CESifo WORKING PAPERS

10526
2023

June 2023

Sibling Spillovers May Enhance the Efficacy of Targeted School Policies

David Figlio, Krzysztof Karbownik, Umut Özek

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 10526*

Sibling Spillovers May Enhance the Efficacy of Targeted School Policies

Abstract

Public policies often target individuals but within-family externalities of such interventions are understudied. Using a regression discontinuity design, we document how a third grade retention policy affects both the target children and their younger siblings. The policy improves test scores of both children while the spillover is up to 30% of the target child effect size. The effects are particularly pronounced in families where one of the children is disabled, for boys, and in immigrant families. Candidate mechanisms include improved classroom inputs and parental school choice.

JEL-Codes: D130, I200, J130.

Keywords: grade retention, sibling spillovers, policy externalities, test scores.

David Figlio
University of Rochester
Rochester / NY / USA
david.figlio@rochester.edu

Krzysztof Karbownik
Emory University
Atlanta / GA / USA
krzysztof.karbownik@emory.edu

Umut Özek
RAND Corporation
Arlington / VA / USA
uozek@rand.org

June 21, 2023
We are grateful to twelve anonymous Florida school districts for providing the confidential administrative data for this project. We thank seminar participants at Emory University and Columbia University Teachers College for helpful comments and feedback. All errors are our own.

Many public policies target specific individuals within a family and in particular select children.[1] Cost-benefit analyses of these policies focus on their effects on the children in question. Despite their prevalence, little is known about how these child-specific policies propagate within families, potentially affecting many more individuals than those considered in the extant literature. If policies that benefit specific children also benefit others in the family, then evaluations systematically understate the benefit-cost ratio.

We study mandatory retention policy (coupled with additional investment in students) that affected about 10% of Florida families whose children were enrolled in third grade during the time frame we examine. Similar policies are common across states and countries, and have been recently studied, with varying results (see Allen et al. (2009) and Goos et al. (2021) for reviews).[2] We investigate whether these policies affect others in the family: Does retention of an older sibling (coupled with enhanced resources) affect outcomes of their younger siblings? How large are these externalities compared to the effects on focal children? Do they differ across groups? And what are some plausible mechanisms behind the observed effects?

Florida's retention policy requiring third grade students to meet a reading test threshold to be promoted to fourth grade provides quasi-random variation for a regression discontinuity design. (We interpret our findings as intent-to-treat effects since (a) not all families comply with the assignment, (b) retention triggers additional instructional support for students, and (c) we show that retention affects an array of outcomes of the focal child.) Twelve anonymous county-level school districts provided us with rich administrative data permitting linkage of children within families.

We find that Florida's policy of grade retention coupled with additional investments increases test scores of the focal children in grades 4 to 5 by about 0.20-0.21σ. This effect spills over to their younger siblings who score 0.05-0.06σ higher than the younger siblings of children not flagged for retention, implying a positive externality of up to 30%. The effects are similar for mathematics and reading, while we do not find any spillover effects on the likelihood of being placed in special education. These spillovers are concentrated in families where either a focal child or a sibling is in special education, in immigrant families, and among boys. We uncover several potential mechanisms that can explain over 40% of the observed spillover effect: In the affected families, parents are more likely to switch away from low-performing schools for a younger sibling and into higher "value-added" school environments, and irrespective of school changes, younger siblings

[1] Education examples include teacher aides (Andersen et al. 2020), tracking (Betts 2011), tutoring (Nickow et al. 2020), special education (Ballis and Heath 2021), and grade retention (Goos et al. 2021), but numerous examples exist in health (e.g., Golberstein et al. (2022)) and welfare (e.g., Deshpande and Mueller-Smith (2022)) as well.

[2] Examples from the extant literature include studies from Florida (Greene and Winters 2007; Winters and Greene 2012; Schwerdt et al. 2017; Figlio and Özek 2020), Chicago (Jacob and Lefgren 2009), New York City (Mariano et al. 2018), Louisiana (Eren et al. 2017; Larsen and Valant 2018; Eren et al. 2022), Indiana (Hwang and Koedel 2022), and Mississippi (Mumma and Winters 2023). We know of no extent work examining potential externalities of these retention policies.

are educated in classrooms with higher-performing peers and teachers. The results are robust to choice of bandwidth, sample, estimator, inclusion of controls, and standard errors methodology.

Our findings contribute to a recent literature on sibling spillovers.[3] Few of these papers study externalities from targeted policies that are ubiquitous in real life. We conjecture that in educational settings such policies could generate very different effects from naturally occurring variation (e.g., peer attributes) or universal and often pre-determined policies (e.g., school entry age) for at least three reasons: (a) targeted policies carry with them a clear and discretely changing information about student ability which is easily observable and salient for parents; (b) at the margin the "shock" is unanticipated by either students or parents which could alter their behavior differently from a predictable event (e.g., in our specific application due to stigma related to grade retention which could be motivational (positive) or disengaging (negative)); and (c) targeted policies are often aimed at specific subsets of the population (e.g., students struggling academically) which can be very different from a "representative student" (e.g., Karbownik and Özek (forthcoming) show that the same universal policy - school entry age - generates spillovers that differ across the SES gradient). Finally, from a policy evaluation perspective, it is imperative to estimate total effects of educational interventions including any potential (positive or negative) externalities they may generate.

The closest extent literature to this paper studies spillovers generated by specific targeted campaigns, such as changes in school curriculum (Joensen and Skyt Nielsen 2018), vaccination campaigns (Alsan 2017), and medical interventions at birth (Daysal et al. 2022). In the development context, researchers have examined gender segregation of schools in Pakistan (Qureshi 2018a), deworming in Kenya (Ozier 2018), and iodine supplementation in Tanzania (Adhvaryu and Nyshadham 2016). Our application offers the opportunity to study a relatively ubiquitous (but unpredictable at the margin for focal families) targeted policy in a developed setting.

Our findings regarding parental responses also shed some light on within-household inequality, social interactions, and differential parental investments in children. Our spillovers can arise either (a) due to direct interactions between siblings or (b) due to parental responses to a high-stakes signal about the academic performance of the older child. The former channel is akin to within-family peer effects (Manski 2000; Sacerdote 2014). The latter channel could imply directing resources towards the struggling child in a compensatory fashion (Pitt et al. 1990; Conley 2008) or away from the said child through reinforcement (Becker and Tomes 1976; Grätz and Torche 2016). Yi

[3] Prior sibling spillovers studies used variation in teacher quality (Qureshi 2018b), peer quality (Nicoletti and Rabe 2019), high school course taking (Gurantz et al. 2020), college admissions (Altmejd et al. 2021), disability (Black et al. 2021), teen pregnancy (Heissel 2021), ADHD diagnoses (Breining 2014; Persson et al. 2021), the 1918 influenza pandemic (Parman 2015), and military service (Bingley et al. 2021). Landersø et al. (2020), Karbownik and Özek (forthcoming), and Zang et al. (forthcoming) study school starting age rules which, although not targeted at specific groups, induce variation in eligibility of a specific child to start schooling.

et al. (2015), Leight and Liu (2020), Berry et al. (2020) all document that parents could engage in even more complex reallocation mechanisms depending on the nature of the shock. While we are unable to test for or quantify the role of the direct effects, we examine if and how parents respond to the high-stakes signal about the academic performance of their older child. Indeed, one plausible mechanism we find support for in our data is differential parental school choice.

Our results have three main policy implications. First, we provide some of the first evidence that targeted educational policies generate within-family externalities. The non-trivial effect size of the spillover (of up to 30% of the main effect) suggests that much of the policy-evaluation literature might be underestimating the benefits of productive interventions or the losses from adverse shocks. On the margin, this could matter for adoption, expansion, or rejection of said policies. Second, we specifically focus on grade retention policies which have been a controversial educational tool, and yet we know nothing about the potential externalities they may generate. Third, our heterogeneity analysis reveals the need for particular attention towards families with special needs children where the spillover effects are particularly pronounced. This is a sizeable population of children accounting for approximately 10% of Florida families served under IDEA part B (U.S. Department of Education 2020) and 14% when we consider state individual education plans (Florida Department of Education 2019).

1 Florida's Third Grade Retention Policy

Enacted in 2002 as part of the broader "Just Read, Florida!" initiative, Florida's third-grade retention policy requires students to meet the Level 2 benchmark (the second lowest of five achievement levels) on the statewide reading test in order to be promoted to fourth grade (Figlio and Özek 2020; Florida Statues 2022). Since there are several "good cause exemptions" that allow students to advance to the next grade despite scoring below the Level 2 cutoff, the policy generates a fuzzy regression discontinuity.[4] A key feature of these exemptions, however, is that they require significant parental and teacher involvement.[5]

Once the child is retained, the legislation requires that schools provide substantial additional instructional support for such students - a plausible reason behind the success of Florida's policy compared with similar policies in other states. In particular, the law requires schools to (1) develop academic improvement plans for students that specifically address their learning needs; (2)

[4]Students are eligible for an exemption (1) if they have certain disabilities and have been already retained once; (2) if they have received intensive reading remediation for two years and have already been retained twice; (3) if they have been in the English learner program for less than two years; (4) by demonstrating that they are reading at a level equal to or above a Level 2 on the statewide reading test by performing at an acceptable level on an alternative standardized reading assessment approved by the State Board of Education; or (5) by demonstrating proficiency through a teacher-developed portfolio.

[5]As such, there is evidence suggesting that students from disadvantaged backgrounds are less likely to receive exemptions than academically-comparable students from more affluent families (Licalsi et al. 2019).

assign these students to high-performing teachers; (3) provide at least 90 minutes of daily reading instruction; and (4) offer summer reading camp at the end of the year that facilitates intensive reading intervention lasting between six and eight weeks for all students who scored below the retention cutoff. As such, in our analysis we estimate the combined effects of grade retention and these instructional supports on targeted students and how these spill over to their siblings.

Approximately 10% of students were retained on annual basis at the end of third grade between 2002 and 2010 with the fraction declining from 15% in 2002 to 6% in 2010. This change is driven by both decreasing share of students scoring below the unchanged retention cutoff as well as increasing share of low-performers receiving the exemptions (Licalsi et al. 2019).

2 Data

We use student-level administrative data from 12 anonymous school districts in Florida covering school years 2002-03 to 2011-12. The information includes Florida Comprehensive Assessment Test (FCAT) scores, student demographics, special education status, and characteristics of child's teachers. We link these data with birth records for all children born in Florida between 1994 and 2002 who subsequently attended a public school in the state which allows us to verify that children are indeed siblings and accurately assign birth order (Figlio et al. 2014). Appendix Table A.1 presents the descriptive statistics. We focus on 7 cohorts of focal students who entered third grade for the first time between 2002-03 and 2008-09 as well as their closest younger sibling.[6] We limit the sample to sibling pairs where the younger sibling is already enrolled in a public school (kindergarten through 2nd grade) at the time when the older child enters third grade for the first time.[7] Compared to all third graders from the select cohorts (Column 1), our preferred estimation sample (Column 5) is negatively selected with students having lower test scores, higher likelihood of subsidized meal eligibility, and higher likelihood of having high school dropout and unmarried (at the time of birth) mother. This makes sense given that the policy specifically targets lowest-performing children who, in Florida, are more likely to live in families with such characteristics.

3 Empirical Strategy

We use third-grade reading scores, a regression discontinuity (RD) design, and student-level data on both individuals directly affected by the policy (focal students i) as well as their younger

[6] Roughly 85 percent of the focal students in our sample have only one younger sibling observed in our matched school records while this percentage increases to 95 when we consider third graders with a younger sibling with a third-grade score. Our results are substantively unchanged when we include all younger siblings (rather than just the closest one) in the analysis.

[7] This additional restriction allows us to observe characteristics of the younger child which we use to document covariates' balance. Furthermore, since test scores in Florida are observed in third grade at the earliest we have fewer observations for families with siblings spaced far apart. Our results are similar when we use this larger, but potentially confounded by endogenous school entry, sample.

siblings (j). We estimate the following equation using ordinary least squares:

$$Y_j = \gamma + \tau B_i + \kappa(S_i) + \kappa(S_i) \times B_i + \varepsilon_j \tag{1}$$

where Y_j are outcomes of younger sibling j, $\kappa(S_i)$ is linear function of focal student's i reading test score in third-grade relative to the Level 2 cutoff (which remained unchanged in Florida during the time frame we examine), B_i is an indicator variable taking value of 1 for focal students who score below the Level 2 cutoff and thus are eligible for retention, and ε_j is Eicker-Huber-White heteroskedasticity-robust standard error. Our primary outcomes of interest are reading and math scores on standardized tests: in grades 4 and 5 for the focal students and in grades 3 through 5 for their younger siblings.[8] In the main results we use bandwidth of 25 points based on selection procedure of Calonico et al. (2017).[9] Our parameter of interest, τ, is an unbiased estimate of an older child scoring below the retention cutoff on their younger siblings' outcomes under two assumptions: (1) all other focal student and their younger sibling attributes are smooth through the cutoff and (2) there is no manipulation of the running variable around the cutoff (McCrary 2008; Lee and Lemieux 2010). In select specifications we replace Y_j with Y_i to document baseline equivalency and to estimate effects on outcomes of the focal child.

Table 1 presents estimates based on Equation 1 where we include pre-determined characteristics as outcomes. Out of 15 estimates, none are statistically significant at conventional levels. Furthermore, we reject the joint statistical significance of these coefficients: p-value of 0.11 for the baseline equivalency of focal students and p-value of 0.15 for younger siblings. Appendix Table A.2 repeats this exercise using non-parametric estimation proposed by Calonico et al. (2017). We have also verified that there are no discontinuities in either the likelihood of observing a younger sibling as a function of focal child's treatment (coefficient of -0.002 with p-value of 0.575) or the likelihood of being observed in specific grades or for specific number of years (Appendix Table A.3). Finally, Appendix Figure A.1 presents distribution of focal students with younger siblings and we reject the hypothesis that it is discontinuous at the cutoff (p-value 0.662). Since we are interested in estimating the spillover effects of an older child's retention, we further require that the policy indeed forces at least some students below the cutoff to repeat third grade (first-stage). Panel A of Appendix Figure A.2 and Column (1) of Panel A of Table 2 shows this exact discon-

[8]The average age difference between sibling pairs in our sample is 2.2 years. As such, we observe outcomes for both the focal student and the younger sibling during similar time frames after the focal student is flagged for retention.

[9]In the Appendix we document that our results are robust to (a) bandwidth choice (Lee and Lemieux 2010), (b) clustering of standard errors (Lee and Card 2008), (c) local polynomial non-parametric estimation methods adjusting for mass points in the running variable (Calonico et al. 2017), (d) honest estimators (Armstrong and Kolesar 2020; Kolesar and Rothe 2018), (e) removing additional covariates, and (f) different sample choices. We prefer our parametric local linear approach since our running variable is continuous, due to its transparency, and since inclusion of control variables improves precision of the estimates.

tinuity for the focal (older) child in the family, implying a change in the likelihood of retention of over 35 percentage points (pp) at the cutoff. Based on these checks, we conclude that τ represents the reduced-form causal spillover of interest.

4 Results

Panel A of Table 2 presents effects of the policy on retention, test scores, and special education designation of the focal (older) child in the family. Column (1) documents a 36 percentage points increase in a probability of being retained in third grade among students who scored below the Level 2 cutoff compared with students who scored above. In subsequent elementary school grades, Columns 2 to 6, students flagged for retention have improved test scores and lower likelihood of grade repetition. The test score gains at about 0.2σ are nontrivial economically given that average test scores in grades 4 and 5 of students who score just above Level 2 cutoff in grade 3 are at -0.8 to -0.7σ.[10] We present scatterplots of the select focal child outcomes in Appendix Figure A.2. Second, Panel B of this table presents outcomes in grades 3 to 5 for the younger siblings of the children examined in Panel A. We present scatterplots of the select younger sibling outcomes in Appendix Figure A.3. The results imply positive, statistically significant, and sizeable sibling spillovers in test scores. Point estimates of 0.05 to 0.06σ correspond to between 22 and 31 of the respective main effects on the focal children. We also investigated if the estimates are similar across grades. We found that, for combined math and reading test scores, they grow from 0.04σ in grade 3 to 0.06σ in grade 6. Although statistically indistinguishable from each other, the increase over grades is consistent with permanent changes in inputs for the younger siblings which we explore in Section 5. We did not find any statistically significant spillover effects on the likelihood that the younger sibling was retained (Column 2) or identified as special education (Column 6) in grades 3 through 5.[11]

Considering the effect sizes of the estimated test score spillovers, at 30% of the focal child effect, they are about three times larger than spillovers from exposure to higher ability peers (Nicoletti and Rabe 2019) but are similar in magnitude to spillovers from being assigned to a more experienced teachers (Qureshi 2018b). In the context of school entry policies, Karbownik and Özek (forthcoming) estimate positive siblings spillovers in the pooled sample of about 40% of the

[10]There are two common ways retention effects are examined in the literature:(1) holding grade constant (i.e., compare retained and promoted students when they reach the same grade) and (2) holding age constant (i.e., compare outcomes in the years following the retention). We follow the former approach in our main analysis as we view the additional year of instruction and support as part of the reduced-form treatment. That said, we present the results using the latter approach in Appendix Table A.4 and our conclusions regarding the effects of retention on the academic achievement of the focal child remain unchanged. The test scores used in Appendix Table A.4 are developmental FCAT scale scores (rather than grade-by-year standardized scores) that are vertically-aligned, which allows us to compare student achievement across grades. They imply positive effect sizes of 0.15-0.22 σ.

[11]We also examined absence rates. We found improvements for both focal children as well as their younger siblings but the latter coefficients are imprecisely estimated.

direct effect while effect sizes in Zang et al. (forthcoming) are slightly smaller at 20% of the direct effect. Our estimates are also very similar in absolute value to the spillovers from exposure to a disabled younger sibling (Black et al. 2021). In contrast to longer-standing gaps in achievement, the externalities estimated here are on the smaller side of about 10 percent of the Black-White test score gap and 6 percent of the gap in test scores between children of mothers who are high school dropouts vs. had at least some college education observed in our data.

4.1 Robustness

This section presents a series of analyses to gauge the robustness of our main results. For brevity, we focus on the third grade retention of the focal child (Column 1 of Table 2) and averaged mathematics and reading test scores for both the focal child as well as their younger sibling (Column 3 of Table 2). We first test the robustness of our main results to different bandwidths and standard errors assumptions. Appendix Figure A.4 presents these results. Irrespective of the outcome, our point estimates are very stable across bandwidths from 15 to 45 points around the cutoff, and, if anything, the spillovers (Panel C) are slightly larger when we focus on siblings of focal students very close to the cutoff. Furthermore, all coefficients are statistically significant at conventional levels whether we use standard errors that are heteroskedasticity-robust or clustered at the running variable. This is expected given that our running variable is continuous.

We provide additional robustness checks in Appendix Table A.5 where Panel A presents estimates for the focal child and Panel B presents the spillover effects on their younger sibling. In particular, we estimate the regressions without additional control variables (Column 2), using local polynomial estimation (Columns 3 and 4), using honest estimation (Column 5), focusing on first- and second-born children only to minimize the quantity-quality trade-offs (Column 6), and including families where the younger child was not yet in schooling at the time when their older sibling entered the third grade for the first time (Column 7). In all those analyses we get very similar point estimates as in our baseline specification (Column 1) with all coefficients being statistically significant at conventional levels.

Another concern with using the regression discontinuity design based on meeting a certain required level of proficiency is that being at or around the cutoff could in itself have an information effect that is independent of any policy consequences that we are attempting to estimate. To rule this out, we replicate our analysis using Level 3 rather than Level 2 reading cutoff. This is a placebo exercise since Level 3 cutoff is not tied to any requirements regarding retention or additional resources dedicated for students. Column 8 of Appendix Table A.5 presents these results and we don't find any sizeable or statistically significant effects on either the focal child or their younger sibling. Overall, we conclude that our results are robust to reasonable permutations of the preferred empirical specification.

4.2 Heterogeneity

In this section, we focus on heterogeneity along three dimensions: presence of children with special education needs in the household, gender of the sibling, and immigration status of the family.[12] The results are presented in Table 3. First, in Panel A, we see limited heterogeneity in the effects on focal child's third grade retention or their subsequent test scores. Even the largest relative difference, for retention among recent vs. not recent immigrant families, is less than 30 percent. Furthermore, all point estimates are statistically significant at conventional levels. On the other hand, when investigating the spillovers in Panel B, we find striking heterogeneity. The externalities are concentrated in families where one of the children is in special education, when the sibling is male, and in recent immigrant families.[13]

Why could we see particularly pronounced spillovers in these three groups? Parents of special education children are part of the student's IEP team and are required to be actively involved in the development of the plan along with the teachers and the school administrators. As such, these parents could respond more strongly to the high-stakes signal about the low performance of the focal child and have the means (e.g., established relationships with school administrators) to alter the educational resources for their younger sibling (e.g., classroom assignments). Gender-wise, our results are consistent with differential sensitivity of boys relative to girls to family and school inputs (Autor et al. 2016, 2019, 2023). The larger spillovers among recent immigrants could be due to the fact that these families are more likely to be involved in school activities that are directly related to their children's achievement and hence are likewise more responsive to signals about the low academic performance (Kao and Tienda 1995). We come back to the heterogeneity analysis below when investigating differential parental school choice decisions and educational resources available to siblings as potential mechanisms for the documented externalities.[14]

[12] Children with special education needs are defined as those with Individualized Education Plans (IEP). Recent immigrant families are defined as those where the mother is foreign born or English is not the main language spoken at home. We also investigated heterogeneity along other dimensions such as race/ethnicity, parental marital status at birth, free or reduced-price eligibility, and maternal education. We did not detect meaningful differences in estimated spillover effects across those groups. One reason for this could be that children in our "cutoff sample" are all very low performers who mostly grow up in disadvantaged families. For example, 80 percent of students are eligible for subsidized meals while only about 20 percent have mothers with some college experience.

[13] The effects are very similar whether we consider special education designation of the focal child or their younger sibling, which provides suggestive evidence that it is not the differential effect of the policy for focal students with disabilities that is driving the larger spillover effects, but rather something about families with special education students. For this reason we classify families as having either vs. neither children in special education.

[14] We have also investigated spillovers by spacing between siblings. Among the siblings already in school the spillover effect is smaller at 0.03σ (SE of 0.02σ) for pairs where the age difference is $\leqslant 2$ years compared with 0.09σ (SE of 0.04σ) for pairs spaced farther apart. This is despite very comparable retention and test score effects for the focal child in these two groups.

5 Mechanisms

There are several ways parents could respond to their older child being flagged for retention that would also affect younger children in the family. While we do not have direct measures of parental involvement or investments, we consider two types of mechanisms that could plausibly explain the observed spillover effects: educational resources available to younger siblings in their classrooms and parental school choice decisions which could be one of the reasons for plausibly observing differential educational resources.

Table 4 documents results pertaining to the first channel where we consider two positive inputs into human capital production function that have been extensively studied in prior literature: teacher effectiveness proxied by their contribution to student test scores (Rivkin et al. 2005; Sass et al. 2012; Chetty et al. 2014) and classroom peers (Burke and Sass 2013; Sacerdote 2014; de Gendre and Salamanca 2020).[15] Our measure of teacher effectiveness is based on leave-out-year value added similar to Chetty et al. (2014), calculated using averaged test scores in reading and math. The peer effects measure is based on classroom-peer prior year math and reading scores that are averaged using the time that a student spends in the classroom (per week) as weights. In calculating the averages we only use classrooms that are designated as self-contained, English language arts (ELA), or math in the course enrollment data.

Column 1 presents pooled estimates for all younger sibling of focal children who score just below vs. just above the Level 2 cutoff while Columns 2 to 7 divide the sample along the dimensions highlighted in Section 4.2. Considering all students, we find statistically significant effects of focal child being flagged for retention on the likelihood that the younger sibling is assigned to classrooms with higher-achieving peers.[16] The evidence on peer effects in compulsory schooling is mixed and average effects vary from positive (de Gendre and Salamanca (2020) for Taiwan at 0.052 SD, Burke and Sass (2013) for Florida at 0.029 SD, and Imberman et al. (2012) for Louisiana at 0.33 SD) to null (Lavy et al. (2012) for England and Imberman et al. (2012) for Houston). Assuming that peer effects exist, our point estimate of 0.032 SD would imply expected test score gains in the rage of 0.1 to 1 percent of a SD, or 2 to 20 percent of the spillover estimated in Column (3) of Table 2. On the other hand, there is no effect on leave-out-year teacher value added.

Both estimates, however, mask considerable heterogeneity with the effects amplified in families with special education needs child, for males, and for recent immigrants (only for peer achievement); exactly the groups for which we documented the largest spillovers in Section 4.2. Taking a closer look at, for example, families with a disabled child reveals 6.7 and 0.9 percent of a SD in-

[15] We have also explored another input commonly considered in education policy - class size - but in this case we did not find any sizeable or statistically significant effects.

[16] We did not find any sizeable or statistically significant changes in other peer characteristics such as race or FRPL eligibility.

creases in peer quality and teacher value added, respectively. The former estimate implies expected test score gains of up to 0.022 SD while the latter (see Hanushek and Rivkin (2010) for a review of value added estimates) of up to 0.003 SD.[17] Our results thus suggest that one mechanism through which the positive spillovers are generated could be due to additional educational resources that younger siblings receive via peers and teachers, with the former appearing quantitatively more important than the latter.

It is challenging to assess importance of the documented changes in educational inputs in the context of potential sibling gains. Nonetheless, to better understand the potential contribution of these changes for test score outcomes, in Table A.6 we explore this question descriptively though mediation analysis which we execute for both pooled sample as well as for the six sub-groups outlined in Table 3. Panel A copies the main findings from Tables 2 and 3 while Panel B replicates them on a sample where we also have complete data for mechanisms considered in Table 4. The results are substantively unchanged with an exception that we now find statistically significant spillovers for native children (albeit still much smaller than for the immigrant families). In Panel C, when we add controls for teacher value added and classroom peers prior achievement, the coefficients meaningfully decline. In the pooled sample (Column 1) the estimate is reduced by 44% which suggests that mechanisms documented in Table 4 plausibly contribute to the observed sibling spillovers although we caution the readers that this analysis cannot be viewed as causal without strong and perhaps untenable assumptions.

There are two ways through which parents could alter educational resources faced by the younger sibling: (1) they might influence classroom assignments in the current school or (2) they might move their younger child to a different school. In Table 5 we specifically focus on the latter channel. Columns 1 to 3 present estimates for the likelihood that the younger child left their current school within two years after their older sibling was flagged for retention. Considering all children in the sample, we do not detect any sizable or statistically significant effects on school switching. We then stratify focal child's school by its quality in prior two years into the highest rated (A-schools) for at least one of the two years and all other schools combined together (B, C, D, and F-schools). We find no effects for children attending higher-performing schools; however, for lower-performing schools, we observe that parents are more likely to move their younger sibling to a different school.[18] That said, we do not find any significant effect on the likelihood that the focal student leaves their low-performing school in the two years after being flagged for retention.

[17] Source numbers for these calculations are the following: 0.022 = 0.067 (Column 2 of Table 4) $\times$ 0.330 (Column 4 of Table 6 in Imberman et al. (2012)) and 0.003 = 0.009 (column 2 of Table 4) $\times$ 0.360 (Table 1 in Hanushek and Rivkin (2010) based on Nye et al. (2004)).

[18] We observe 90% of younger children being in the same school as their older sibling at the time when the latter enters third grade for the first time. Therefore, average school quality for both children at the baseline is almost identical.

This provides evidence that mobility results for the younger sibling are driven by parents exercising school choice for their younger child rather than Tiebout choice, which would likely lead to a school change for both siblings.[19] It suggests that (1) the high-stakes signal about the academic performance of the older sibling is particularly salient in low-performing schools (perhaps because parents believe that overall school quality could be the reason behind their older child's academic struggles) and (2) the threat of retention is an important motivator for parents.

Do these families move their younger child to higher-performing schools with more resources? We explore this question in Columns 4 to 8 of Table 5 which limit the sample to families where focal child attended relatively lower performing school. For these families, we find that the focal child being flagged for retention increases the likelihood that the younger sibling moves to schools with improved reading gains, increased share of highly effective teachers, and particularly those specializing in reading.[20] We likewise find positive coefficients for mathematics but these are not statistically significant at conventional levels. Overall, it appears that parents indeed try to divert their younger children to somewhat higher performing and better resourced schools; particularly when it comes to reading resources which is the domain specifically targeted by the policy.

6 Conclusions

Many public policies are targeted at specific individuals within a family and yet we know relatively little about how such interventions can propagate across household members. Although policy evaluations typically focus on targeted individuals when assessing cost effectiveness, our findings suggest that failing to account for within-family externalities in these calculations could yield misleading conclusions. In our context, considering the effects of a beneficial policy on just the focal student would understate the total benefits to the family.

Importantly, these findings could likewise indicate that a targeted policy with negative effects on focal children could have larger overall negative effects. For instance, some studies have found negative effects of grade retention policies when not accompanied by increased investments and there could be a (negative) multiplier effect within the family in these circumstances for the same reasons that we find a positive multiplier effect in our application. The results of our paper suggest that targeted policies are likely to have more diffuse impacts within a family, which should be kept in mind when determining the benefits and costs of any given policy or intervention.

[19] Related to the heterogeneity by spacing discussed in Section 4.2, we further find that these school changes are much more likely to occur in families where age difference between siblings is more than 2 years. This makes sense if changing school is easier and more productive (due to longer exposure to a higher quality school) in KG compared with grade 2.

[20] We define an effective teacher as one whose leave-out-year value-added score is in the top quartile of the value-added distribution in a given year. We also studied the likelihood of moving to an A-rated school but we do not find any statistically significant effects for this outcome.

References

Adhvaryu, Achyuta and Anant Nyshadham (2016) "Endowments at birth and parents' investments in children," *Economic Journal*, Vol. 126, No. 593, pp. 781–820.

Allen, Chiharu, Qi Chen, Victor Willson, and Jan Hughes (2009) "Quality of research design moderates effects of grade retention on achievement: A meta-analytic, multilevel analysis," *Educational Evaluation and Policy Analysis*, Vol. 31, No. 4, pp. 480–499.

Alsan, Marcella (2017) "The gendered spillover effects of young children's health on human capital: Evidence from Turkey."

Altmejd, Adam, Andres Barrios-Fernandez, Marin Drlje, Joshua Goodman, Michael Hurwitz, Dejan Kovac, Christine Mulhern, Christopher Neilson, and Jonathan Smith (2021) "O brother, where start thou? Sibling spillovers on college and major choice in four countries," *Quarterly Journal of Economics*, Vol. 136, No. 3, pp. 1831–1886.

Andersen, Simon, Louise Beuchert, Helena Skyt Nielsen, and Mette Thomsen (2020) "The effect of teacher's aides in the classroom: Evidence from a randomized trial," *Journal of the European Economic Association*, Vol. 18, No. 1, pp. 469–505.

Armstrong, Timothy and Michal Kolesar (2020) "Simple and honest confidence intervals in nonparametric regression," *Quantitative Economics*, Vol. 11, No. 1, pp. 1–39.

Autor, David, David Figlio, Krzysztof Karbownik, Jeffrey Roth, and Melanie Wasserman (2016) "School quality and the gender gap in educational achievement," *American Economic Review*, Vol. 106, No. 5, pp. 289–295.

——— (2019) "Family disadvantage and the gender gap in behavioral and educational outcomes," *American Economic Journal: Applied Economics*, Vol. 11, No. 3, pp. 338–381.

——— (2023) "Males at the tails: How socioeconomic status shapes the gender gap," *Economic Journal*, Vol. forthcoming.

Ballis, Briana and Katelyn Heath (2021) "The long-run impacts of special education," *American Economic Journal: Economic Policy*, Vol. 13, No. 4, pp. 72–111.

Becker, Gary and Nigel Tomes (1976) "Child endowments and the quantity and quality of children," *Journal of Political Economy*, Vol. 84, No. 4, pp. S143–S162.

Berry, James, Rebecca Dizon-Ross, and Maulik Jagnani (2020) "Not playing favorites: An experiment on parental fairness preferences."

Betts, Julian (2011) "The economics of tracking in education," in Eric Hanushek, Stephen Machin, and Ludger Woessmann eds. *Handbook of the Economics of Education*, Vol. 3: Elsevier, pp. 341–381.

Bingley, Paul, Petter Lundborg, and Stephanie Lyk-Jensen (2021) "Brothers in arms: Spillovers from a draft lottery," *Journal of Human Resources*, Vol. 56, No. 1, pp. 225–268.

Black, Sandra, Sanni Breining, David Figlio, Jonathan Guryan, Krzysztof Karbownik, Helena Skyt Nielsen, Jeffrey Roth, and Marianne Simonsen (2021) "Sibling spillovers," *Economic Journal*, Vol. 131, No. 633, pp. 101–128.

Breining, Sanni (2014) "The presence of ADHD: Spillovers between siblings," *Economics Letters*, Vol. 124, No. 3, pp. 469–473.

Burke, Mary and Tim Sass (2013) "Classroom peer effects and student achievement," *Journal of Labor Economics*, Vol. 31, No. 1, pp. 51–82.

Calonico, Sebastian, Matias Cattaneo, Max Farrell, and Rocio Titiunik (2017) "rdrobust: Software for regression-discontinuity designs," *Stata Journal*, Vol. 17, No. 2, pp. 372–404.

Chetty, Raj, John Friedman, and Jonah Rockoff (2014) "Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood," *American Economic Review*, Vol. 104, No. 9, pp. 2633–2679.

Conley, Dalton (2008) "Bringing sibling differences in: Enlarging our understanding of the transmission of advantage in families," in Annette Lareau and Dalton Conley eds. *Social class: How does it work?*: Ruseel Sage Foundation, pp. 179–200.

Daysal, Meltem, Marianne Simonsen, Mircea Trandafir, and Sanni Breining (2022) "Spillover effects of early-life medical interventions," *Review of Economic and Statistics*, Vol. 104, No. 1, pp. 1–16.

Deshpande, Manasi and Michael Mueller-Smith (2022) "Does welfare prevent crime? The criminal justice outcomes of youth removed from SSI," *Quarterly Journal of Economics*, Vol. 137, No. 4, pp. 2263–2307.

Eren, Ozkan, Briggs Depew, and Stephen Barnes (2017) "Test-based promotion policies, dropping out, and juvenile crime," *Journal of Public Economics*, Vol. 152, pp. 9–31.

Eren, Ozkan, Michael Lovenheim, and Naci Mocan (2022) "The effect of grade retention on adult crime: Evidence from a test-based promotion policy," *Journal of Labor Economics*, Vol. 40, No. 2, pp. 361–395.

Figlio, David, Jonathan Guryan, Krzysztof Karbownik, and Jeffrey Roth (2014) "The effects of poor neonatal health on childnren's cognitive development," *American Economic Review*, Vol. 104, No. 12, pp. 3921–3955.

Figlio, David and Umut Özek (2020) "An extra year to learn English? Early grade retention and the human capital development of English learners," *Journal of Public Economics*, Vol. 186, p. 104184.

Florida Department of Education (2019) "2019 SEA Profile."

Florida Statues (2022) "Florida Statue 1001.215."

de Gendre, Alexandra and Nicolas Salamanca (2020) "On the mechanisms of ability peer effects."

Golberstein, Ezra, Irina Zainullina, Aaron Sojourner, and Mark Sander (2022) "Effects of school-based mental health services on youth outcomes in Hennepin County."

Goos, Mieke, Joana Pipa, and Francisco Peixoto (2021) "Effectiveness of grade retention: A systematic review and meta-analysis," *Educational Research Review*, Vol. 34, p. 100401.

Grätz, Michael and Florencia Torche (2016) "Compensation or reinforcement? The stratification of parental responses to children's early ability," *Demography*, Vol. 53, No. 6, pp. 1883–1904.

Greene, Jay and Marcus Winters (2007) "Revisiting grade retention: An evaluation of Florida's test-based promotion policy," *Education Finance and Policy*, Vol. 2, No. 4, pp. 319–340.

Gurantz, Oded, Michael Hurwitz, and Jonathan Smith (2020) "Sibling effects on high school exam taking and performance," *Journal of Economic Behavior & Organization*, Vol. 178, pp. 534–549.

Hanushek, Eric and Steven Rivkin (2010) "Generalization about using value-added measures of teacher quality," *American Economic Review*, Vol. 100, No. 2, pp. 267–271.

Heissel, Jennifer (2021) "Teen fertility and siblings' outcomes: Evidence of family spillovers using matched samples," *Journal of Human Resources*, Vol. 56, No. 1, pp. 40–72.

Hwang, NaYoung and Cory Koedel (2022) "Holding back to move forward: The effects of retention in the third grade on student outcomes."

Imberman, Scott, Adriana Kugler, and Bruce Sacerdote (2012) "Katrina's children: Evidence on the structure of peer effects from hurricane evacuees," *American Economic Review*, Vol. 102, No. 5, pp. 2048–2082.

Jacob, Brian and Lars Lefgren (2009) "The effect of grade retention on high school completion," *American Economic Journal: Applied Economics*, Vol. 1, No. 3, pp. 33–58.

Joensen, Juanna and Helena Skyt Nielsen (2018) "Spillovers in education choice," *Journal of Public Economics*, Vol. 157, pp. 158–183.

Kao, Grace and Marta Tienda (1995) "Optimism and achievement: The educational performance of immigrant youth," *Social Science Quarterly*, Vol. 76, No. 1, pp. 1–19.

Karbownik, Krzysztof and Umut Özek (forthcoming) "Setting a good example? Examining sibling spillovers in educational achievement using a regression discontinuity design," *Journal of Human Resources*.

Kolesar, Michal and Christoph Rothe (2018) "Inference in regression discontinuity designs with a discrete running variable," *American Economic Review*, Vol. 108, No. 8, pp. 2277–2304.

Landersø, Rasmus, Helena Skyt Nielsen, and Marianne Simonsen (2020) "Effects of school starting age on the family," *Journal of Human Resources*, Vol. 55, No. 4, pp. 1258–1286.

Larsen, Matthew and Jon Valant (2018) "The long-term effects of grade retention: Eviddence on persistence through high school and college."

Lavy, Victor, Olmo Silva, and Felix Weinhardt (2012) "The good, the bad, and the average: Evidence on ability peer effects in schools," *Journal of Labor Economics*, Vol. 30, No. 2, pp. 367–414.

Lee, David and David Card (2008) "Regression discontinuity inference with specification error," *Journal of Econometrics*, Vol. 142, No. 2, pp. 655–674.

Lee, David and Thomas Lemieux (2010) "Regression discontinuity designs in economics," *Journal of Economic Literature*, Vol. 48, No. 2, pp. 281–355.

Leight, Jessica and Elaine Liu (2020) "Maternal education, parental investment, and noncognitive characteristics in rural China," *Economic Development and Cultural Change*, Vol. 69, No. 1, pp. 213–251.

Licalsi, Christina, Umut Özek, and David Figlio (2019) "The uneven implementation of universal school policies: Maternal education and Florida's mandatory grade retention," *Education Finance and Policy*, Vol. 14, No. 3, pp. 383–413.

Manski, Charles (2000) "Economic Analysis of Social Interactions," *Journal of Economic Perspectives*, Vol. 14, No. 3, pp. 115–136.

Mariano, Louis, Paco Martorell, and Tiffany Berglund (2018) "The effects of grade retention on high school outcomes: Evidence from New York City schools."

McCrary, Justin (2008) "Manipulation of the running variable in the regression discontinuity design: A density test," *Journal of Econometrics*, Vol. 142, No. 2, pp. 698–714.

Mumma, Kirsten and Marcus Winters (2023) "The effect of retention under Mississippi's test-based promotion policy."

Nickow, Andre, Philip Oreopoulos, and Vincent Quan (2020) "The impressive effects of tutoring on PreK-12 learning: A systematic review and meta-analysis of the experimental evidence."

Nicoletti, Cheti and Birgitta Rabe (2019) "Sibling spillover effects in school achievement," *Journal of Applied Econometrics*, Vol. 34, No. 4, pp. 482–501.

Nye, Barbara, Spyros Konstantopoulos, and Larry Hedges (2004) "How large are teacher effects?" *Educational Evaluation and Policy Analysis*, Vol. 26, No. 3, pp. 237–257.

Ozier, Owen (2018) "Exploiting externalities to estimate the long-term effects of early childhood deworming," *American Economic Journal: Applied Economics*, Vol. 10, No. 3, pp. 235–262.

Parman, John (2015) "Childhood health and sibling outcomes: Nurture reinforcing nature during the 1918 influenza pandemic," *Exploration in Economic History*, Vol. 58, pp. 22–43.

Persson, Petra, Xinyao Qiu, and Maya Rossin-Slater (2021) "Family spillover effects of marginal diagnoses: The case of ADHD."

Pitt, Mark, Mark Rosenzweig, and Nazmul Hassan (1990) "Inequality in the intrahousehold distribution of food in low-income countries," *American Economic Review*, Vol. 80, No. 5, pp. 1139–1156.

Qureshi, Javaeria (2018a) "Additional returns to investing in girls' education: Impact on younger sibling human capital," *Economic Journal*, Vol. 128, No. 616, pp. 3285–3319.

——— (2018b) "Siblings, teachers and spillovers in academic achievement," *Journal of Human Resources*, Vol. 53, No. 1, pp. 272–297.

Rivkin, Steven, Eric Hanushek, and John Kain (2005) "Teachers, schools, and academic achievement," *Econometrica*, Vol. 73, No. 2, pp. 417–458.

Sacerdote, Bruce (2014) "Experimental and quasi-experimental analysis of peer effects: Two steps forward?" *Annual Reviews Economics*, Vol. 6, pp. 253–272.

Sass, Tim, Jane Hannaway, Zeyu Xu, David Figlio, and Li Feng (2012) "Value added of teachers in high-poverty schools and lower poverty schools," *Journal of Urban Economics*, Vol. 72, No. 2-3, pp. 104–122.

Schwerdt, Guido, Martin West, and Marcus Winters (2017) "The effects of test-based retention on student outcomes over time: Regression discontinuity evidence from Florida," *Journal of Public Economics*, Vol. 152, pp. 154–169.

U.S. Department of Education (2020) "IDEA Part B Child Count and Educational Environments Collection."

Winters, Marcus and Jay Greene (2012) "The medium-run effects of Florida's test-based promotion policy," *Education Finance and Policy*, Vol. 7, No. 3, pp. 305–330.

Yi, Junjian, James Heckman, Junsen Zhang, and Gabriella Conti (2015) "Early health shocks, intra-household resource allocation and child outcomes," *Economic Journal*, Vol. 125, No. 588, pp. F347–F371.

Zang, Emma, Poh Lin Tan, and Phillip Cook (forthcoming) "Sibling spillovers: Having an academically successful older sibling may be more important for children in disadvantaged families," *American Journal of Sociology*.

7 Tables

Table 1. Balancing Check: Discontinuities in background characteristics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
			Child				Mother	
	3rd grade math score	Black	Hispanic	Male	Not an ELL student	FRPL eligible	Married	Some college or higher
	Panel A: Focal student with younger sibling							
Focal child below the retention cutoff	-0.015	0.013	-0.010	0.002	0.002	-0.008	-0.003	0.004
	(0.021)	(0.014)	(0.011)	(0.014)	(0.007)	(0.011)	(0.014)	(0.012)
Control group mean	-0.697	0.528	0.192	0.545	0.944	0.788	0.377	0.214
Implied imbalance (%)		2.462	5.208	0.367	0.212	1.015	0.796	1.869
Observations	19,725							
	Panel B: Younger sibling							
Focal child below the retention cutoff	N/A	0.014	-0.004	-0.018	0.010	-0.006	0.005	0.007
		(0.014)	(0.011)	(0.014)	(0.009)	(0.011)	(0.014)	(0.012)
Control group mean		0.521	0.192	0.543	0.900	0.781	0.470	0.263
Implied imbalance (%)		2.687	2.083	3.315	1.111	0.768	1.064	2.662
Observations	19,725							

Note: Local linear regressions with bandwidth +/- 25 points based on equation 1. Sample includes older focal children (Panel A) and their younger adjacent siblings (Panel B) conditional on the former being observed around Level 2 reading cutoff in third grade and the latter being observed in school records (KG to grade 2) at the same time. There is a single observation per individual in this table. Outcome variables are: 3rd grade mathematics test scores (Column 1), indicator for Black student (Column 2), indicator for Hispanic student (Column 3), indicator for male student (Column 4), indicator for student who is not in English Language Learner program (Column 5), indicator for free or reduced price eligibility in the current school year (Column 6), indicator for whether mother was married at the time of child's birth (Column 7), and indicator for whether mother had some college or higher educational attainment at the time of child's birth (Column 8). Control group mean is defined as mean of an outcome above the cutoff. Hetroskedasticity robust errors in parentheses. ***, ** and * mark estimates statistically different from zero at the 99, 95 and 90 percent confidence level.

Table 2. Main results: Effects of being flagged for third grade retention on focal child and their younger sibling

	(1) Retained	(2) Retained	(3) Math+Reading	(4) Math	(5) Reading	(6) In special education
			Panel A: Focal child with younger sibling			
	Grade 3		Grades 4 to 5			
Focal child below the cutoff	0.363***	-0.043***	0.205***	0.199***	0.210***	0.001
	(0.011)	(0.006)	(0.014)	(0.017)	(0.016)	(0.008)
Control group mean	0.032	0.084	-0.774	-0.710	-0.838	0.262
			Panel B: Younger sibling			
			Grades 3 to 5			
Focal child below the cutoff	N/A	-0.011	0.054***	0.062***	0.046**	-0.001
		(0.011)	(0.020)	(0.022)	(0.021)	(0.010)
Control group mean		0.174	-0.458	-0.400	-0.516	0.217
N	19,725	19,238	19,526	19,523	19,526	19,238

Note: Local linear regressions with bandwidth +/- 25 points based on equation 1. Sample includes older focal children (Panel A) and their younger adjacent siblings (Panel B) conditional on the former being observed around Level 2 reading cutoff in third grade and the latter being observed in school records (KG to grade 2) at the same time. There is a single observation per individual in this table. Outcome variables are measured in grades 4 to 5 for the focal child (except for Column 1 where it is grade 3 retention) and in grades 3 to 5 for their younger sibling. Outcome variables are: probability of being retained (Columns 1 and 2), averaged mathematics and reading test scores (Column 3), mathematics test scores (Column 4), reading test scores (Column 5), and probability of being enrolled in special education program (Column 6). Outcomes are aggregated over multiple grades in Columns (2) to (6) meaning that for test scores we compute average across the given grades while for indicator variables we define the event as ever occurring in given grades. Control group mean is defined as mean of an outcome above the cutoff. Hetroskedasticity robust errors in parentheses. ***, ** and * mark estimates statistically different from zero at the 99, 95 and 90 percent confidence level.

Table 3. Effects of being flagged for third grade retention on focal child and their younger sibling: Heterogeneity analysis

	(1) Focal or sibling special education	(2) Neither is special education	(3) Sibling male	(4) Sibling female	(5) Recent immigrant	(6) Not recent immigrant
			Panel A. Focal child with younger sibling			
Retained (grade 3)	0.307***	0.382***	0.369***	0.356***	0.457***	0.353***
	(0.019)	(0.013)	(0.015)	(0.015)	(0.024)	(0.011)
p-value	0.010		0.333		0.001	
Control group mean	0.033	0.031	0.045	0.016	0.056	0.024
Average test score	0.185***	0.210***	0.227***	0.183***	0.227***	0.179***
(Grades 4-5)	(0.028)	(0.016)	(0.019)	(0.020)	(0.029)	(0.014)
p-value	0.284		0.129		0.132	
Control group mean	-0.880	-0.731	-0.791	-0.754	-0.626	-0.830
			Panel B. Younger sibling			
Average test score	0.184***	-0.000	0.097***	0.008	0.144***	0.035
(Grades 3-5)	(0.041)	(0.023)	(0.029)	(0.027)	(0.050)	(0.024)
p-value	0.004		0.056		0.070	
Control group mean	-0.674	-0.414	-0.526	-0.449	-0.439	-0.532
N	5,638	13,888	9,847	9,679	3,098	13,742

Note: Local linear regressions with bandwidth +/- 25 points based on equation 1. Sample includes older focal children (Panel A) and their younger adjacent siblings (Panel B) conditional on the former being observed around Level 2 reading cutoff in third grade and the latter being observed in school records (KG to grade 2) at the same time. There is a single observation per individual in this table. Outcome variables are in rows and include probability of being retained in third grade for the focal child (first row), average mathematics and reading test scores in grades 4 to 5 for the focal child (second row), and average mathematics and reading test scores in grades 3 to 5 for the younger sibling of the focal child (third row). Test scores are computed as average scores in grades 4 to 5 for the focal child and average scores in grades 3 to 5 for their younger sibling. The main sample is divided along three dimensions: whether the household has children in special education program (Columns 1 and 2), gender of the younger sibling (Columns 3 and 4), and immigration status of the household (Column 5 and 6). Children in special education are defined as those with Individualized Education Plan (IEP). Recent immigrants are defined as those where the mother is foreign born or English is not the main language spoken at home. Control group mean is defined as mean of an outcome above the cutoff. Hetroskedasticity robust errors in parentheses. ***, ** and * mark estimates statistically different from zero at the 99, 95 and 90 percent confidence level.

Table 4. Mechanisms: Educational resources of younger sibling

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				Younger sibling outcomes			
	Pooled	Focal or sibling special education	Neither is special education	Sibling male	Sibling female	Recent immigrant	Not recent immigrant
				Panel A: Teacher value added			
Focal child below the cutoff	0.003	0.009**	0.001	0.006**	0.000	0.001	0.003
	(0.002)	(0.004)	(0.003)	(0.003)	(0.003)	(0.006)	(0.003)
p-value		0.127		0.159		0.413	
Control group mean	0.001	-0.018	0.002	-0.011	0.005	0.013	-0.005
				Panel B: Classroom peers prior achievement			
Focal child below the cutoff	0.032**	0.067**	0.016	0.044**	0.021	0.056*	0.026
	(0.013)	(0.029)	(0.014)	(0.020)	(0.018)	(0.033)	(0.016)
p-value		0.117		0.273		0.270	
Control group mean	-0.172	-0.119	-0.193	-0.164	-0.151	-0.115	-0.216
N	19,671	5,638	13,888	9,847	9,847	3,098	13,742

Note: Local linear regressions with bandwidth +/- 25 points based on equation 1. All outcomes variables are for the younger sibling of the focal child affected by the Level 2 third grade reading cutoff. There is a single observation per individual in this table. Sample includes younger siblings adjacent to their older focal sibling conditional on the latter being observed around Level 2 reading cutoff in third grade and the former being observed in school records (KG to grade 2) at the same time. Outcome variables are: teacher value added (Panel A) and achievement of classroom peers (Panel B). Teacher value added is measured as leave-out-year value added and computed using Stata command *vam* based on teacher-student linked data for 2002/03 to 2011/12 school years. Average achievement of classroom peers is computed as leave-self-out average math and reading test score using the time that a student spends in the classroom (per week) as weights. Column (1) presents pooled estimates for the full sample, Columns (2) and (3) divide the sample based on the presence of children in special education in the household, Columns (4) and (5) divide the sample based on gender of the younger sibling, and Columns (6) and (6) divide the sample based on immigration status of the household. See Table 3 for definitions of the stratification variables. Control group mean is defined as mean of an outcome above the cutoff. Hetroskedasticity robust errors in parentheses. ***, ** and * mark estimates statistically different from zero at the 99, 95 and 90 percent confidence level.

Table 5. Mechanisms: Parental school choice

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Younger sibling				
		Left school in 2 years		Below A school in prior two years: Sibling moved in 2 years to a school with higher…				
				Accountability score components		Teacher effectiveness		
	Pooled	A school at least once in prior two years	Below A school in prior two years	Reading gain score	Math gain score	Share of highly effective teachers	Share of highly effective reading teachers	Share of highly effective math teachers
Focal child below the cutoff	0.009	-0.017	0.041**	0.038**	0.015	0.052***	0.030*	0.023
	(0.013)	(0.019)	(0.021)	(0.019)	(0.019)	(0.018)	(0.018)	(0.017)
Control group mean	0.414	0.392	0.437	0.313	0.289	0.229	0.269	0.229
N	19,815	10,095	8,731	8,259	8,259	8,259	8,259	8,259

Note: Local linear regressions with bandwidth +/- 25 points based on equation 1. All outcomes variables are for the younger sibling of the focal child affected by the Level 2 third grade reading cutoff. There is a single observation per individual in this table. Sample in columns (1) to (3) includes younger siblings adjacent to their older focal sibling conditional on the latter being observed around Level 2 reading cutoff in third grade and the former being observed in school records (KG to grade 2) at the same time. Sample in Columns (4) to (8) is based on individuals whose older (focal) sibling was in a school with below A rating in the prior two years (Column 3). Outcomes variables are: probability of leaving the current school within two years (Columns 1 to 3), state-reported school-level accountability measures based on test score growth (Columns 4 and 5), school-level teacher effectiveness measures (Columns 6 to 8). We define an effective teacher as one whose leave-out-year value-added score (or subject specific score) is in the top quartile of the value-added distribution in a given year. Control group mean is defined as mean of an outcome above the cutoff. Hetroskedasticity robust errors in parentheses. ***, ** and * mark estimates statistically different from zero at the 99, 95 and 90 percent confidence level.

Appendix: For Online Publication

Appendix Tables

Table A.1. Descriptive statistics

	(1)	(2)	(3)	(4)	(5)
			Third graders		
	All	+ with birth records	+ with a younger sibling	+ with a younger sibling in school when focal child in 3rd grade	+ focal child within 25 points around retention cutoff
			Panel A: Student characteristics		
Black	0.266	0.291	0.328	0.337	0.519
	(0.442)	(0.454)	(0.469)	(0.473)	(0.500)
Hispanic	0.280	0.219	0.201	0.196	0.199
	(0.449)	(0.414)	(0.401)	(0.397)	(0.399)
Male	0.513	0.505	0.506	0.509	0.517
	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)
Not an ELL	0.907	0.961	0.960	0.960	0.927
	(0.290)	(0.193)	(0.195)	(0.197)	(0.260)
FRPL eligible	0.541	0.510	0.573	0.593	0.798
	(0.498)	(0.500)	(0.495)	(0.491)	(0.402)
In special education	0.148	0.150	0.145	0.153	0.204
	(0.355)	(0.357)	(0.352)	(0.360)	(0.403)
3rd grade math score	0.036	0.070	0.035	0.016	-0.664
	(1.014)	(0.997)	(1.036)	(1.051)	(0.764)
			Panel B: Maternal characteristics		
Some college or higher		0.399	0.400	0.398	0.221
		(0.490)	(0.490)	(0.490)	(0.415)
Married at birth		0.614	0.572	0.570	0.389
		(0.487)	(0.495)	(0.495)	(0.488)
Observations	824,640	448,974	140,919	109,028	21,953

Note: Column 1 includes all children who attended public schools in 12 anonymous school districts at any point in time between school years 2000-01 to 2011-12 and who entered third grade for the first time between school years 2002-03 and 2008-09. Column 2 restricts the sample to only children born in Florida for whom we can observe birth certificate information. Column 3 further restricts the sample to children who have a younger sibling. Column 4 restricts the sample to younger siblings that are observed in grades KG to G2 at the time when the focal child is in the third grade for the first time. Column 5 restricts the sample from Column 4 to only include families where the older focal child is within 25 points of the Level 2 reading cutoff. Student characteristics in Panel A are based on school records while maternal characteristics in Panel B are based on birth certificates.

Table A.2. Balancing check: Non-parametric estimates

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
			Child				Mother	
	3rd grade math score	Black	Hispanic	Male	Not an ELL student	FRPL eligible	Married	Some college or higher
				Panel A: Focal student with younger sibling				
Focal child below the retention cutoff	-0.026	0.022	-0.023**	-0.011	0.004	-0.002	-0.015	-0.003
	(0.025)	(0.015)	(0.010)	(0.017)	(0.008)	(0.013)	(0.016)	(0.012)
Control group mean		0.336	0.198	0.510	0.960	0.590	0.573	0.400
Implied imbalance (%)		6.548	11.616	2.157	0.417	0.339	2.618	0.750
Bandwidth	29	37	51	29	33	30	29	33
Observations	23,367	29,893	41,778	22,525	26,559	24,176	23,346	26,559
				Panel B: Younger sibling				
Focal child below the retention cutoff		-0.023	0.007	0.007	-0.009	-0.003	-0.021	-0.009
		(0.016)	(0.012)	(0.013)	(0.008)	(0.013)	(0.016)	(0.013)
Control group mean		0.335	0.201	0.503	0.919	0.589	0.644	0.425
Implied imbalance (%)		6.866	3.483	1.392	0.979	0.509	3.261	2.118
Bandwidth		34	44	54	53	33	37	41
Observations		29,893	41,778	22,525	26,559	24,176	23,346	26,559

Note: This table is a non-parametric version of results presented in Table 1 estimated using method developed by Calonico et al. (2017). Control group mean is defined as mean of an outcome above the cutoff. Hetroskedasticity robust errors in parentheses. ***, ** and * mark estimates statistically different from zero at the 99, 95 and 90 percent confidence level.

Table A.3. Differential sample attrition

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Focal child with younger sibling observed				Younger sibling of the focal child observed		
	in grade 4	in grade 5	one year later	two years later	in grade 3	in grade 4	in grade 5
Focal child below the	-0.003	-0.007	0.002	0.003	0.000	0.001	0.002
cutoff	(0.005)	(0.006)	(0.003)	(0.003)	(0.001)	(0.008)	(0.009)
Control group mean	0.950	0.942	0.987	0.985	0.920	0.821	0.681
Implied imbalance (%)	0.316	0.743	0.203	0.305	0.000	0.122	0.294
N	19,425	19,425	19,425	19,425	19,866	19,866	19,866

Note: Local linear regressions with bandwidth +/- 25 points based on equation 1. Sample includes older focal children (Columns 1 to 4) and their younger adjacent siblings (Columns 5 to 7) conditional on the former being observed around Level 2 reading cutoff in third grade and the latter being observed in school records (KG to grade 2) at the same time. There is a single observation per individual in this table. Outcomes are: probability of observing the focal child in grade 4 (Column 1), probability of observing the focal child in grade 5 (Column 2), probability of observing the focal child one year after the first time they entered third grade (Column 3), probability of observing the focal child two years after the first time they entered third grade (Column 4), probability of observing the younger sibling in grade 3 (Column 5), probability of observing the younger sibling in grade 4 (Column 6), and probability of observing the younger sibling in grade 5 (Column 7). Control group mean is defined as mean of an outcome above the cutoff. Hetroskedasticity robust errors in parentheses. ***, ** and * mark estimates statistically different from zero at the 99, 95 and 90 percent confidence level.

Table A.4. Focal child estimates: Follow-up years

	(1) Retained	(2) Retained	(3) Math+Reading	(4) Math	(5) Reading	(6) In special education
	Grade 3			Next two years		
Focal child below the cutoff	0.363***	-0.039***	35.370***	28.018***	42.814***	0.005
	(0.011)	(0.007)	(3.755)	(3.998)	(4.957)	(0.007)
Control group mean	0.032	0.055	1410	1445	1376	0.252
SD of Y			164.6	188	191.5	
Effect size (% of test score SD)			21.5	14.9	22.4	
N	19,725	19,401	19,363	19,358	19,361	19,401

Note: Local linear regressions with bandwidth +/- 25 points based on equation 1. Sample includes older focal children with younger adjacent siblings conditional on the former being observed around Level 2 reading cutoff in third grade and the latter being observed in school records (KG to grade 2) at the same time. All outcomes are for the focal (older) child in the family. There is a single observation per individual in this table. Column (1) replicates estimate from Column (1) of Table 2. Outcomes in subsequent columns are: likelihood of ever being retained in the two years after the first time we observe focal child in grade 3 (Column 2), averaged test scores in the two years after the first time we observe focal child in grade 3 (Column 3 to 5), likelihood of the student ever being in special education in the two years after the first time we observe focal child in grade 3 (Column 6). Test scores used in this analysis are developmental FCAT scale scores that are vertically-aligned, which allow us to compare student achievement across grades. Column (3) presents averaged mathematics and reading test scores, Column (4) presents mathematics test scores, and Column (5) presents reading test scores. Control group mean is defined as mean of an outcome above the cutoff. Heteroskedasticity robust errors in parentheses. ***, ** and * mark estimates statistically different from zero at the 99, 95 and 90 percent confidence level.

Table A.5. Robustness analysis: Estimation and sample choices

	(1) Baseline estimates	(2) Without control variables	(3) Calonico et al. (2017) Conventional	(4) Calonico et al. (2017) Bias-corrected	(5) Honest RD estimator	(6) 1st and 2nd-born children only	(7) Include all possible siblings	(8) Level 3 placebo cutoff
Panel A. Focal child with younger sibling								
Retained (grade 3)	0.363***	0.362***	0.356***	0.350***	0.341***	0.365***	0.378***	-0.001
	(0.011)	(0.011)	(0.011)	(0.012)	(0.015)	(0.012)	(0.009)	(0.003)
Average test score	0.205***	0.196***	0.207***	0.203***	0.175***	0.213***	0.203***	0.005
(Grades 4-5)	(0.014)	(0.016)	(0.013)	(0.016)	(0.022)	(0.015)	(0.014)	(0.010)
N	19,725	19,864	24,652	24,652	10,229	15,654	27,064	28,346
Panel B. Younger sibling								
Average test score	0.054***	0.057**	0.049**	0.054**	0.048*	0.057**	0.039**	0.018
(Grades 3-5)	(0.020)	(0.022)	(0.019)	(0.022)	(0.025)	(0.022)	(0.020)	(0.017)
N	19,725	19,864	27,405	27,405	15,713	15,654	26,020	28,346

Note: This table presents various robustness tests of the main results from Table 2 for three outcome variables that are in rows and include probability of being retained in third grade for the focal child (first row), average mathematics and reading test scores in grades 4 to 5 for the focal child (second row), and average mathematics and reading test scores in grades 3 to 5 for the younger sibling of the focal child (third row). Panel A presents estimates for the older focal child and Panel B presents estimates for their younger sibling. Column (1) replicates results from Columns (1) and (3) of Panel A and Column (3) of Panel B of Table 2, Column (2) removes control variables from regression, Columns (3) and (4) estimate the results using the non-parametric method developed by Calonico et al. (2017) with adjustment for mass points in the running variable and present conventional (Column 3) and bias-adjusted (Column 4) coefficients, Column (5) present estimates based on honestly and nearly-optimally calculated estimators, Column (6) limits the sample to families where the older focal child is first-born while the younger sibling is second-born in the family, Column (7) expands the sample to include all families with younger siblings including those who are not yet in KG at the time that the older sibling is in third grade for the first time, and Column (8) presents a placebo estimates that use Level 3 instead of Level 2 cutoff in reading. Hetroskedasticity robust errors in parentheses for Columns (1), (2), (6), (7), and (8). Calonico et al. (2017) robust standard errors in Columns (3) and (4). Nearest neighbor standard errors in Column (5). ***, ** and * mark estimates statistically different from zero at the 99, 95 and 90 percent confidence level.

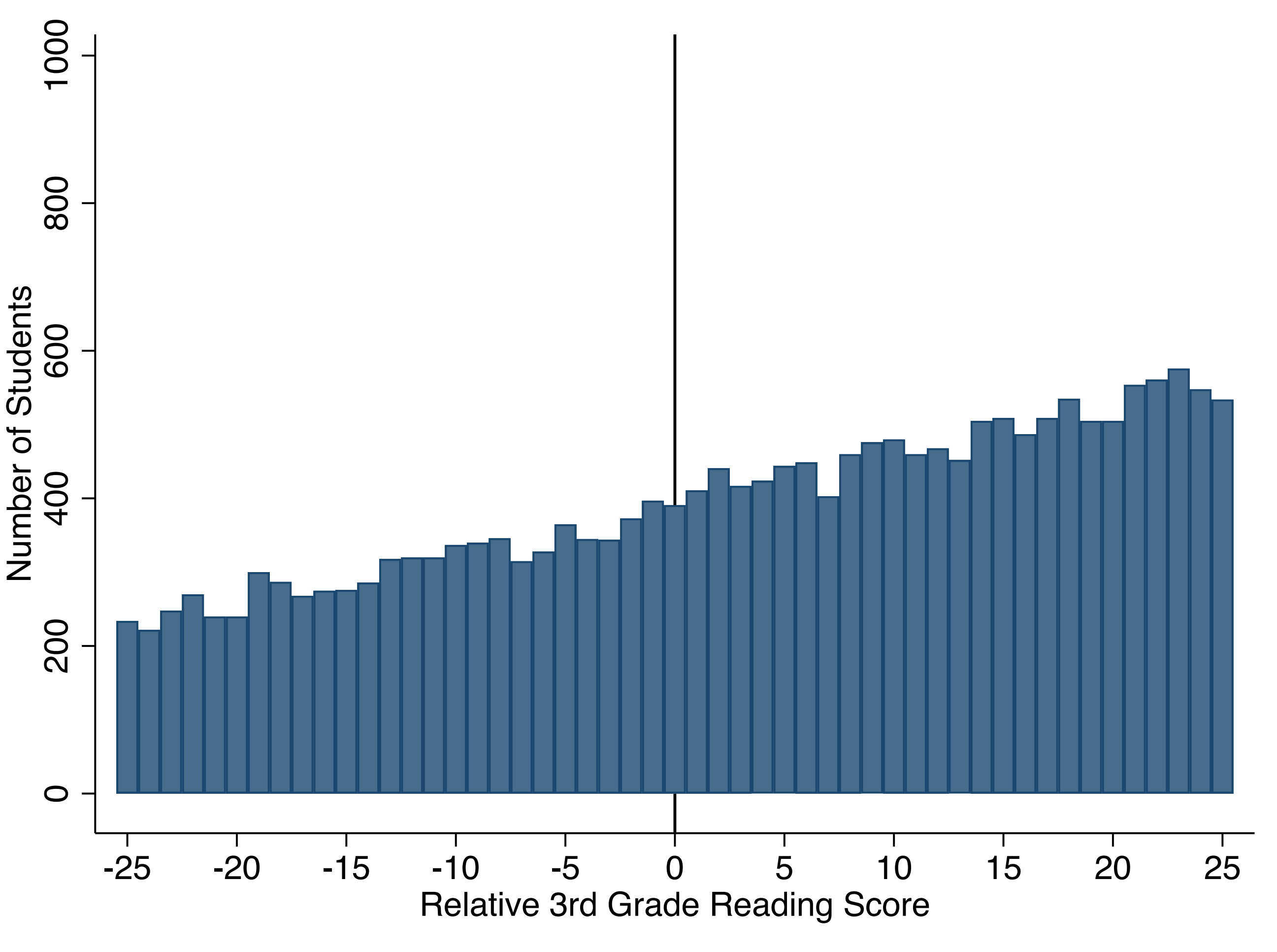
Table A.6. Descriptive mediation analysis

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				Younger sibling outcomes			
	Pooled	Focal or sibling special education	Neither is special education	Sibling male	Sibling female	Recent immigrant	Not recent immigrant
			Panel A: Replication of the results from Table 2 (column 3 of panel B) and Table 3 (panel B)				
Focal child below the cutoff	0.054***	0.184***	-0.000	0.097***	0.008	0.144***	0.035
	(0.020)	(0.041)	(0.023)	(0.029)	(0.027)	(0.050)	(0.024)
Control group mean	-0.458	-0.674	-0.414	-0.526	-0.449	-0.439	-0.532
N	19,526	5,638	13,888	9,847	9,679	3,098	13,742
			Panel B. Replicate panel A estimates for the sample for which mechanisms' observations are available				
Focal child below the cutoff	0.068***	0.185***	0.017	0.118***	0.012	0.112**	0.065***
	(0.021)	(0.043)	(0.024)	(0.030)	(0.029)	(0.053)	(0.025)
			Panel C: Descriptive mediation analysis				
Focal child below the cutoff	0.038**	0.122***	0.005	0.078***	-0.004	0.069	0.041*
	(0.019)	(0.037)	(0.021)	(0.027)	(0.026)	(0.046)	(0.022)
Control group mean	-0.441	-0.585	-0.382	-0.478	-0.397	-0.285	-0.495
N	16,822	4,767	12,055	8,329	8,493	2,662	11,686

Note: Panel A of this table reproduces results from Column 3 of Panel B in Table 2 (Column 1) and from panel B of Table 3 (Columns 2 to 7). Panel B runs the same analyses as in Panel A but on a sample restricted to individuals for whom we also observe mechanisms (teacher value added and peer composition) explored in Table 4. Panel C runs analyses from panel B while additionally controlling for teacher value added and peer composition. Control group mean is defined as mean of an outcome above the cutoff. Hetroskedasticity robust errors in parentheses. ***, ** and * mark estimates statistically different from zero at the 99, 95 and 90 percent confidence level.

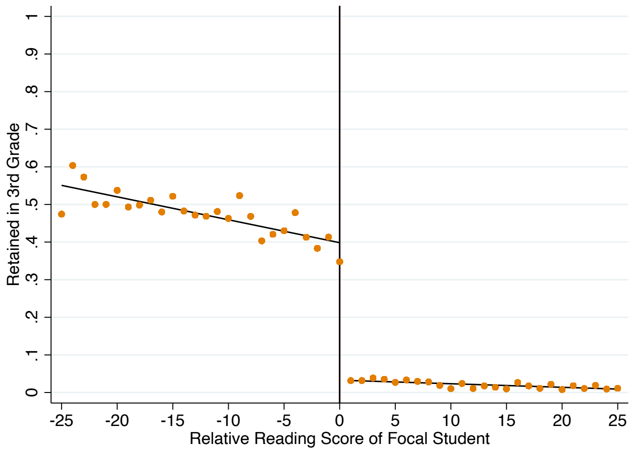
Appendix Figures

Figure A.1. Density test

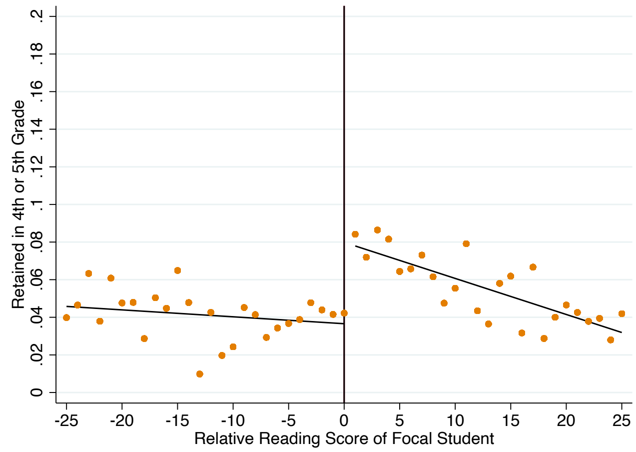


Note: Sample includes older focal children with younger adjacent siblings conditional on the former being observed around Level 2 reading cutoff in third grade and the latter being observed in school records (KG to grade 2) at the same time. Figures present histogram of density of focal children around Level 2 third grade reading cutoff (+/- 25 points). Bin width is set to 1 point. The solid vertical line indicates Level 2 cutoff.

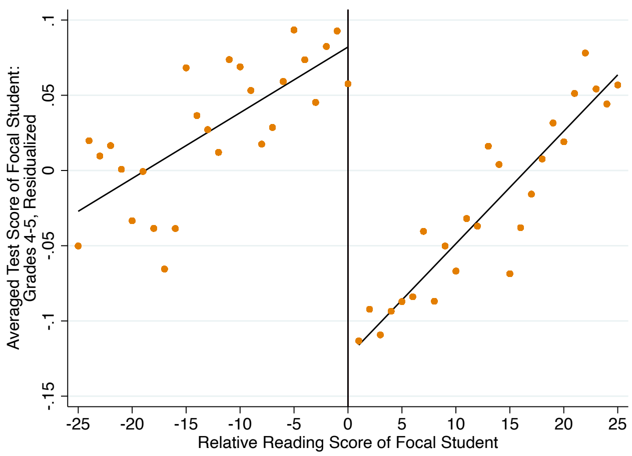
Figure A.2. Focal child outcomes: Scatterplots



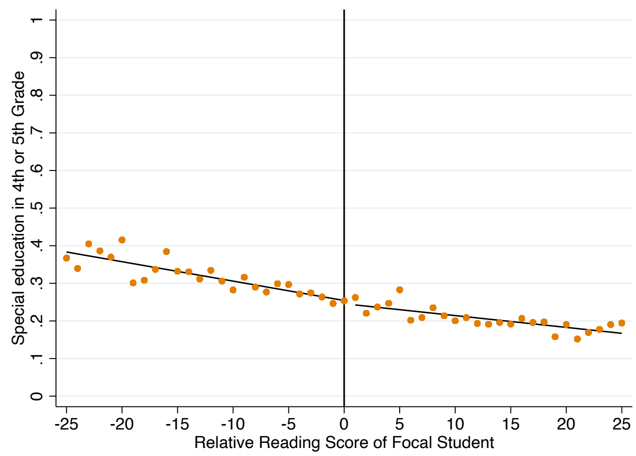
(a) P(retained in grade 3)



(b) P(ever retained in grades 4 to 5)



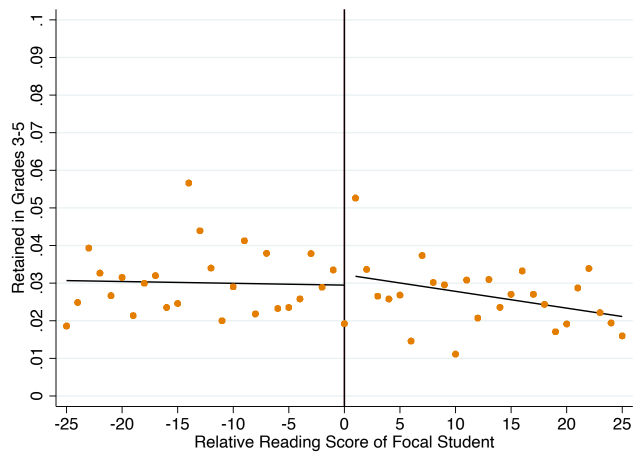
(c) Average math + reading test scores in grades 4 to 5



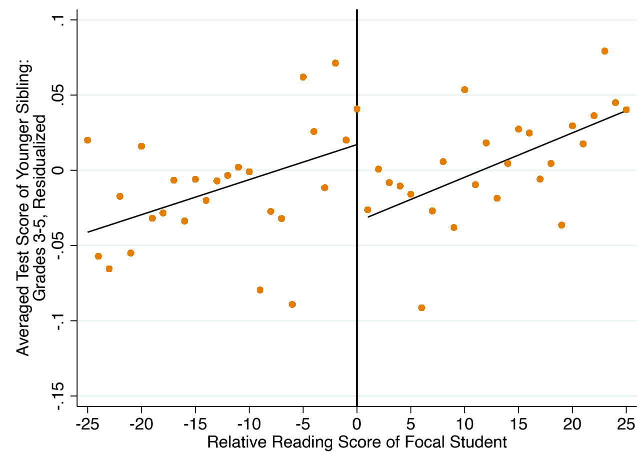
(d) P(ever in special education in grades 4 to 5)

Note: Sample includes older focal children with younger adjacent siblings conditional on the former being observed around Level 2 reading cutoff in third grade and the latter being observed in school records (KG to grade 2) at the same time. Outcomes are for the focal child (Panel A of Table 2). Each panel presents outcome variable means collapsed at the single-point level (orange circles) for the preferred bandwidth of +/- 25 points. We fit local linear regressions through these data points to the left and to the right of the cutoff. Vertical line at zero denotes Level 2 third grade reading cutoff of the focal child. Outcome variables are: probability of the focal child being retained in third grade (Panel A), probability of the focal child ever being retained in grades 4 to 5 (Panel B), averaged math and reading test scores of the focal child in grades 4 to 5 (Panel C), probability of the focal child being in special education (Panel D).

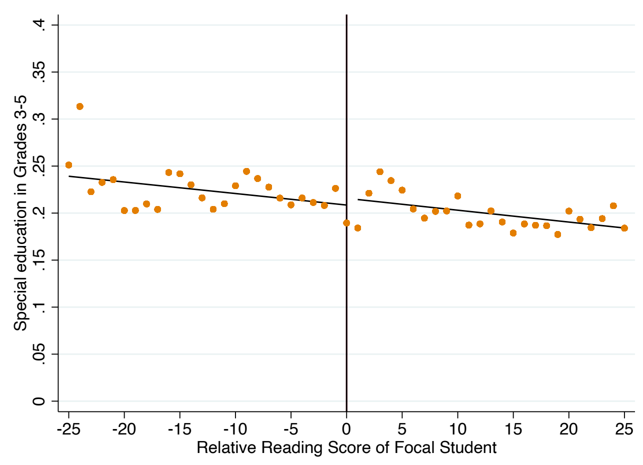
Figure A.3. Younger sibling outcomes: Scatterplots



(a) P(ever retained in grades 3 to 5)



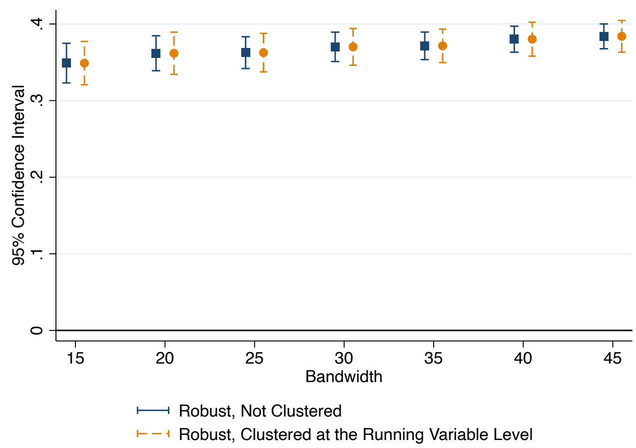
(b) Average math + reading test scores in grades 3 to 5



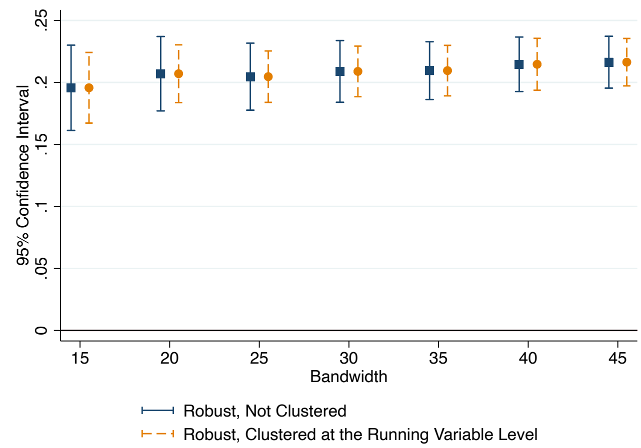
(c) P(ever in special education in grades 3 to 5)

Note: Sample includes older focal children with younger adjacent siblings conditional on the former being observed around Level 2 reading cutoff in third grade and the latter being observed in school records (KG to grade 2) at the same time. Outcomes are for the younger sibling of the focal child (Panel B of Table 2). Each panel presents outcome variable means collapsed at the single-point level (orange circles) for the preferred bandwidth of +/- 25 points. We fit local linear regressions through these data points to the left and to the right of the cutoff. Vertical line at zero denotes Level 2 third grade reading cutoff of the focal child. Outcome variables are: probability of the younger sibling ever being retained in grades 3 to 5 (Panel A), averaged math and reading test scores of the younger sibling in grades 3 to 5 (Panel B), probability of the younger sibling being in special education (Panel C).

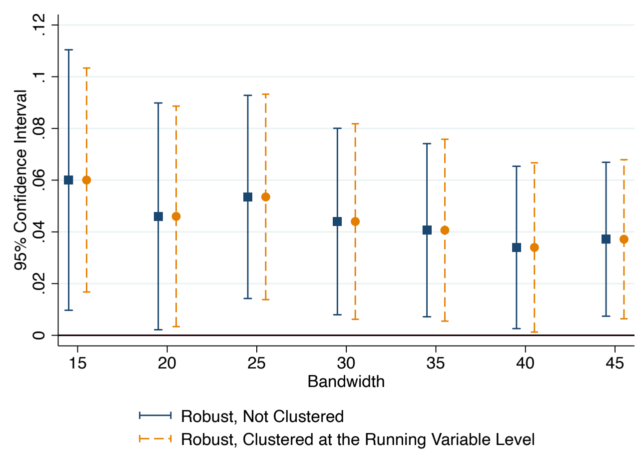
Figure A.4. Robustness to bandwidth and standard error choices



(a) P(focal child retained in grade 3)



(b) Average math + reading test scores of the focal child in grades 4 to 5



(c) Average math + reading test scores of younger sibling in grades 3 to 5

Note: This figure presents robustness analyses for the main outcome variables of interest. Sample includes older focal children with younger adjacent siblings conditional on the former being observed around Level 2 reading cutoff in third grade and the latter being observed in school records (KG to grade 2) at the same time. Outcome variables are: probability of the focal child being retained in third grade (Panel A), averaged math and reading test scores of the focal child in grades 4 to 5 (Panel B), and averaged math and reading test scores of the younger sibling in grades 3 to 5 (Panel C). Local linear regressions with varying bandwidth based on equation 1. The bandwidth varies between 15 and 45 points every 5 points. The preferred specification from Table 2 uses bandwidth of +/- 25 points. Blue squares mark estimates and 95% confidence intervals (whiskers) based on Eicker-Hueber-White heteroskedasticity robust standard errors. Orange circles mark estimates and 95% confidence intervals (whiskers) based on standard errors clustered at running variable level.