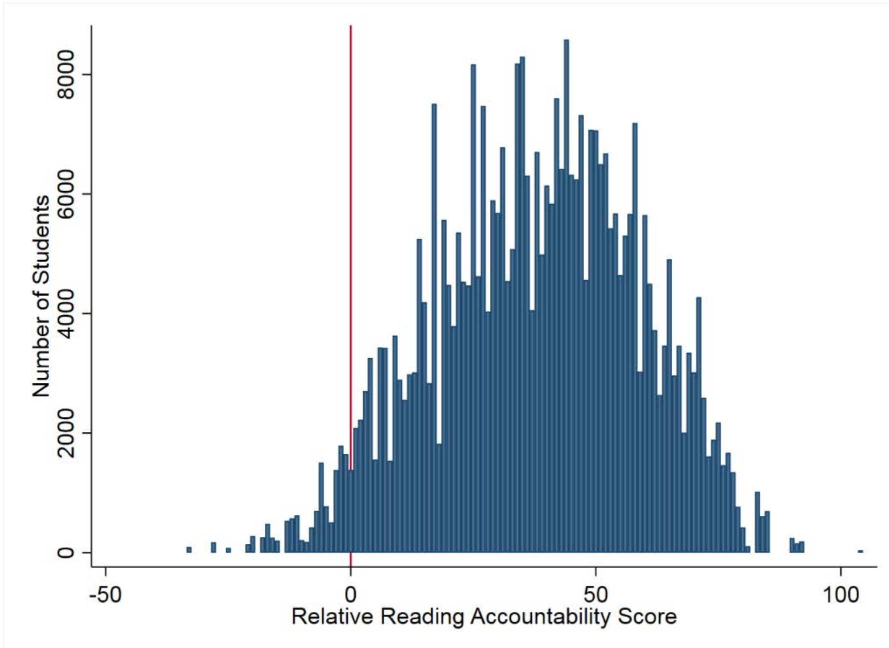
Figure 3
Robustness to Bandwidth Selection



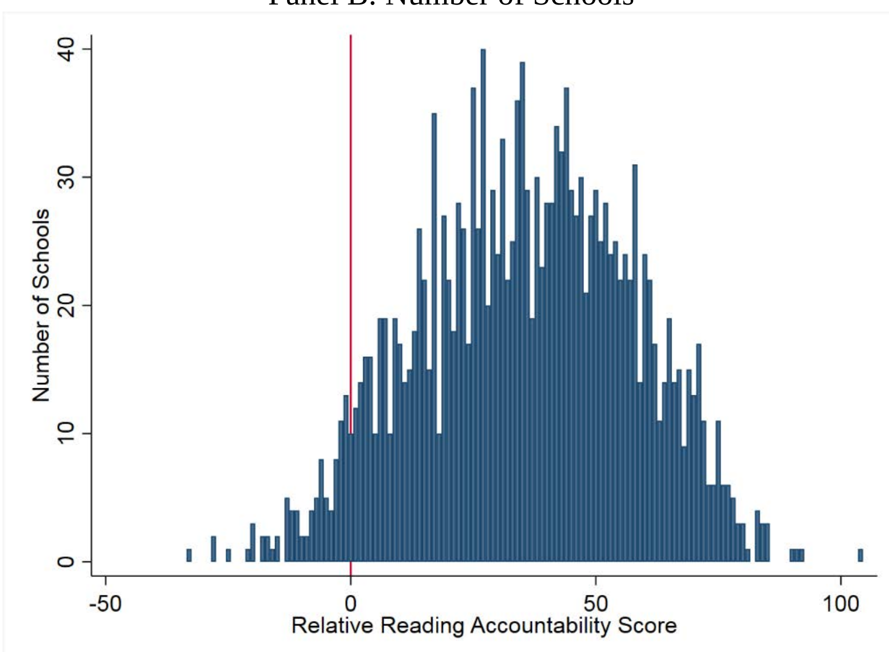
Notes: The figure plots the estimates effects of extended school day on reading scores for the five predicted groups based on the regression of prior year reading scores on FRPL eligibility history, maternal education, maternal marital status, maternal immigrant origin, and maternal age. The vertical lines represent the 95 percent confidence interval for the estimates.

Figure 4
Reading Accountability Score Distribution

Panel A: Number of Students



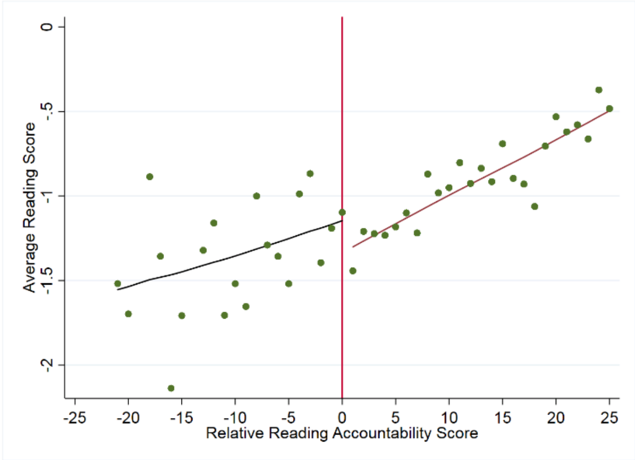
Panel B: Number of Schools



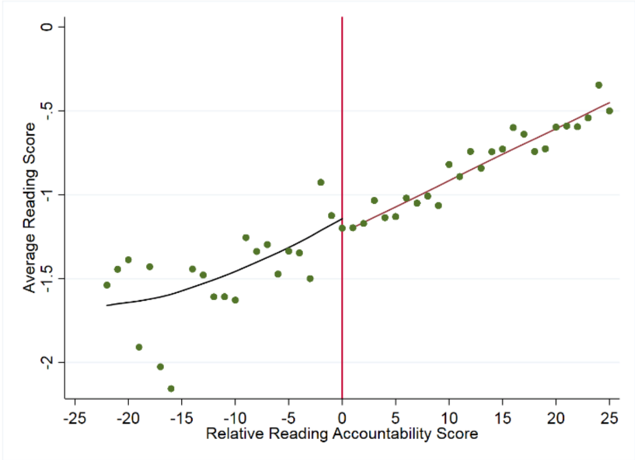
Notes: The figures present the number of schools (panel A) and the number of students (panel B) in each reading accountability score bin. The ESD cutoff is shown by the vertical line.

Figure 5
First Year Effects of ESD Program on School-Level Reading Test Scores
Year of Implementation

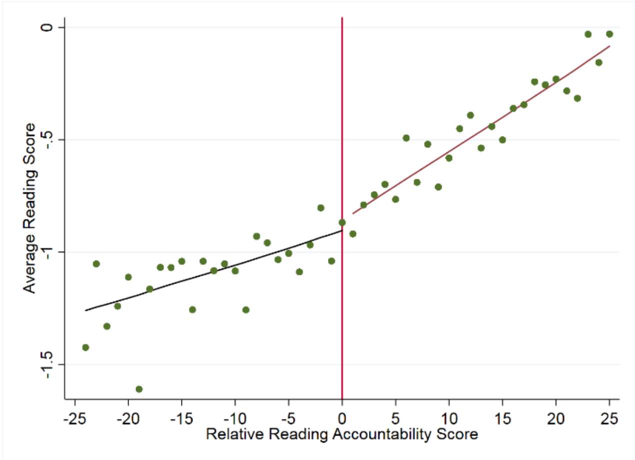
Panel A: 2012-13 School Year



Panel B: 2013-14 School Year



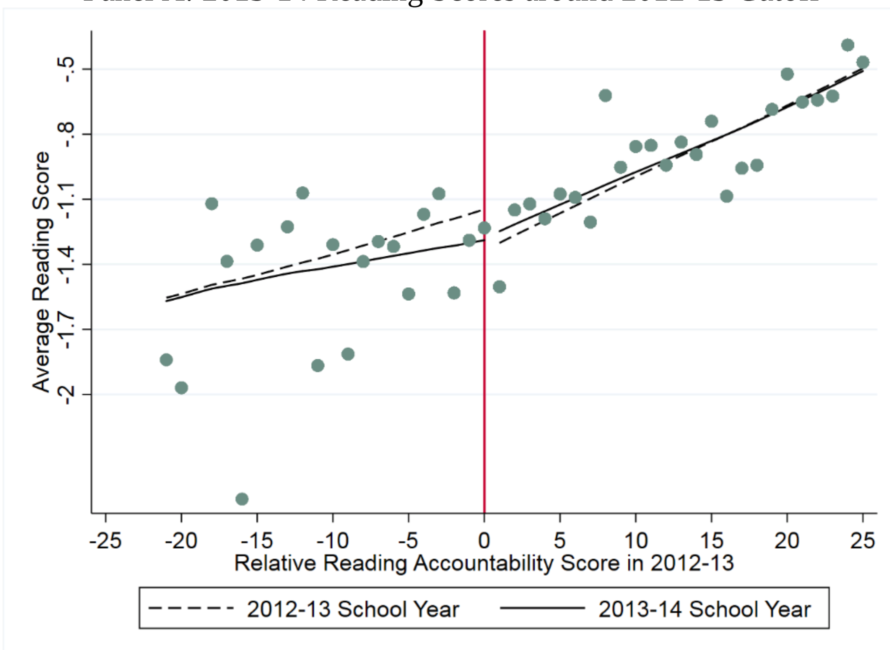
Panel C: 2014-15 School Year



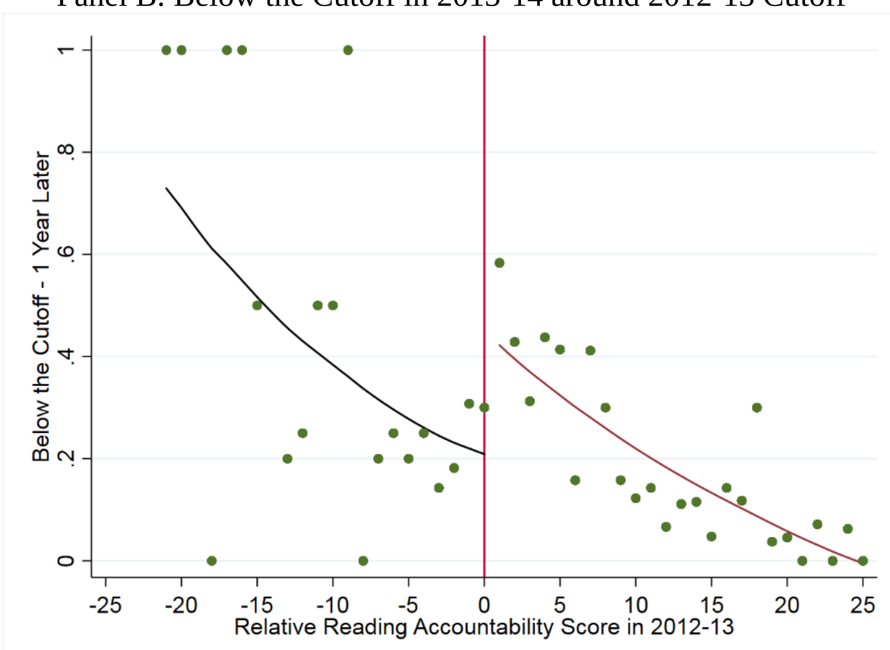
Notes: Each graph present the local linear smoothing of current year average school reading scores on the relative reading accountability score of the school separately for the left of the cutoff and the right. The triangle kernel and a bandwidth of 25 points are used in the estimation. The solid circles represent raw cell means.

Figure 6
Second Year Effects of ESD Program on School-Level Reading Test Scores

Panel A: 2013-14 Reading Scores around 2012-13 Cutoff



Panel B: Below the Cutoff in 2013-14 around 2012-13 Cutoff



Notes: The top panel present the local linear smoothing of current year and 1-year later average school reading scores on the relative reading accountability score of the school separately for the left of the cutoff in 2012-13 and the right. The bottom panel replicates the same analysis using an indicator for schools below the cutoff in 2013-14 as the outcome. The triangle kernel and a bandwidth of 25 points are used in the estimation. The solid circles in the top panel represent raw cell means of 2013-14 reading scores, and the bottom panel represent the raw cell means of the below the cutoff in 2013-14 indicator.

Table 1
Descriptive Statistics

	Student Characteristics	
	ESD Schools	Non-ESD Schools
Prior year reading score	-0.706	-0.0396
	(0.863)	(1.013)
Prior year math score	-0.591	-0.0279
	(0.890)	(1.007)
FRPL eligible	0.951	0.651
	(0.215)	(0.477)
Persistently FRPL eligible	0.828	0.477
	(0.378)	(0.499)
Limited English proficient	0.126	0.107
	(0.331)	(0.309)
Ever limited English proficient	0.201	0.209
	(0.401)	(0.407)
English non-native	0.282	0.335
	(0.450)	(0.472)
Foreign born	0.0660	0.0918
	(0.248)	(0.289)
Special education student	0.159	0.146
	(0.365)	(0.354)
Gifted	0.0216	0.0809
	(0.145)	(0.273)
White	0.0895	0.411
	(0.285)	(0.492)
Hispanic	0.185	0.300
	(0.388)	(0.458)
African American	0.691	0.222
	(0.462)	(0.416)
Male	0.515	0.515
	(0.500)	(0.500)
Old for grade	0.252	0.135
	(0.434)	(0.342)
Number of schools	99	1,686
Number of students	21,392	558,254
Number of students with test scores	14,034	362,479
	Maternal Characteristics	
Maternal years of education	11.03	12.41
	(2.382)	(2.641)
Maternal education-less than HS	0.464	0.248
	(0.499)	(0.432)
Maternal education-HS degree	0.404	0.371
	(0.491)	(0.483)
Maternal education-some college	0.110	0.213

	(0.313)	(0.409)
Maternal education-Bachelor's degree	0.0227	0.168
	(0.149)	(0.374)
Mother married	0.236	0.554
	(0.425)	(0.497)
Mother's age at pregnancy	24.46	27.10
	(6.044)	(6.329)
Mother had a teenage pregnancy	0.211	0.116
	(0.408)	(0.320)
Mother foreign born	0.0798	0.0948
	(0.271)	(0.293)
Number of schools	99	1,686
Number of students with maternal char.	8,487	184,807

Notes: Standard deviations are given in parentheses.

Table 2
Extended School Day and Reading Achievement

	(I)	(II)	(III)	(IV)	(V)
Effect in 2012-13	0.049*	0.051*	0.056***	0.047***	0.045**
	(0.029)	(0.028)	(0.017)	(0.018)	(0.018)
Number of students	162,887	106,609	106,609	58,894	58,894
Number of schools	578	578	578	578	578
Pseudo effect in 2011-12	0.012	0.013	0.003	-0.001	0.001
	(0.026)	(0.025)	(0.016)	(0.015)	(0.015)
Number of students	144,898	93,418	93,418	68,489	68,489
Number of schools	522	522	522	522	522
Sample restrictions:					
Students with prior test scores	No	Yes	Yes	Yes	Yes
Students with maternal attributes	No	No	No	Yes	Yes
Covariates:					
Student characteristics	No	No	Yes	Yes	Yes
Maternal characteristics	No	No	No	Yes	Yes

Notes: Robust standard errors, clustered at the school level, are given in parentheses. Discontinuity estimates are obtained parametrically using a bandwidth of 25 points and linear functional form. Column labeled as (I) presents estimates from the base specification using all students; column (II) presents estimates from the base specification using students with prior year test; column (III) presents estimates with student characteristics reported in Table 1 using students with prior year test scores; column (IV) presents estimates with student characteristics reported in Table 1 using students with prior year test scores and maternal characteristics; and column (V) presents estimates with student and maternal characteristics reported in Table 1 using students with prior year test scores and maternal characteristics. The top panel presents the estimates for 2012-13, and the bottom panel presents the pseudo effects using student test scores in reading and the school reading accountability scores in 2011-12. *, ** and *** represent statistical significance at 10, 5 and 1 percent respectively.

Table 3
Extended School Day and Baseline Student and Maternal Characteristics

Student Characteristics	
Prior year reading score	-0.015
	(0.028)
Prior year math score	0.018
	(0.041)
FRPL eligible	-0.004
	(0.012)
Persistently FRPL eligible	-0.014
	(0.019)
Limited English proficient	0.014
	(0.031)
Ever limited English proficient	-0.016
	(0.048)
English non-native	-0.069
	(0.054)
Foreign born	0.005
	(0.012)
Special education student	-0.001
	(0.010)
Gifted	0.001
	(0.007)
White	-0.005
	(0.026)
Hispanic	-0.014
	(0.051)
African American	0.016
	(0.056)
Male	-0.008
	(0.006)
Old for grade	-0.018
	(0.011)
Number of students	162,887
Number of students with test scores	106,609
Number of schools	578
Maternal Characteristics	
Maternal years of education	0.133
	(0.208)
Maternal education-less than HS	-0.023
	(0.026)
Maternal education-HS degree	-0.002
	(0.020)
Maternal education-some college	0.021*
	(0.011)

Maternal education-Bachelor's degree	0.005
	(0.005)
Mother married	0.001
	(0.024)
Mother's age at pregnancy	0.225
	(0.226)
Mother had a teenage pregnancy	-0.001
	(0.009)
Mother foreign born	0.009
	(0.026)
Number of students with maternal char.	58,894
Number of schools	578

Notes: Robust standard errors, clustered at the school level, are given in parentheses. The estimates represent the discontinuities in student and maternal characteristics at the ESD cutoff in 2012-13, obtained parametrically using a bandwidth of 25 points and linear specification. *, ** and *** represent statistical significance at 10, 5 and 1 percent respectively.

Table 4
Extended School Day and Student Mobility

	(I)	(II)	(III)	(IV)	(V)
At the end of 2011-12…					
Changed schools	0.003	0.002	0.002	0.001	0.002
	(0.014)	(0.014)	(0.013)	(0.012)	(0.012)
Left FL public schools	-0.012	-0.016*	-0.016*	-0.012	-0.012
	(0.007)	(0.009)	(0.009)	(0.010)	(0.010)
Number of students	111,812	55,404	55,404	41,755	41,755
Number of schools	578	578	578	578	578
Sample restrictions:					
Students with prior test scores	No	Yes	Yes	Yes	Yes
Students with maternal attributes	No	No	No	Yes	Yes
Covariates:					
Student characteristics	No	No	Yes	Yes	Yes
Maternal characteristics	No	No	No	Yes	Yes

Notes: Robust standard errors, clustered at the school level, are given in parentheses. Discontinuity estimates are obtained parametrically using a bandwidth of 25 points and linear functional form. Column labeled as (I) presents estimates from the base specification using all students; column (II) presents estimates from the base specification using students with prior year test; column (III) presents estimates with student characteristics reported in Table 1 using students with prior year test scores; column (IV) presents estimates with student characteristics reported in Table 1 using students with prior year test scores and maternal characteristics; and column (V) presents estimates with student and maternal characteristics reported in Table 1 using students with prior year test scores and maternal characteristics. The top panel presents the estimates for 2012-13, and the bottom panel presents the pseudo effects using student test scores in reading and the school reading accountability scores in 2011-12. *, ** and *** represent statistical significance at 10, 5 and 1 percent respectively.

Table 5
Extended School Day and Reading Achievement
Heterogeneous Treatment Effects

Prior year reading achievement level	
Lowest [N= 32,182]	0.026
	(0.021)
Level 2 [N= 34,602]	0.077***
	(0.022)
Level 3 and higher [N= 40,019]	0.062***
	(0.020)
Socioeconomic status	
Persistently FRPL eligible [N= 79,610]	0.060***
	(0.018)
Not persistently FRPL eligible [N= 27,193]	0.039
	(0.026)
Mother less than HS diploma [N= 24,409]	0.044**
	(0.021)
Mother HS diploma or higher [N= 34,598]	0.049**
	(0.021)
Mother married [N= 21,159]	0.066***
	(0.023)
Mother not married [N= 38,381]	0.042**
	(0.021)
Student characteristics	Yes

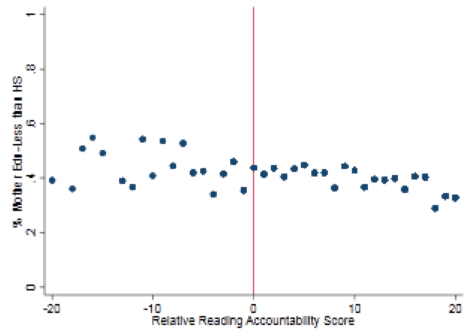
Notes: Robust standard errors, clustered at the school level, are given in parentheses. Discontinuity estimates are obtained parametrically using a bandwidth of 25 points and linear functional form separately for each subgroup using students with prior year test scores. Each regression includes 578 schools. All regressions control for student characteristics reported in Table 1. *, ** and *** represent statistical significance at 10, 5 and 1 percent respectively.

Appendix A

Figure A.1

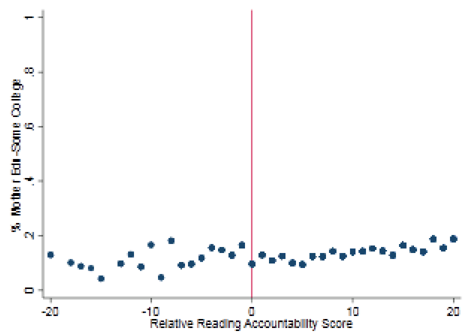
Additional Instruction Time and Student Achievement

Fraction Mother < High School

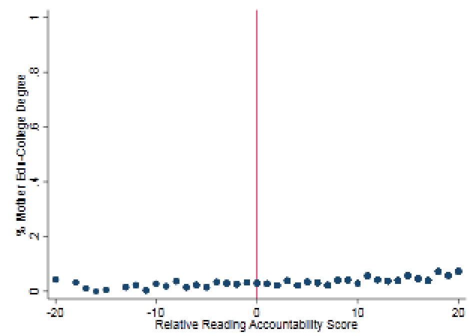


Fraction Mother High School

Fraction Mother Some College

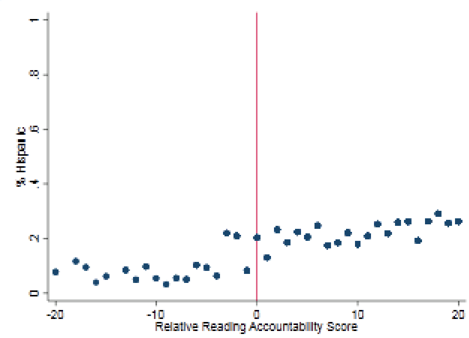


Fraction Mother Bachelor's Degree



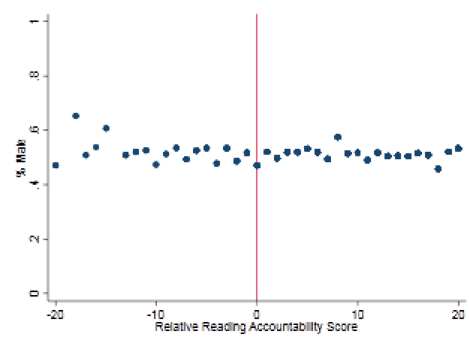
Fraction Black

Fraction Hispanic

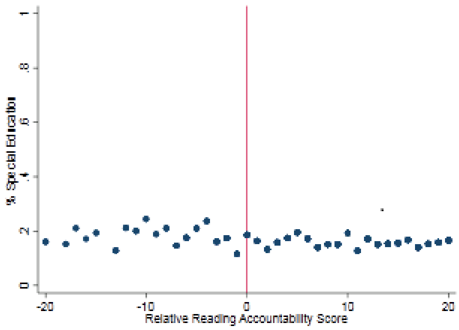


Fraction White

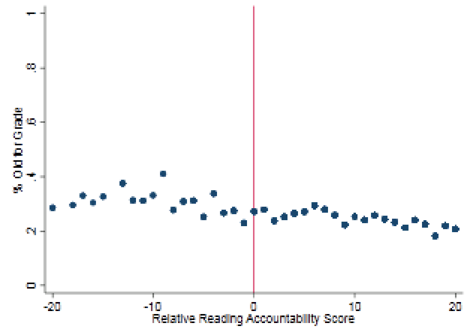
Fraction Male



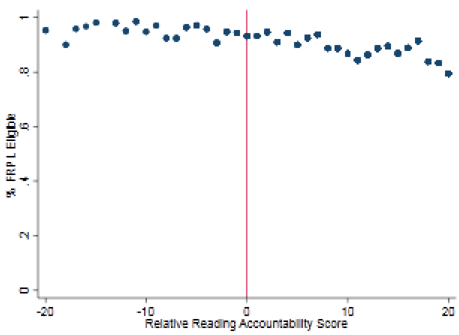
Fraction Special Education



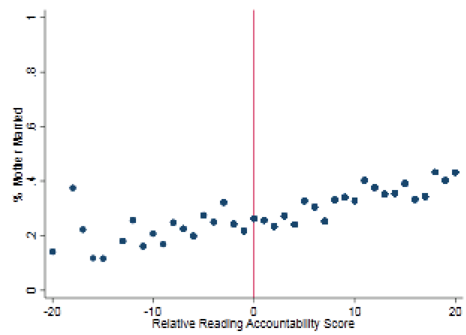
Fraction Old for Grade



FRPL Eligible

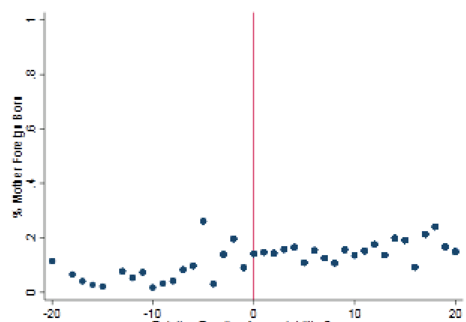


Fraction Married Mother

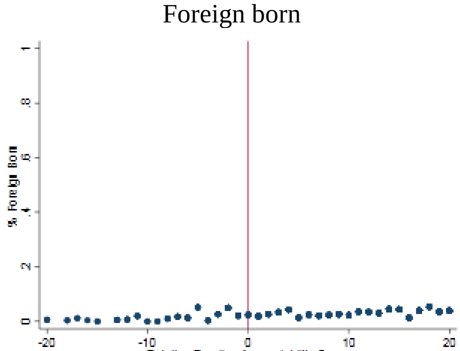


Persistently FRPL Eligible

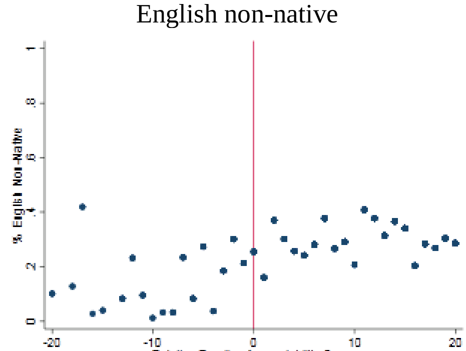
Mother foreign born



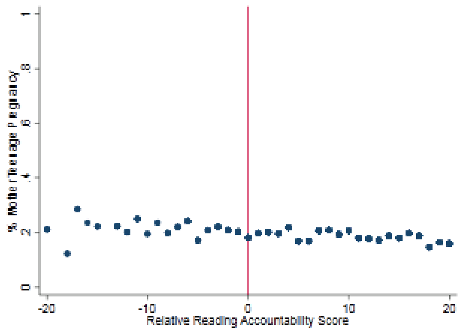
Foreign born



English non-native



Mother teenage pregnancy



Notes: Each figure plots the average student or maternal characteristic by relative reading accountability score, with the vertical line indicating the ESD cutoff. The solid circles represent raw cell means.