NBER WORKING PAPER SERIES

AN EXTRA YEAR TO LEARN ENGLISH? EARLY GRADE RETENTION AND THE HUMAN CAPITAL DEVELOPMENT OF ENGLISH LEARNERS

David N. Figlio
Umut Özek

Working Paper 25472
http://www.nber.org/papers/w25472

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2019

We are grateful to the Laura and John Arnold Foundation for financial support. All errors and omissions are those of the authors, and do not reflect the opinions of the twelve anonymous Florida school districts who provided us with these data for this study, nor those of the National Bureau of Economic Research.

An Extra Year to Learn English? Early Grade Retention and the Human Capital Development of English Learners

David N. Figlio and Umut Özek

NBER Working Paper No. 25472
January 2019
JEL No. I24

ABSTRACT

In this study, we use microdata from 12 Florida county-level school districts and a regression discontinuity design to examine the effects of early grade retention on the short-, medium-, and long-term outcomes of English learners. We find that retention in the third-grade substantially improves the English skills of these students, reducing the time to proficiency by half and decreasing the likelihood of taking a remedial English course in middle school by one-third. Grade retention also roughly doubles the likelihood of taking an advanced course in math and science in middle school, and more than triples the likelihood of taking college credit-bearing courses in high school for English learners. We also find that these benefits are larger for foreign born students, students with higher latent human capital in third grade as proxied by their math scores, students whose first language is Spanish, and students in lower-poverty elementary schools.

David N. Figlio
Institute for Policy Research
Northwestern University
2040 Sheridan Road
Evanston, IL 60208
and NBER
figlio@northwestern.edu

Umut Özek
American Institutes for Research/CALDER
1000 Thomas Jefferson St. NW
Washington, DC 20007
uozek@air.org

An Extra Year to Learn English? Early Grade Retention and the Human Capital Development of English Learners

1. Introduction

The United States is in the midst of the second largest wave of immigration in its history, and immigrant integration remains to be a major point of contention in political debates. A number of studies have examined the extent to which the educational and economic outcomes of these recent immigrants converge to those of natives across generations.[1] An overarching conclusion of this extant literature is that attaining English proficiency early in life, especially before middle school when students are exposed to more challenging course content, is a major predictor of future success for immigrants. As such, reducing the time to English proficiency for English learners[2]—students who come from homes where a language other than English is mainly spoken and who need additional instructional support to access the mainstream curriculum—is a top education policy priority.

This study addresses whether a commonly implemented strategy—early grade retention for English learners who have not mastered basic English literacy skills by third grade—is an effective policy lever to improve the English skills of English learners, reduce the time to proficiency, and improve their human capital accumulation in middle and high school. While grade retention has non-trivial costs—an additional year of school costs, one fewer year of potential earnings, and possibly increased probabilities of high school dropout—a grade retention

[1] For a review of the literature that examines the degree to which economic outcomes of immigrants converge to those of natives across generations, see Abramitzky and Boustan (2017). Figlio and Ozek (forthcoming) provides a review of the literature on cross-generational differences in educational outcomes of immigrants in the U.S.

[2] A number of studies have shown that it takes many English learners 4 to 10 years or more to attain proficiency, and this path is even longer for Spanish-speaking English learners and English learners from disadvantaged backgrounds (Conger, 2010; Grissom, 2004; Hakuta, Butler, & Witt, 2000; Parrish et al., 2006; Salazar, 2007; Slama, 2014; Thompson, 2012).

strategy might be worth it if it leads to accelerated English language proficiency and heightened human capital accumulation.

Specifically, we investigate the causal effects of Florida's early grade retention policy, which requires students who fail to score above the lowest achievement level on the third-grade reading test to be retained, on the English skills of English learners as measured by their reading test scores, their time to proficiency (as indicated by reclassification out of English learner status) and human capital accumulation as proxied by advanced course-taking in middle and high school using regression discontinuity (RD) design. Given the recent evidence showing the significant benefits of advanced courses in high school on postsecondary access and completion (Smith et al., 2017), taking advanced courses in middle and high school could be regarded as an early indicator of postsecondary success, and subsequent labor market gains.

Reducing the time to English proficiency is important for at least two reasons. First, the long-term adverse effects of English deficiencies while in school are expected to be more pronounced in later grades as English learners attempt to develop academic English and learn more challenging course content simultaneously. For instance, Figlio and Ozek (forthcoming) show that first-generation immigrants who enter the public school system in Florida in later grades are significantly less likely to graduate from high school, and even when they graduate, are less likely to be college-ready than observationally similar first-generation immigrants who enter in earlier grades, and these gaps are mostly explained by differences in English proficiency in later grades. Similarly, Huang et al. (2016) find that long-term English learners (students who are classified as an English learner in elementary school and fail to attain proficiency by middle school) in Arizona are almost half as likely to graduate from high school in four years compared

to students who have never been classified as an English learner, even after controlling for demographic and socioeconomic differences.

Second, educating English learners is costly. In the four states with the largest English learner populations (California, Texas, Florida, and New York), per-pupil spending for English learners is 10 (in Texas) to 50 percent (in New York) higher than the average native English speaker (Millard, 2015).[3] And these costs are expected to grow rapidly, given that English learners currently account for 10 percent of all public school students in the United States (and more than 20 percent in high-immigration districts such as Los Angeles Unified and Miami-Dade), and represent the fastest growing student population over the past two decades.[4] Therefore, reducing the time to proficiency even by one year translates into significant cost savings for many states and school districts.

Early grade retention policies have become increasingly popular in the United States over the last two decades as a general tool for student accountability (and not specifically targeted at English learners), with 16 states and the District of Columbia currently requiring the retention of third-grade students who do not meet grade-level expectations in reading. The promise of early grade retention policies for English learners is that they provide additional time for these students to acquire the necessary English skills before they are exposed to more challenging course content. Furthermore, as we describe below, Florida's third-grade retention policy might be particularly beneficial for English learners to attain proficiency earlier as it provides substantial

[3] In the 34 states that use the state's funding formula to fund English learner education, the average pupil spending for ELLs is 30 percent higher than non-English natives.

[4] In 2014, English learners constituted roughly 10 percent of all public school students in the United States. While it is difficult to find historical numbers for English learners in the U.S. due to differences in ELL definitions across agencies, Ballantyne et al. (2008) has shown that the English learner population nationwide has grown by almost 57 percent between 1995 and 2005, whereas non-English learners grew by 3.7 percent during the same time frame.

instructional support in reading for retained students, including at least 90 minutes of reading instruction each day using effective instruction strategies and high-performing teachers.

There exists extensive literature examining the effects of grade retention on student outcomes, yet very little is known about their effects on English learners. Recent studies using quasi-experimental designs to identify the causal effect of retention find significant academic benefits of grade retention, especially in early grades in the short term (Roderick & Nagaoka, 2005; Jacob & Lefgren, 2004; Jacob & Lefgren, 2009; Greene & Winters, 2007; Greene & Winters, 2012; Schwerdt et al., 2017). For instance, in an RD framework, Jacob and Lefgren (2004, 2009) find that retention and mandatory summer school had a small positive effect on achievement in the short term for third graders but not for sixth graders in Chicago.

There are several studies to date that examine the effects of the third-grade retention policy in Florida, on which we focus in this study. For instance, Schwerdt et al. (2017) find that although retained students outperform their same-age peers in the short term, these achievement gains fade out entirely after 6 years. However, retained students under Florida's retention policy significantly outperform their promoted peers when they reach the same grade level (Greene & Winters, 2007; Greene & Winters, 2012; and Schwerdt et al., 2017). Retained students are also less likely to be retained in a later grade and no more or less likely to graduate from high school (Schwerdt et al., 2017). That said, to the best of our knowledge, there is no study to date that examines the causal effects of test-based early grade retention on English learners.

We find significant benefits of early grade retention on English learners. In particular, we find significant effects of retention on the reading scores of English learners that persist through middle school, reducing the time to proficiency by about a year, decreasing the likelihood of taking a remedial English course in middle school by one-third, almost doubling the likelihood

of taking an advanced course in math and science in middle school, and more than tripling the likelihood of taking college credit-bearing courses in high school. These benefits are larger for foreign born English learners, English learners with higher latent human capital in third grade as proxied by their math scores, English learners attending lower-poverty elementary schools, and Spanish-speaking English learners.

2. Third-Grade Retention Policy in Florida

Since 2002, all third-grade students in Florida are required to meet the Level 2 benchmark or higher (the second lowest of five achievement levels) on the statewide reading test in order to be promoted to fourth grade. The main objective of this policy, which was part of the broader "Just Read, Florida!" initiative, is to ensure that all students meet the reading benchmarks described in Florida's Sunshine State Standards before they reach the fourth grade, when students traditionally begin to "read to learn" rather than "learn to read." The legislation requires that schools provide development strategies for retained students. These include proven-effective teaching strategies, assigning retained students to high-performing teachers, participation in summer reading camps, and at least 90 minutes of reading instruction each day. If the retained student can demonstrate the required reading level before the beginning of the following school year or during the school year, he/she might be eligible for mid-year grade promotion.

There are a number of "good cause exemptions" that allow students to be promoted to the fourth grade despite failing to score at the Level 2 benchmark or above. The most relevant exemption for the purposes of this study is the English learner exemption, which allows students to be promoted to fourth grade if they have been in the English learner program for less than two

years. As such, the third-grade retention policy impacts only "long term" English learners who have not been able to attain proficiency in at least 3 years. Students are also eligible for a good-cause exemption if they have certain disabilities, or have received intensive reading remediation for two years and have already been retained twice between kindergarten and third grade. Additionally, students are able to obtain an exemption by demonstrating that they are reading at a level equal to or above a Level 2 on the FCAT by performing at an acceptable level on an alternative standardized reading assessment approved by the State Board of Education, or by demonstrating proficiency through a teacher-developed portfolio.

3. Data

In our analysis, we make use of detailed longitudinal, student-level administrative data that cover school years between 2000–01 and 2011–12 from 12 anonymous, county-level school districts in Florida. In particular, the student-level administrative data contain FCAT reading and math scores for all students between grades three and ten in these districts during that time frame along with a wealth of student characteristics including student demographics (e.g., race/ethnicity, gender), whether the student is eligible for subsidized meals, measures of English proficiency (limited English proficiency status indicator and language spoken at home), country of birth, and special education status. We are also able to link these student-level administrative data with course enrollment data that provide detailed information about courses taken and course type (e.g., subject; remedial, regular, or advanced) in middle and high school.

In our analysis, we focus on four cohorts of English learner students who enter third grade for the first time between 2002–03 (the first year of the grade retention policy) and 2005–06 school years, and follow them until the 2011–12 school year when the retained students in our youngest cohort reach the eighth grade. Roughly 65 percent of these students had been in the

English learner program for at least 2 years when they first entered the third grade and hence were not eligible for the good-cause exemption. Table 1 compares the students in our sample with their non-English native peers in the same cohorts. English learner students have significantly lower reading and math scores in the third grade, are almost three times more likely to score below the retention cutoff, and twice as likely to be retained at the end of the third grade compared to non-English native students. Furthermore, English learner students are significantly more likely to come from disadvantaged backgrounds: 80 percent of them are classified as eligible for subsidized meals (compared to 52 percent among non-English native students), are more than three times more likely to be Hispanic, and are roughly seven times more likely to be foreign born.

The main challenge with obtaining rigorous evidence on the causal effects of grade retention on student outcomes is that the students to be retained are not selected randomly. Therefore, regression-adjusted differences based on observable student attributes between promoted and retained students are likely to yield biased inferences. Table 2 illustrates these differences between English learner students who were retained and English learner students who were promoted at the end of the third grade, and shows that, compared to promoted English learner students, English learner students who were retained at the end of the third grade had significantly lower test scores in third grade, were significantly more likely to come from disadvantaged backgrounds, and more likely to be special education students. To deal with this selection issue, we utilize the non-linearity created by the retention policy and compare students who scored right below and right above the promotion cutoff in an RD framework. In what follows, we detail this empirical approach.

4. Empirical Framework

In our analysis, we rely on RD design using the student-level treatment cutoffs based on the third-grade reading scores of students. Let S_i denote the difference between the score of student *i* in the third-grade reading test and the retention cutoff—with negative values indicating scores below cutoff—and T_i denote an indicator for retained students. A common regression model representation of this evaluation problem would then become:

$$Y_i = \alpha + \beta T_i + \epsilon_i \quad (1)$$

where Y_i is the outcome of interest (e.g., test scores, time to proficiency, advanced or remedial course-taking). Since students on both sides of the retention cutoff can be promoted or retained under the Florida policy, we utilize a fuzzy RD design where the effect of the treatment on the treated is given by:

$$\beta = \frac{lim_{S\uparrow 0}E[Y|S] - lim_{S\downarrow 0}E[Y|S]}{lim_{S\uparrow 0}E[T|S] - lim_{S\downarrow 0}E[T|S]}. \quad (2)$$

There are several ways to estimate β in this context. The first is to estimate Equation (2) non-parametrically using kernel-weighted local polynomial smoothing. However, when the selection variable is discrete—as in this case—the non-parametric estimator might lead to biased estimates, since 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 β parametrically using the following two-stage least squares (2SLS) framework:

$$T = \gamma + \delta B + k(S) + k(S) * B + \upsilon \quad (3)$$

and the fitted value of *T* is used in a second stage:

$$Y = \rho + \tau \hat{T} + k(S) + k(S) * \hat{T} + \varepsilon. \quad (4)$$

where $k(S)$ is a polynomial function of the relative test score, and B is an indicator for students below cutoff. We estimate this model using the linear polynomial specification and different bandwidths (i.e., restrict the estimation sample to students within different score ranges)

to check the robustness of our findings, and two-way cluster the standard errors at the relative test score and school levels (Lee and Card, 2008).

In this empirical framework, τ will yield unbiased estimates of the causal effect of the treatment if (a) there is a significant discontinuity in treatment at the cutoff (large denominator in Equation (2)); and (b) all other student attributes are smooth around the cutoff. Figure 1 empirically checks to make sure that the first condition is satisfied, and presents the percentage of retained English learners by their relative third-grade reading score. This figure shows that while the retention rate declines as we approach the cutoff from the left, English learners whose reading scores fell just below the retention cutoff were still approximately 25 percentage points more likely to be retained compared to students right above the cutoff.

While the second condition cannot be definitively proven, we conduct several tests. First, we examine if the observable characteristics of students are continuous by estimating the following models:

$$X = \alpha + \beta B + k(S) + k(S) * D + \varepsilon \quad (5)$$

where X represents baseline student characteristics including third-grade math scores. In Table 3, we present falsification tests where each row represents a separate regression using the identified variable as the dependent variable estimated using a bandwidth of 10 and 20 points and the linear polynomial specification, and the estimated coefficient indicates the size of the discontinuity. Figure 2 presents further evidence plotting the mean value of observable student characteristics against reading scores in third grade close to the cutoff. Consistent with our identification assumption, we observe no significant discontinuity at the cutoff.

Third, we check for the possibility of selection variable manipulation as noted in McCrary (2008), even though this is very unlikely in this context since FCAT scores are assessed

without any teacher, student, or principal involvement. Figure 3 presents the distribution of students around the retention cutoff, and shows that the number of students in each bin seems to be increasing as the retention cutoff falls on the left tail of the normally distributed reading scores, but the results present no unusual discontinuity at the cutoff and hence no evidence of strategic sorting around the cutoff. We reject the hypothesis on discontinuity in the density of the distribution at the cutoff, with a p-value of 0.993 (Frandsen, forthcoming).

5. Results

5.1. Estimated Effects on Reading Achievement in Elementary and Middle School

We first examine the effects of early grade retention on English learners' English skills as proxied by their reading test scores in elementary and middle school. Similar to Schwerdt et al. (2017), we conduct (1) same-age analysis, comparing the test scores of students around the retention cutoff in the years following the first time these students enter third grade, and (2) same-grade analysis, comparing their reading achievement when both retained and promoted students reach the same grade level. In both analysis, we use vertically aligned developmental test scores that are comparable across grades. We present the 2SLS results of the same-age analysis in the top panel of Table 4 using the linear polynomial specification and bandwidths of 10 and 20 points, whereas the top panel of Figure 4 provides a visual inspection of the discontinuity in reading scores in the years following third grade at the retention cutoff using local linear smoothing with a bandwidth of 20 points. The results suggest large benefits of retention on the reading scores of English learners in the short run, starting with 26 to 30 percent of the standard deviation in the first year, peaking at roughly 0.6σ within two years, and reverting back to 0.35σ to 0.4σ within three years. These benefits dissipate quickly when the

retained students enter middle school, with fourth-year effects of 0.08σ to 0.17σ, and fifth-year effects of 0.05σ to 0.10σ. While three out of the four estimates in years 4 and 5 are no longer statistically different than zero at conventional levels, they still represent meaningful-magnitude estimated gains in reading achievement in middle school. Middle school students in our data typically gain about 30 percent of a standard deviation per year on the current-year reading test, so these effect sizes, at minimum, represent gains comparable to about a month of instruction.[5]

The bottom panels of Table 4 and Figure 4 repeat the same analysis, this time comparing the developmental test scores of English learners around the retention cutoff when they reach the same grade level. We find sizable effects on reading scores that persist through eighth grade, with effect sizes of roughly 0.5σ in eighth grade, providing evidence that just-retained students are better prepared for more challenging course content in middle school than their just-promoted peers. These test-score findings for English learners are mostly in line with—and, in some cases, slightly larger than—the overall effects reported in Schwerdt et al. (2017).

5.2. Estimated Effects on Time to Proficiency

There are two important implications of improved English skills for retained English learners. First, these students could attain English proficiency and leave the English learner program earlier than their promoted peers. The top panel of Table 5 presents the 2SLS estimates where the dependent variable is the number of years between the first time an English learner enters the third grade and his/her reclassification out of English learner status, and Figure 5 presents the local linear smoothing of the years-to-proficiency variable on the relative reading

[5] Based on authors' calculations using vertically aligned developmental reading scores for middle school students in our sample. Specifically, we calculated the average year-to-year change (difference between year $t+1$ and year t) in developmental reading scores for sixth and seventh graders in terms of the standard deviation of the current year (year t) reading test scores.

score, calculated separately for each side of the cutoff using the triangle kernel and the bandwidth of 20 points, with the solid circles representing the averaged outcome variable for each test score. The results indicate that retention reduces time to proficiency by a year for English learners. Given that English learners whose third-grade reading scores fell right above the cutoff spend an additional 2 years in the English learner program, this effect size implies that grade retention reduces additional time spent in English learner program by half.

The second panel of Table 5 and the middle panel of Figure 5 repeat the same analysis, replacing the time-to-proficiency variable with a series of English learner indicators in the years following the first time students enter the third grade. The findings indicate that retention reduces the likelihood of being identified as an English learner by 7 to 20 percentage points in the first year (or 10 to 30 percent of the control mean), by 15 to 18 percentage points within 2 years (or 31 to 38 percent of the control mean), and by 21 to 22 percentage points (or roughly 64 percent of the control mean) in the third year. These benefits also persist in middle school even though the estimated effects are not as precisely estimated in some cases—retention reduces the likelihood of being identified as an English learner in the fourth year by almost 70 percent of the control mean. The results of the same-grade analysis, reported in the third panel of Table 5 and the bottom panel of Figure 5, reveal similar findings and show that retention significantly increases the likelihood of exiting the English learner program by the time English learners enter middle school.

5.3. Estimated Effects on Middle and High School Course-Taking

Another potential benefit of Florida's early grade retention policy for English learners is that it might reduce the need for future remediation by helping these students acquire the

necessary English skills for more challenging course content. For example, during the time frame we examine in our study, Florida required middle and high school students who scored in the lowest two achievement levels on prior-year reading tests to take remedial reading courses.[6] Therefore, early grade retention could reduce the likelihood of taking remedial reading courses in middle school due to the aforementioned positive effects of retention on the reading scores of English learners. The first row of Table 6 presents the estimated effects of grade retention on the likelihood of taking at least one remedial reading course in middle school (we present graphical evidence in the first panel of Figure 6).[7] The results indicate that early grade retention leads to a decline in remedial course-taking in middle school by 15 to 25 percentage points, roughly equivalent to a 20 to 35 percent reduction relative to the control mean at the cutoff.

The reduced need for English remediation could also free up time and allow these students to take more advanced courses in middle school.[8] We explore this possibility in rows (II) to (IV) of Table 6 and panels (B) to (D) of Figure 6, where we examine the effects of early grade retention on whether the student (I) took an advanced language arts course in middle school, (II) took an advanced math course in middle school, and (III) took an advanced science course in middle school. These findings point to significant effects of grade retention on the middle school course-taking behavior among English learners. In particular, we find that grade retention increases the likelihood of taking an advanced language arts course by 55 to 68 percent,

[6] This is similar to many other college readiness initiatives recently implemented nationwide, which require students that are deemed not on track based on their standardized test scores to take remedial courses. For example, under the Targeted Interventions Program in Kentucky, one of the earliest adopters of college- and career-readiness benchmarks, performance standards have been established for students in grades 8, 10, and 11 using standardized tests, and students deemed not on track to meet subsequent college readiness indicators are required by law to take transitional courses before they exit high school.

[7] We identify remedial courses as those that are labeled "Remedial" in the Florida Department of Education's course code directory.

[8] Advanced courses in middle school are those labeled level-3 courses (the most advanced out of three categories) in the Florida Department of Education's Course Code Directory (CCD) for years between 2005–06 and 2011–12.

roughly doubles the likelihood of taking an advanced math course, and increases the likelihood of taking an advanced science course by 69 to 90 percent among English learners.

These effects on middle school course-taking behavior could better prepare English learners for high school courses and lead them to take college credit-bearing courses in high school such as Advanced Placement[9] or International Baccalaureate (IB) courses, potentially better preparing them for college.[10] While we are unable to observe course-taking behavior throughout high school for any of our cohorts, we examine the effects of grade retention in third grade on the likelihood of taking an AP or IB course before the 12th grade for our oldest cohort (students who entered third grade for the first time in 2002).[11] Once again, our results point to a sizable effect of early grade retention on course-taking among English learners—the estimates, presented in the last row of Table 6 (graphical inspection provided in the last panel of Figure 6) reveal that early grade retention increases the likelihood of taking college credit-bearing courses in high school for English learners by 31 to 44 percentage points, more than tripling the probability of taking these courses compared to a control mean of 15 percent at the cutoff.

5.4. Subgroup Analysis

Do certain English learner subgroups benefit more from grade retention compared to others? While we do not have sufficient power to statistically distinguish the estimated effects of

[9] Advanced Placement® and AP® are trademarks registered and owned by the College Board, which is not affiliated with, and does not endorse, this publication.

[10] Florida offers a wide array of high school acceleration programs for students to earn college credit while in high school, including Advanced Placement (AP), International Baccalaureate (IB), Advanced International Certificate of Education (AICE), and dual enrollment (DE) courses. In our main analysis, we focus on AP, IB, and AICE course-taking in math, science, English, and social studies because these courses are offered in the K-12 system and hence included in our data.

[11] That said, taking at least one AP/IB course by the end of 11th grade is a good predictor of the likelihood of taking one of these courses in high school. For example, when we examine the earlier high school cohorts in our sample (those that entered third grade before Florida's grade retention policy), less than 10 percent of students who did not take at least one AP/IB course by the end of 11th grade took one of these courses in 12th grade.

retention across different English learner groups, panels (A) through (J) in Figure 7 present the results of an exploratory analysis providing the 2SLS estimates (along with their 95% confidence intervals) for different student groups of interest.

For instance, Florida's grade retention policy could be more effective in improving the outcomes of immigrant students who recently relocated to the U.S. and might simply need more time to learn English before facing more challenging course content. In contrast, the academic struggles of native-born English learners might be driven by factors beyond English deficiencies. The first set of bars in each figure presents the breakdown by student nativity. While the short-term effects of retention on reading scores are comparable for U.S. born and immigrant English learners, the longer-term benefits are much larger for the latter group and are statistically distinguishable from zero. The effects on remedial English course-taking in middle school and AP/IB course-taking in high school are also larger for foreign born English learners, yet the effects are comparable for all other outcomes of interest.

One of the overarching conclusions of the extant literature on English learners is that their pathways to proficiency vary significantly by their native language (e.g., Conger, 2010; Slama, 2014; Thompson, 2012). The second set of bars in each figure compares the estimated effects for Spanish speakers (who constitute roughly 70 percent of all English learners in our sample) with speakers of other languages. The benefits of retention are slightly larger for the former group in many cases, with the exception of the effects on remedial English-taking in middle school and AP/IB course-taking in high school, where we observe significantly larger effect sizes for Spanish-speaking English learners that are statistically distinct from zero.

Finally, the effects of grade retention could be larger for English learners with larger latent human capital in third grade as proxied by their third-grade math scores, or for students in

elementary schools in more affluent neighborhoods (as measured by the percentage of students who are eligible for subsidized meals) that could have more resources to provide the instructional support for retained students. The third and fourth set of bars investigate whether the estimated effects of retention on English learners are different based on the student's third-grade math proficiency or the poverty of the school the student attended in third grade. The results provide strong evidence that the benefits of grade retention are higher for English learners with larger latent human capital. For instance, we find that students with higher math scores in third grade experience a decline of 1.5 years (or almost an 80 percent drop compared to the control mean at the cutoff) in time to proficiency because of retention (compared to a decline of 0.6 years for students who were not proficient in math in third grade). The results also suggest that the benefits of retention are substantially larger for English learners in schools serving more affluent students. For example, retention leads to considerably larger increases in reading test scores for these students who persist longer, reduces time to proficiency by 1 year (compared to 0.6 years for students in high-poverty schools), and leads to a much larger increase in the likelihood of taking advanced courses in high school.

5.5. Robustness Checks

Figure 8 checks the sensitivity of our main findings to bandwidth selection and provides the effects (and the 95% confidence intervals) of retention estimated using different bandwidths and the linear polynomial specification. For all of our outcomes, the estimated discontinuities at the cutoff are extremely robust to bandwidth selection, and our previous conclusions remain unchanged.

Another potential challenge to identification in this context is the possibility that students whose third-grade reading test scores fall in the lowest achievement level are subject to other interventions targeting low-performing students. This might imply, for instance, that the observed differences between students around the cutoff are driven by other student-level policies instead of grade retention. To test this hypothesis, we constructed "pseudo" cutoffs using earlier cohorts who were not subject to the retention policy (English learners who first entered third grade in 2001–02) and checked the discontinuities in outcomes. Figure 9 presents these findings using the linear specification and different bandwidths, and shows that there are no statistically significant discontinuities in our main outcome variables at these pseudo-cutoffs. These findings provide further evidence that the observed differences in outcomes between students around the retention cutoff are indeed driven by the retention policy.

Finally, differential attrition might lead to biased estimates if retained students leave the sample at higher rates than their promoted peers. To test this possibility, Figure 10 checks the discontinuities in attrition rates around the cutoff in the 5 years following the first time students enter the third grade, and by the time students enter 11th grade (for the 2002 cohort only). We find no significant discontinuities in these attrition rates around the cutoff, providing evidence that our results are not driven by differential attrition around the retention cutoff.

6. Concluding Remarks

English learners represent the fastest growing student group in the United States. As such, the achievement gaps between English learners and non-English natives have been well documented in the literature (Genesee et al., 2005; Hakuta, 2011; Mitchell, 2015; Saunders and Marcelletti, 2013; and Huang, 2016). An overarching conclusion of this literature is that these

gaps are significantly larger between long-term English learners (students who are classified as English learners in elementary school and fail to attain proficiency by middle school) and their English-proficient peers. Therefore, in order to close English learner achievement gaps, it is crucial to reduce the time to proficiency among these students.

In this study, we examine the effects of Florida's third-grade retention policy on English learners. This policy is particularly relevant for English learners as the retention decision is mainly based on student achievement on reading tests, and because the retained students receive substantial instructional support in reading (including at least 90 minutes of reading instruction each day using effective instruction strategies and high-performing teachers). We find that early grade retention significantly improves the English skills as proxied by reading test scores and reduces time to proficiency among English learners. Furthermore, it reduces the need for future remediation in middle school and increases the likelihood of taking advanced courses in middle and high school, potentially better preparing English learners for college. We also find that these benefits are larger for recent immigrants who were born in another country and relocated to the United States, for students with higher latent human capital in third grade as proxied by their third-grade math scores, and for students in lower-poverty schools.

How about cost effectiveness? The main argument against grade retention policies is that they are costly to taxpayers. For example, Warren et al. (2014) estimate that roughly 1.5 percent of the 50 million public school students were retained in 2009–10. With average per-pupil spending of approximately $12,000, the estimated cost of retention to taxpayers in 2009–10 was $9 billion and affected about 750,000 children nationwide. This estimate ignores two additional costs. First, grade retention may impose private costs, such as the delay of entry into the labor force. Eide and Goldhaber (2005) estimate that being retained in the third grade in the 1990s cost

the average student $22,251 to $35,475 due to foregone wages, depending on the educational attainment of the student. Furthermore, grade retention might impose a significant emotional burden on students because they are stigmatized as failing and they face the challenges of adjusting to new peers, which might in turn lead to student disengagement from schooling.

That said, these costs could be justified if grade retention policies increase student achievement and human capital, which in turn could lead to school districts/states spending less on remedial education for vulnerable student populations (e.g., if fewer students take remedial courses in middle or high school, or if grade retention leads to English learners being reclassified out of English learner programs sooner), and/or lead to higher wages in the long run because of higher human capital accumulation. We find evidence supporting this hypothesis. Therefore, early grade retention policies with instructional support might be a cost-effective way to improve the long-term outcomes of English learners. Our results also provide evidence that grade retention for English learners could be more efficacious in cases where the second language is still a dominant language (e.g., Spanish in the Florida context).

References

Abramitzky, Ran and Leah Platt Boustan. 2017. Immigration in American Economic History. Journal of Economic Literature 55(4): 1311-1345.

Ballantyne, K. G., Sanderman, A. R., & Levy, J. (2008). Educating English language learners: Building teacher capacity. Washington, DC: National Clearinghouse for English Language Acquisition. Retrieved from http://www.ncela.gwu.edu/files/uploads/3/EducatingELLsBuildingTeacherCapacityVol1.pdf

Conger, D. (2010). Does bilingual education interfere with English-language acquisition? Social Science Quarterly, 91(4), 1103–1122.

Eide, E.R., and Goldhaber, D.G. (2005). Grade Retention: What Are the Costs and Benefits? Journal of Education Finance, 31(2), 195–214.

Figlio, D. & Ozek, U. (forthcoming). Cross-generational Differences in Educational Outcomes in the Second Great Wave of Immigration. Education Finance and Policy.

Frandsen, B. (forthcoming). Party bias in union representation elections: testing the manipulation in the regression discontinuity design when the running variable is discrete. Advances in Econometrics.

Genesee, F., Lindholm-Leary, K., Saunders, W., & Christian, D. (2005). English learners in U.S. schools: An overview of research findings. Journal of Education for Students Placed at Risk, 10(4), 363–385.

Greene, J.P. & Winters, M. A. (2007). Revisiting grade retention: An evaluation of Florida's test-based promotion policy. *Education Finance and Policy,* 2(4), 319–340.

Greene, J.P. & Winters, M. A. (2012). The medium-run effects of Florida's test-based promotion policy. *Education Finance and Policy,* 7(3), 305–330.

Grissom, J. B. (2004). Reclassification of English learners. Education Policy Analysis Archives, 12(36), 1–36.

Hakuta, K. (2011). Educating language minority students and affirming their equal rights: Research and practical perspectives. Educational Researcher, 40(4), 163–174.

Hakuta, K., Butler, Y., & Witt, D. (2000). How long does it take English learners to attain proficiency? (No. 2000-1). Berkeley, CA: University of California Linguistic Minority Research Institute.

Huang, M., Haas, E., Zhu, N., and Tran, L. (2016). High school graduation rates across English

learner student subgroups in Arizona (REL 2017–205). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory West. Retrieved from http://ies.ed.gov/ncee/edlabs.

Jacob, Brian A. and Lars Lefgren. 2004. Remedial education and student achievement: A regression-discontinuity analysis. The Review of Economics and Statistics 86(1): 226–244.

Jacob, B.A. & Lefgren, L. (2009). The effect of grade retention on high school completion. *American Economic Journal: Applied Economics, 1(3),* 33-58.

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

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

Millard, M. (2015). State funding mechanisms for English language learners. Denver, CO: Education Commission of the States.

Mitchell, C. (2015, June 10). English language learners head to the top of the class in Houston. Education Week, p. 9.

Parrish, T., Merickel, A., Perez, M., Linquanti, R., Socias, M., Spain, A., & Delancey, D. (2006). Effects of the implementation of Proposition 227 on the education of English learners, K-12: Findings from a five-year evaluation. Final report for AB 56 and AB 1116. Washington, DC: American Institutes for Research and WestEd.

Roderick, M. & Nagaoka, J. (2005). Retention under Chicago's high stakes testing program: Helpful, harmful, or harmless? *Educational Evaluation and Policy Analysis* 27(1), 309–340.

Saunders, W. M., & Marcelletti, D. J. (2013). The gap that can't go away: The catch-22 of reclassification in monitoring the progress of English learners. Educational Evaluation and Policy Analysis, 35(2), 139–156.

Salazar, J. J. (2007). Master plan evaluation report for English learners—2005/2006. Los Angeles, CA: Los Angeles Unified School District, Program Evaluation and Assessment Branch.

Schwerdt, Guido, Martin R. West, and Marcus A. Winters. 2017. The Effects of Test-Based Retention on Student Outcomes over Time: Regression Discontinuity Evidence from Florida. Journal of Public Economics, 152, 154–169.

Slama, R.B. (2014). Investigating Whether and When English Learners Are Reclassified Into

Mainstream Classrooms in the United States: A Discrete-Time Survival Analysis. American Educational Research Journal, 51(2), 220–252.

Smith, J., Hurwitz, M., & Avery, C. 2017. Giving college credit where it is due: Advanced Placement exam scores and college outcomes. Journal of Labor Economics, 35(1), 67–147.

Thompson, K. (2012). Are we there yet?: Exploring English learners' journey to reclassification and beyond. Unpublished manuscript, Stanford University.

Warren, J.R., Hoffman, E., & Andrew, M. (2014). Patterns and Trends in Grade Retention Rates in the United States, 1995–2010. Educational Researcher, 43(9), 433–443.

Table 1. Student Characteristics of First-Time Third Graders Between 2002–03 and 2005–06 School Years, by English Learner Status

	English Learners	Non-English Learner Students
Third-grade reading score	-0.819	0.0202
	(1.009)	(0.974)
Third-grade math score	-0.744	-0.0803
	(1.020)	(0.972)
Below the retention cutoff	0.451	0.165
	(0.498)	(0.371)
Retained	0.185	0.0912
	(0.388)	(0.288)
Eligible for subsidized meals	0.801	0.520
	(0.399)	(0.500)
White	0.0638	0.433
	(0.244)	(0.496)
Black	0.159	0.275
	(0.366)	(0.447)
Hispanic	0.727	0.226
	(0.445)	(0.418)
Male	0.525	0.508
	(0.499)	(0.500)
Special education	0.129	0.139
	(0.335)	(0.346)
Foreign born	0.491	0.0658
	(0.500)	(0.248)
Age (in months)	104.9	104.2
	(7.135)	(5.640)
Spanish speaker	0.714	0.181
	(0.452)	(0.385)
Number of unique students	40,418	392,121

Notes: Standard deviations are given in parentheses. Reading and math scores are standardized at the grade-year level to zero mean and unit variance.

Table 2. Characteristics of First-Time Third Grade English Learners Between 2002–03 and 2005–06 School Years, by Grade Level in the Following Year

	Repeating Third Grade	In Fourth Grade
Third-grade reading score	-1.708	-0.611
	(0.704)	(0.953)
Third-grade math score	-1.533	-0.564
	(0.843)	(0.961)
Years in English learner program	2.764	2.171
	(1.152)	(1.455)
Eligible for subsidized meals	0.891	0.790
	(0.311)	(0.407)
White	0.0391	0.0666
	(0.194)	(0.249)
Black	0.216	0.151
	(0.412)	(0.358)
Hispanic	0.716	0.729
	(0.451)	(0.444)
Male	0.581	0.512
	(0.493)	(0.500)
Special education	0.226	0.109
	(0.418)	(0.311)
Foreign born	0.357	0.512
	(0.479)	(0.500)
Age (in months)	104.7	104.9
	(6.554)	(7.186)
Spanish speaker	0.704	0.715
	(0.457)	(0.451)
Number of unique students	7,141	31,413

Notes: Standard deviations are given in parentheses. Reading and math scores are standardized at the grade-year level to zero mean and unit variance.

Table 3. Baseline Equivalency of Student Characteristics Around the Retention Cutoff

	Bandwidth: 10 points	Bandwidth: 20 points
Third-grade math score	0.011	0.037
	(0.033)	(0.025)
Years in English learner program	-0.063	-0.027
	(0.072)	(0.049)
Eligible for subsidized meals	-0.017	0.006
	(0.018)	(0.012)
White	-0.003	0.005
	(0.011)	(0.011)
Black	0.024*	0.001
	(0.014)	(0.012)
Hispanic	-0.025	-0.004
	(0.019)	(0.017)
Male	0.017	-0.013
	(0.025)	(0.022)
Special education	0.002	-0.019
	(0.013)	(0.011)
Foreign born	0.023	0.012
	(0.020)	(0.016)
Age (in months)	-0.896*	-0.409
	(0.491)	(0.333)
Number of unique students	5,002	9,921

Notes: Robust standard errors, two-way clustered at the school and relative reading score level, are given in parentheses. The estimates represent the discontinuities in student characteristics at the retention cutoff, obtained parametrically using linear polynomial specification and a bandwidth of 10 and 20 points. *, **, and *** represent statistical significance at 10, 5, and 1 percent, respectively.

Table 4. Effects of Grade Retention on Reading Test Scores for English Learners

	Same-Age Analysis	
	Bandwidth: 10 points	Bandwidth: 20 points
1 year later (SD of Y = 352)	92.312***	103.732***
	(34.522)	(26.548)
2 years later (SD of Y = 316)	183.632***	182.531***
	(37.446)	(27.566)
3 years later (SD of Y = 311)	107.230***	122.477***
	(35.980)	(25.068)
4 years later (SD of Y = 302)	24.238	52.990***
	(26.087)	(20.203)
5 years later (SD of Y = 253)	13.492	27.263
	(35.136)	(23.608)
	Same-Grade Analysis	
	BW: 10 points	BW: 20 points
Fourth grade (SD of Y = 318)	263.910***	286.082***
	(38.338)	(27.750)
Fifth grade (SD of Y = 302)	222.664***	224.965***
	(38.846)	(27.843)
Sixth grade (SD of Y = 305)	158.277***	195.954***
	(35.090)	(25.268)
Seventh grade (SD of Y = 274)	125.854***	153.267***
	(27.825)	(18.721)
Eighth grade (SD of Y = 221)	111.142***	111.529***
	(34.578)	(24.250)

Notes: All regressions control for the baseline student characteristics listed in Table 3, and standard errors that are two-way clustered at the school and relative reading score level are given in parentheses. *, **, and *** represent statistical significance at 10, 5, and 1 percent, respectively.

Table 5. Effects of Grade Retention on Time to Proficiency for English Learners

	Bandwidth: 10 points	Bandwidth: 20 points
Years to proficiency	-1.077***	-0.795***
	(0.321)	(0.234)
Control mean at cutoff	2.08	
	Same-Age Analysis	
Classified as English learner…		
1 year later	-0.209**	-0.074
	(0.089)	(0.070)
Control mean at cutoff	0.738	
2 years later	-0.184**	-0.152**
	(0.087)	(0.062)
Control mean at cutoff	0.479	
3 years later	-0.220***	-0.208***
	(0.072)	(0.058)
Control mean at cutoff	0.327	
4 years later	-0.162***	-0.141***
	(0.062)	(0.048)
Control mean at cutoff	0.225	
5 years later	-0.077	-0.072*
	(0.058)	(0.041)
Control mean at cutoff	0.155	
	Same-Grade Analysis	
Classified as English learner in…		
Fourth grade	-0.482***	-0.375***
	(0.085)	(0.066)
Control mean at cutoff	0.733	
Fifth grade	-0.320***	-0.319***
	(0.084)	(0.061)
Control mean at cutoff	0.470	
Sixth grade	-0.341***	-0.271***
	(0.077)	(0.057)
Control mean at cutoff	0.319	
Seventh grade	-0.187***	-0.159***
	(0.062)	(0.050)
Control mean at cutoff	0.220	
Eighth grade	-0.091	-0.072*
	(0.064)	(0.044)
Control mean at cutoff	0.142	

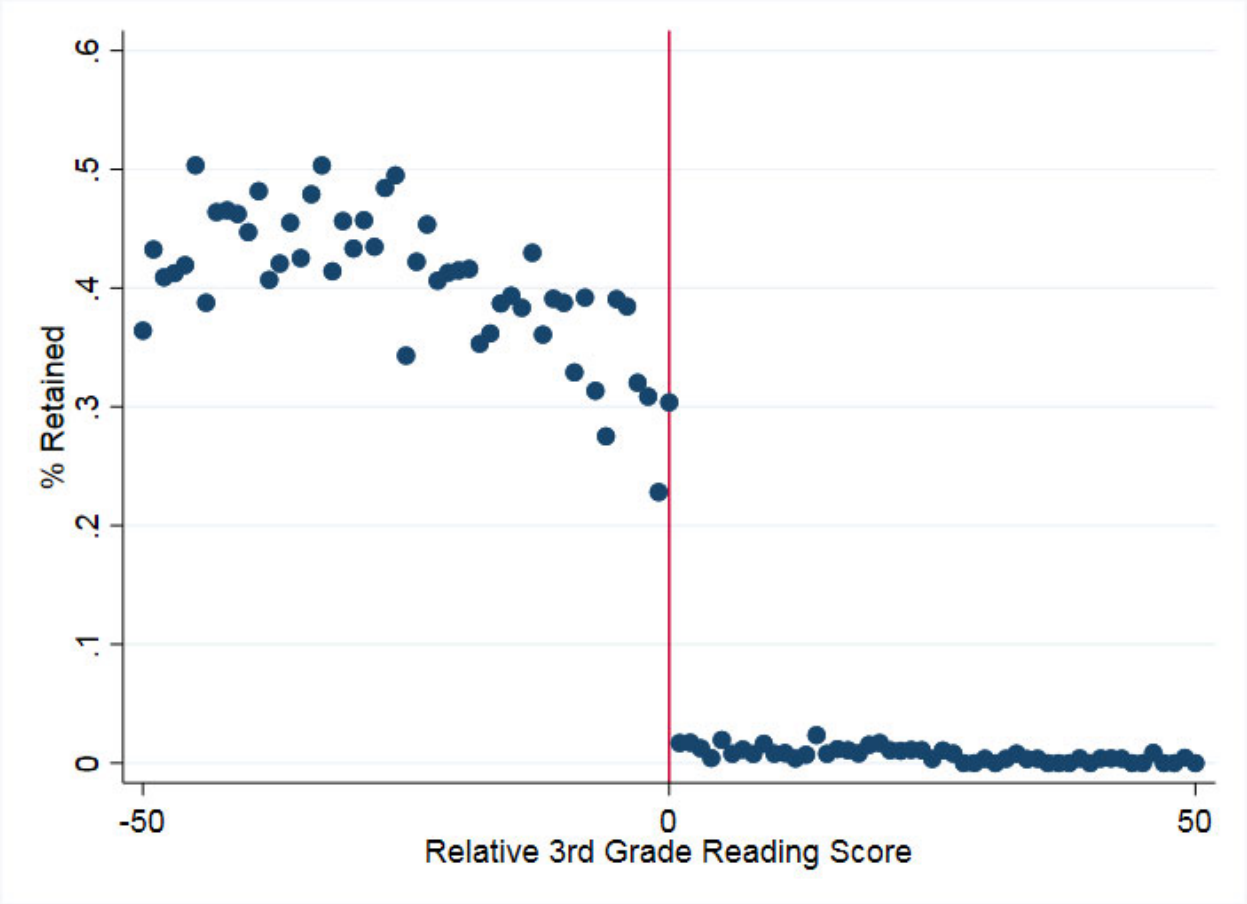
Notes: All regressions control for the baseline student characteristics listed in Table 3, and standard errors that are two-way clustered at the school and relative reading score level are given in parentheses. *, **, and *** represent statistical significance at 10, 5, and 1 percent, respectively.

Table 6. Effects of Grade Retention on Middle School and High School Course-Taking Among English learners

	Bandwidth: 10 points	Bandwidth: 20 points
Remedial language arts course in MS	-0.145**	-0.247***
	(0.073)	(0.054)
Control mean at cutoff	0.721	
Advanced language arts course in MS	0.214**	0.174***
	(0.106)	(0.068)
Control mean at cutoff	0.308	
Advanced math course in MS	0.349***	0.250***
	(0.103)	(0.070)
Control mean at cutoff	0.299	
Advanced science course in MS	0.258***	0.200***
	(0.097)	(0.062)
Control mean at cutoff	0.294	
College-credit bearing course by 12th grade	0.438***	0.308***
	(0.144)	(0.094)
Control mean at cutoff	0.153	

Notes: All regressions control for the baseline student characteristics listed in Table 3, and standard errors that are two-way clustered at the school level and the relative reading score level are given in parentheses. *, **, and *** represent statistical significance at 10, 5, and 1 percent, respectively.

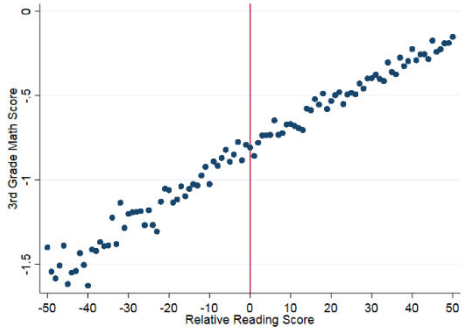
Figure 1. Grade Retention and Third-Grade Reading Score Among English Learners



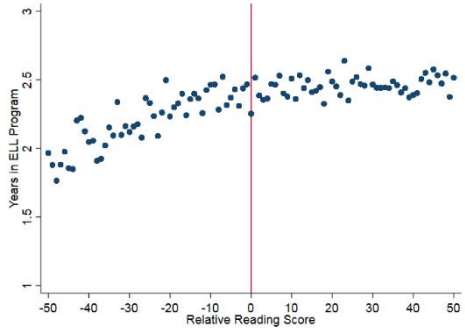
Notes: The figure presents the raw cell means of the retention indicator for each reading score between 50 points below and 50 points above the retention cutoff.

Figure 2. The Relationship Between Reading Scores in Third Grade and English Learner Student Characteristics

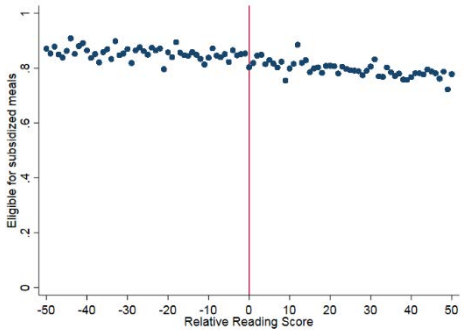
Third-Grade Math Score



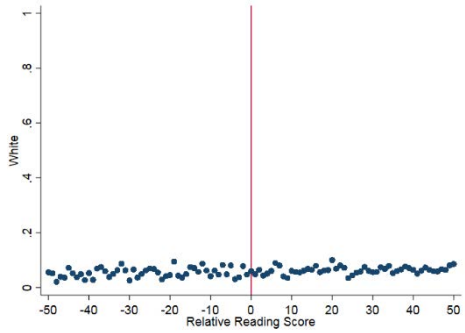
Years in EL Program



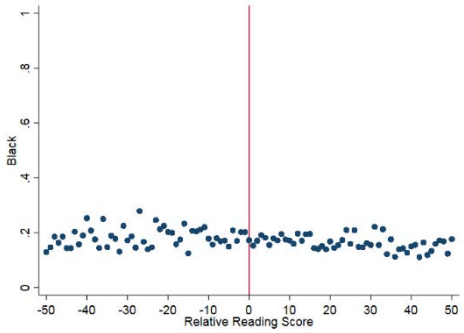
Eligible for Subsidized Meals



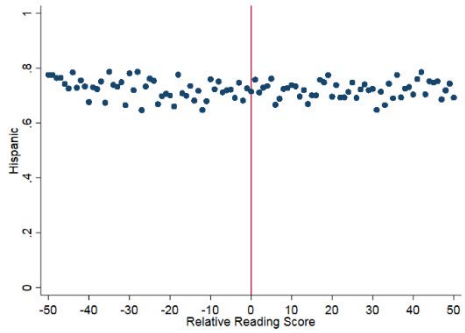
White



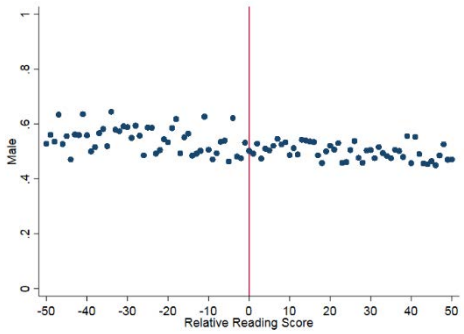
Black



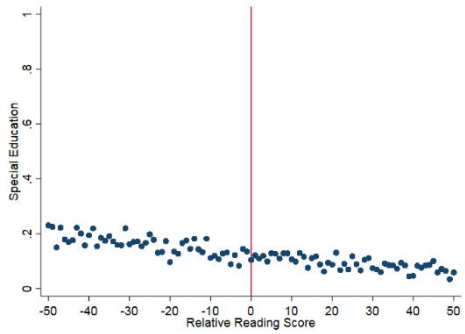
Hispanic



Male

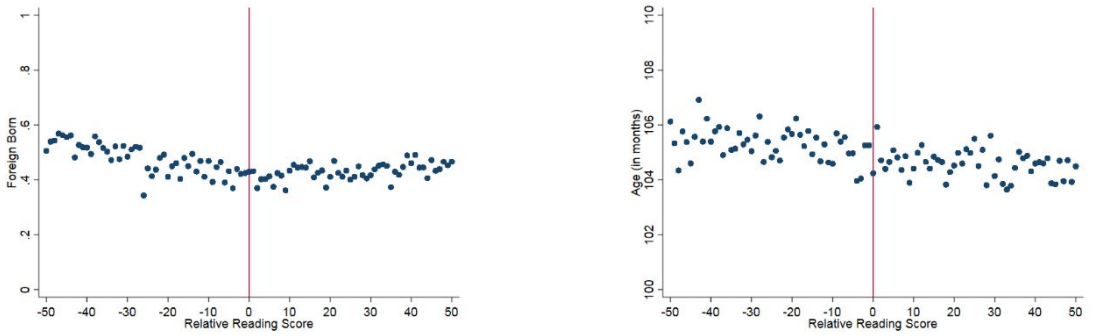


Special Education



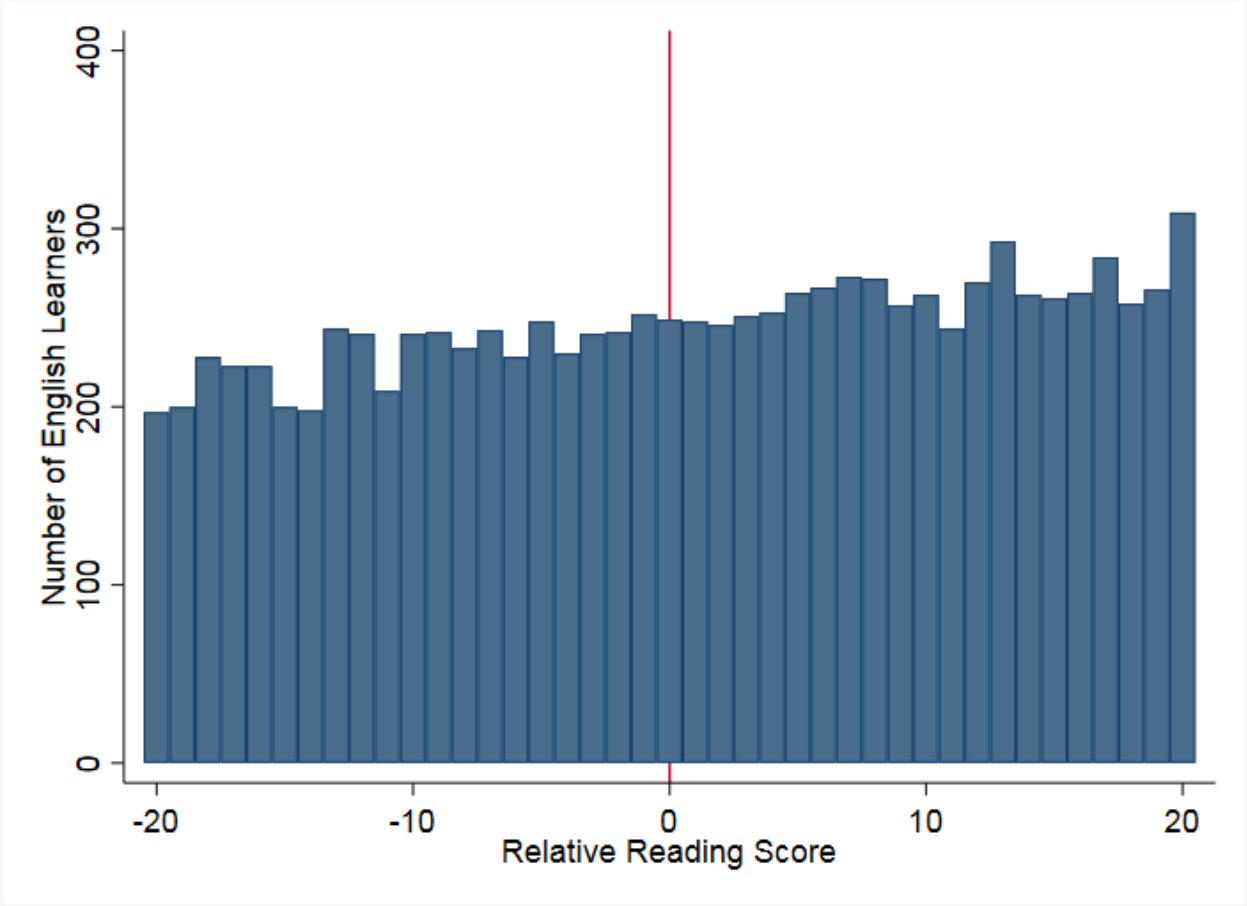
Foreign Born

Age (in Months)



Notes: The figure presents the raw cell means of the given student attribute for each reading score between 50 points below and 50 points above the retention cutoff.

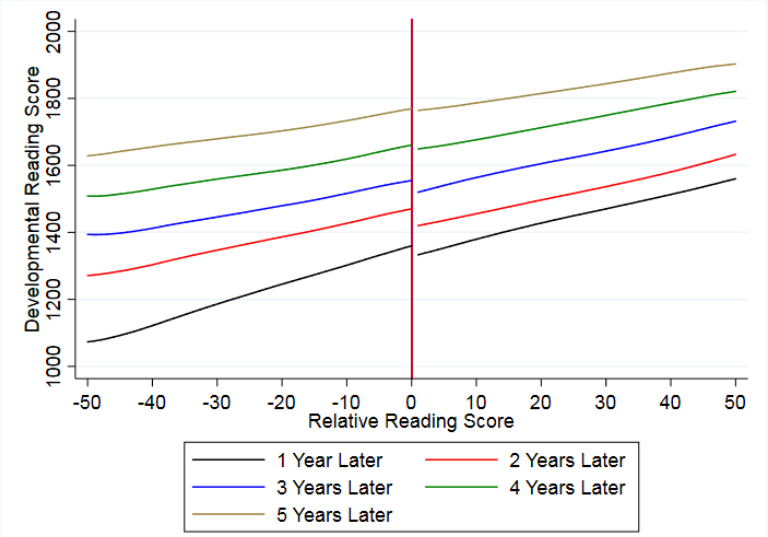
Figure 3. Distribution of Third-Grade Reading Scores, English Learners



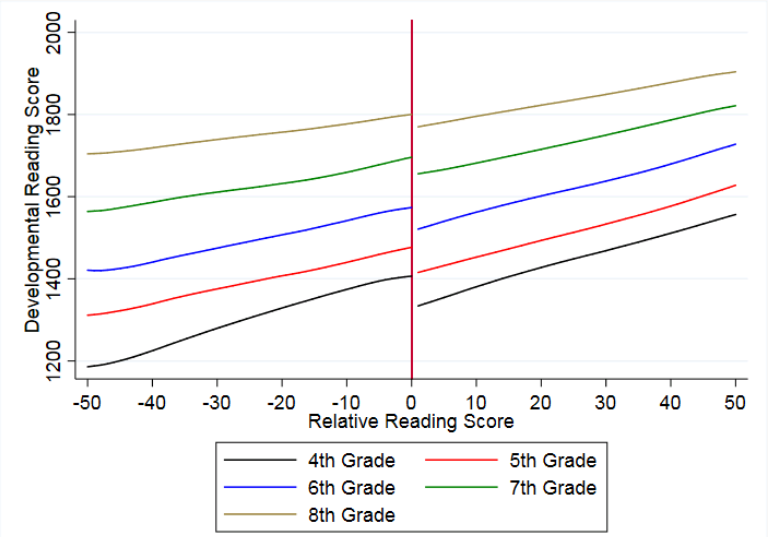
Notes: The figure presents the number of students in each reading score bin between 20 points below and above the retention cutoff, which is shown by the vertical line.

Figure 4. Retention and Reading Achievement Among English Learners

(I) Same-Age Analysis



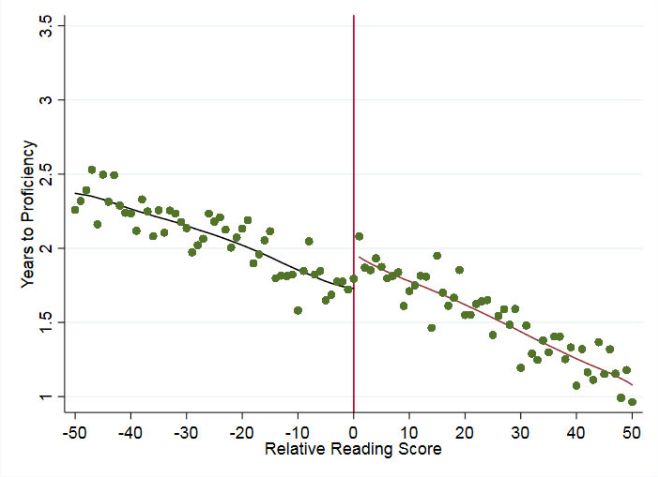
(II) Same-Grade Analysis



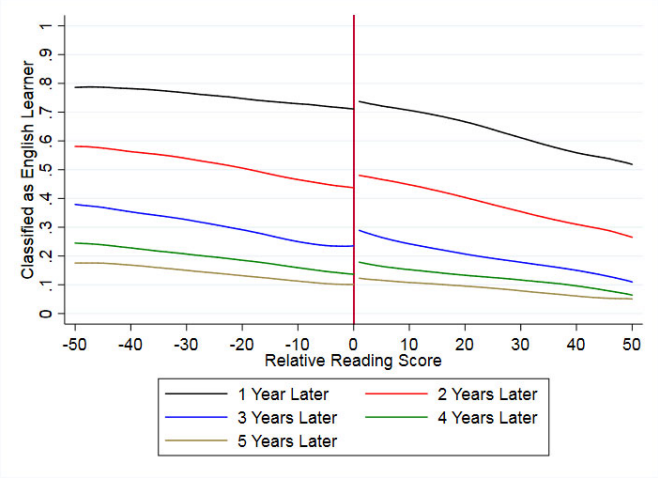
Notes: The figures present the local linear smoothing of student developmental scores in reading in the years following third grade, and by grade level on relative reading score of the student separately for the left of the cutoff date and the right. The triangle kernel and a bandwidth of 20 points are used in the estimation.

Figure 5. Retention and Time to Proficiency

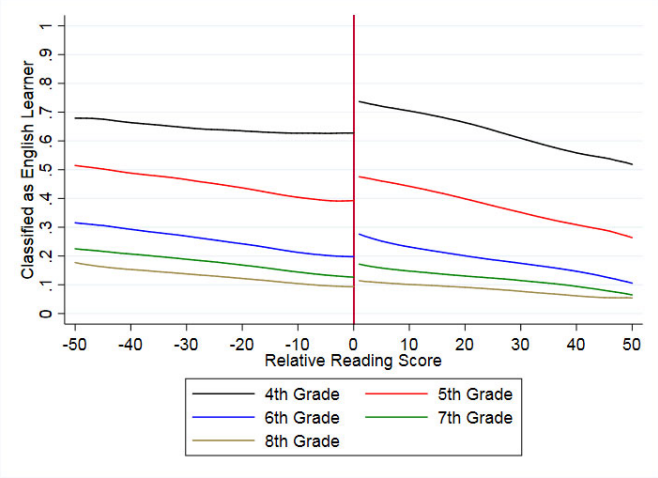
(A) Years to Proficiency



(B) English Learner Classification by Year



(C) English Learner Classification by Grade

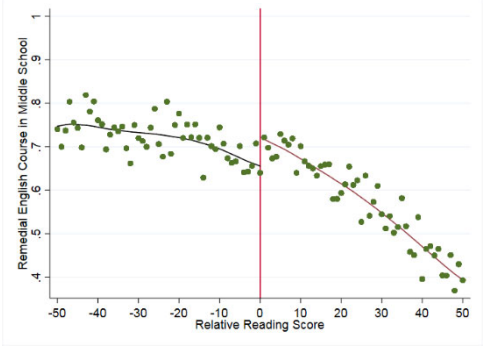


Notes: The three panels present the local linear smoothing of years to proficiency in panel (A), English learner classification in the years following the first time students enter third grade in panel (B), and English learner classification by grade in panel (C), on relative reading score of the student separately for the left of the cutoff date

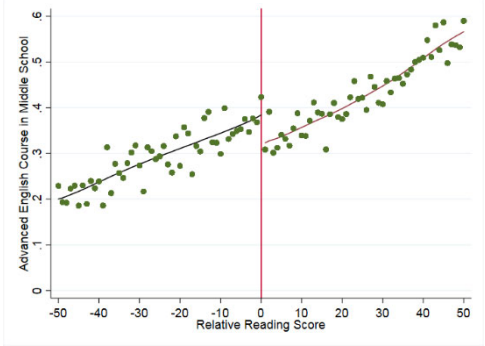
and the right. The triangle kernel and a bandwidth of 20 points are used in the estimation. The solid circles represent raw cell means in panel (A).

Figure 6. Retention and Course-Taking in Middle and High School

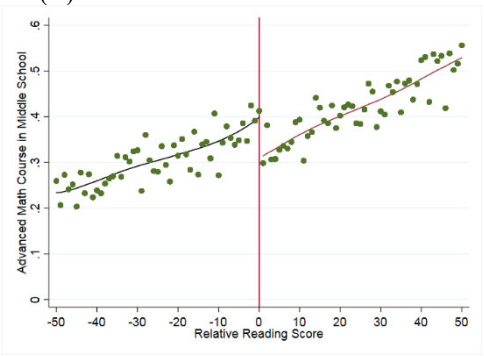
(A) Remedial Language Arts in Middle School



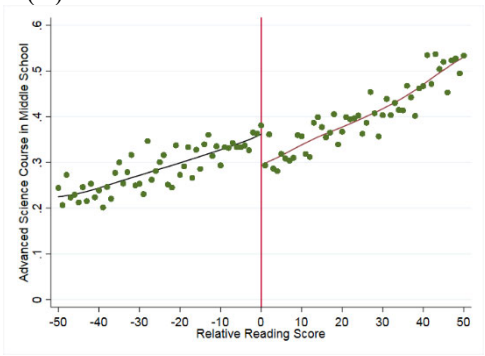
(B) Advanced Language Arts in Middle School



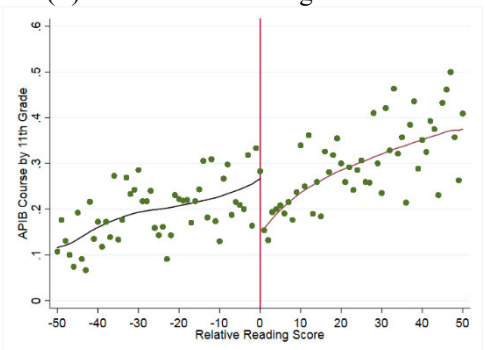
(C) Advanced Math in Middle School



(D) Advanced Science in Middle School



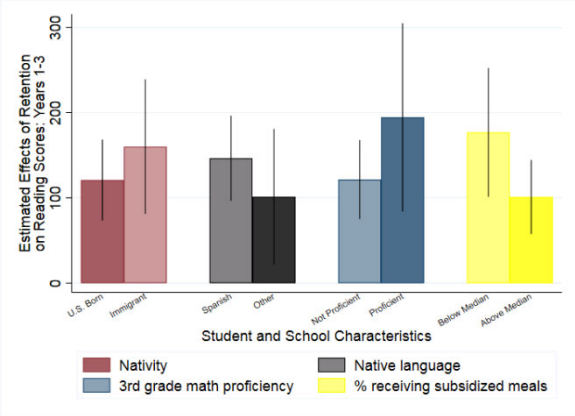
(E) AP-IB Course in High School



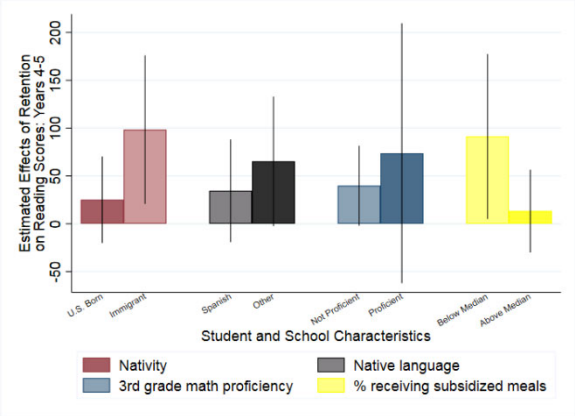
Notes: The figures present the local linear smoothing of student course-taking indicators on relative reading score of the student separately for the left of the cutoff date and the right. The triangle kernel and a bandwidth of 20 points are used in the estimation. The solid circles represent raw cell means.

Figure 7. Effects of Grade Retention on English Learners, by Subgroup

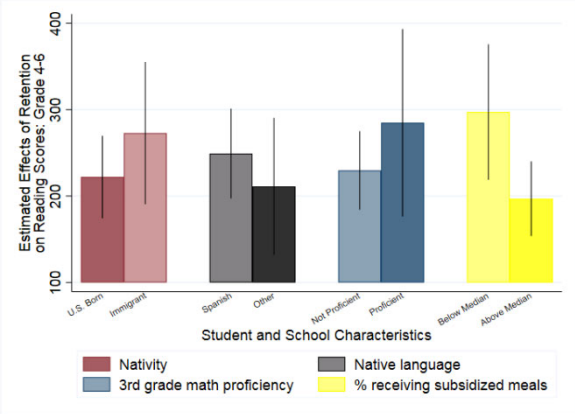
(A) Reading scores: Years 1–3



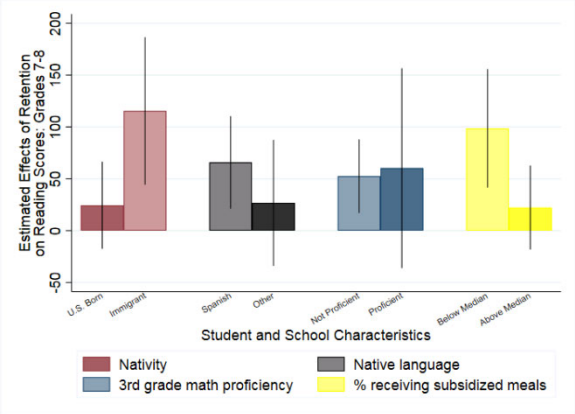
(B) Reading scores: Years 4–5



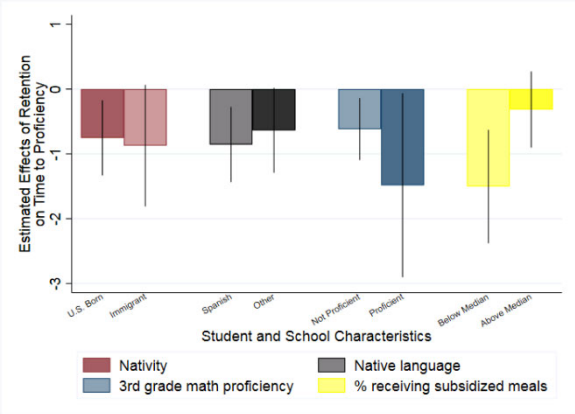
(C) Reading scores: Grades 4–6



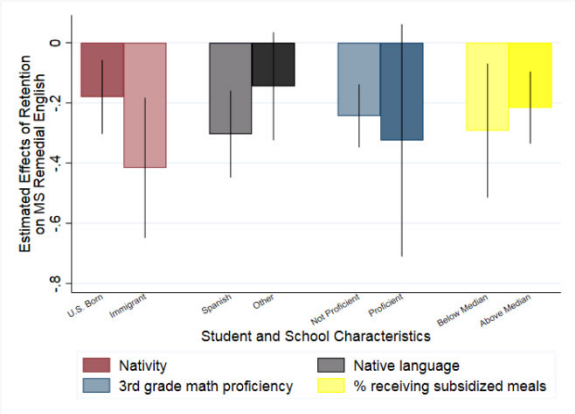
(D) Reading scores: Grades 7–8



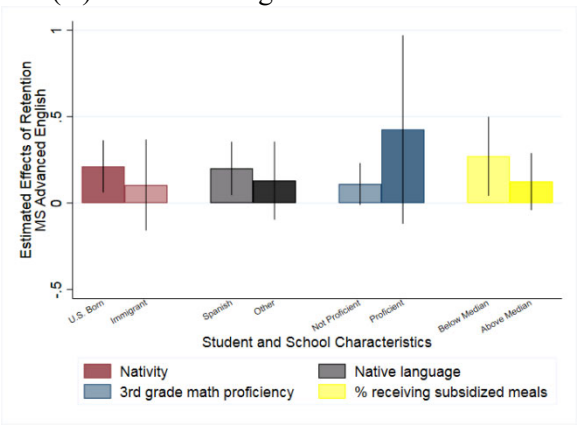
(E) Time to Proficiency



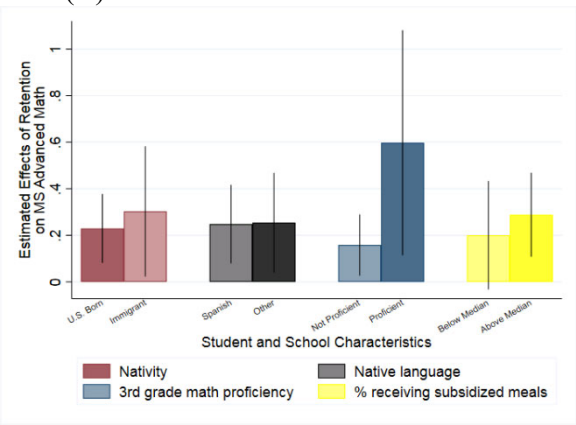
(F) Remedial English in Middle School



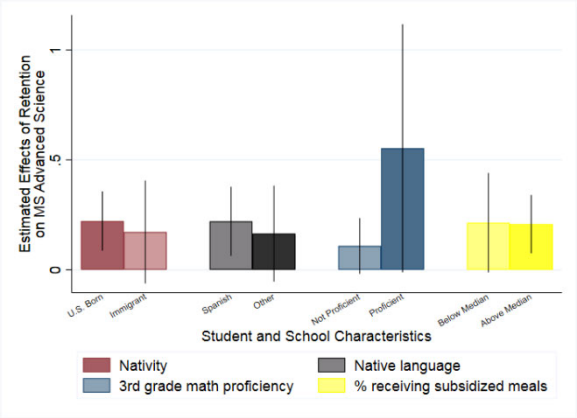
(G) Advanced English in Middle School



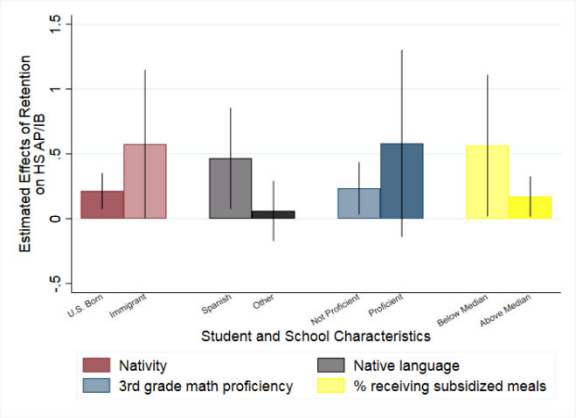
(H) Advanced Math in Middle School



(I) Advanced Science in Middle School



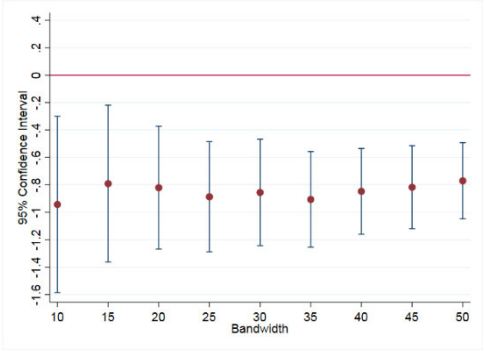
(J) AP/IB in High School



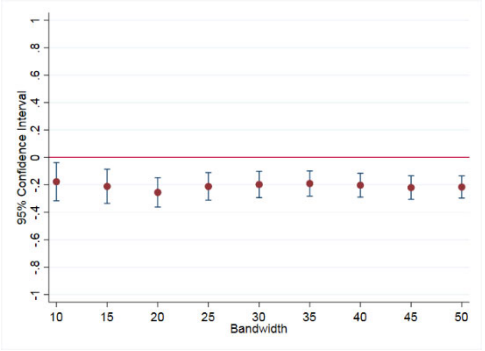
Notes: Bars in each figure represent the treatment effect (τ) on the corresponding outcome for the given subgroup of English learner students with a bandwidth of 20 points around the cutoff. Spikes in each figure provide the 95% confidence interval for the corresponding estimate. All regressions control for the baseline student characteristics listed in Table 3, and standard errors are two-way clustered at the school and relative reading score levels.

Figure 8. Robustness to Bandwidth Selection

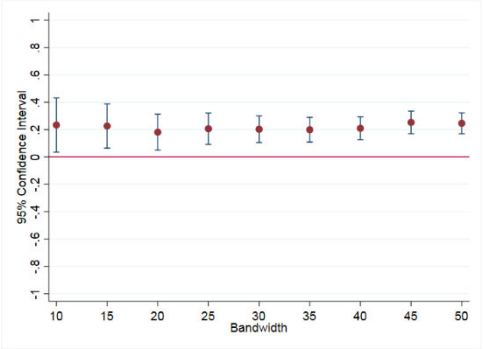
(A) Time to Proficiency



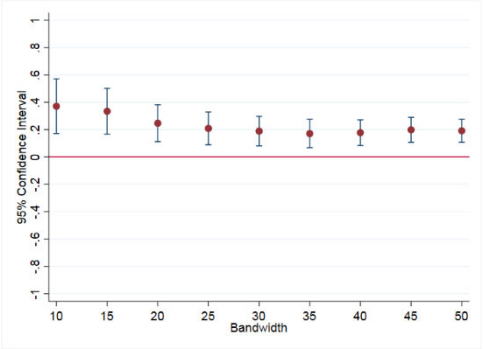
(B) Remedial Language Arts in Middle School



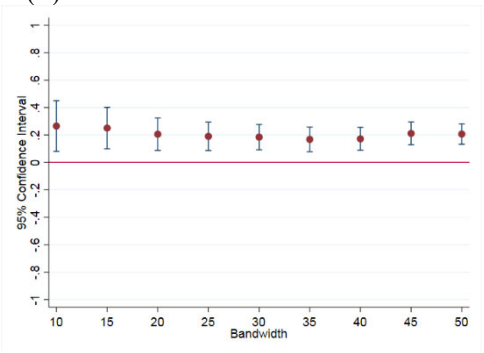
(C) Advanced Language Arts in Middle School



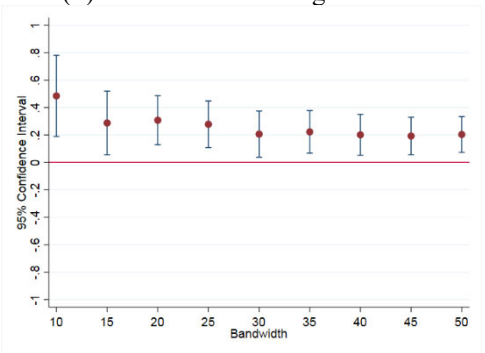
(D) Advanced Math in Middle School



(E) Advanced Science in Middle School



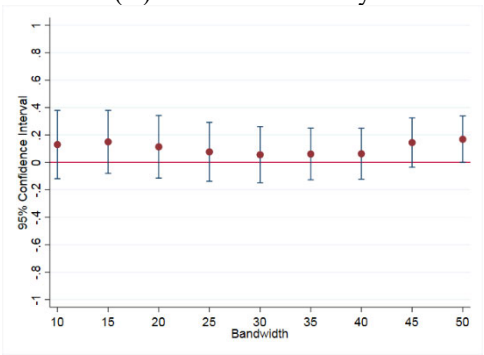
(F) AP/IB Course in High School



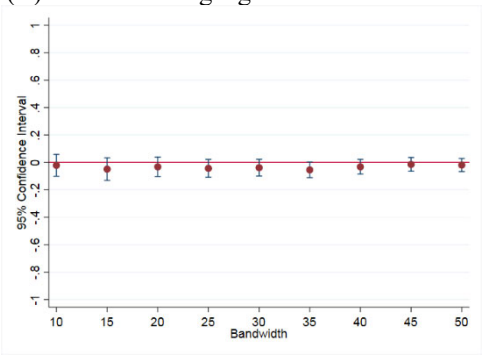
Notes: The figures present the treatment effect (τ) and the 95% confidence interval estimated using the bandwidth shown and linear specification.

Figure 9. Falsification Test—Pseudo-Effects in Pre-Policy Years

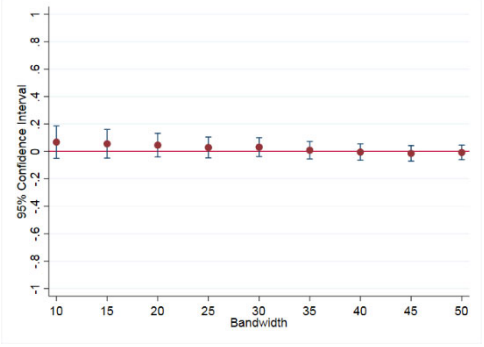
(A) Time to Proficiency



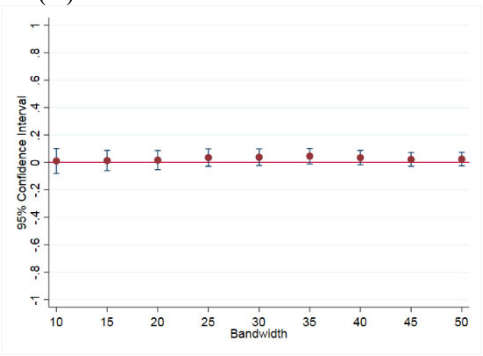
(B) Remedial Language Arts in Middle School



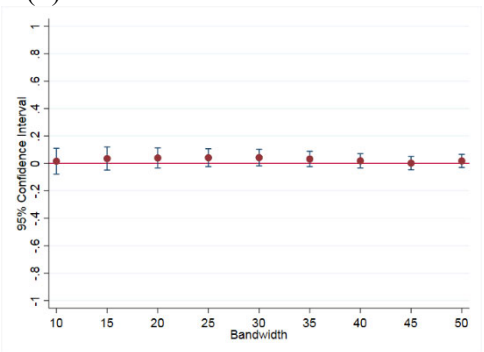
(C) Advanced Language Arts in Middle School



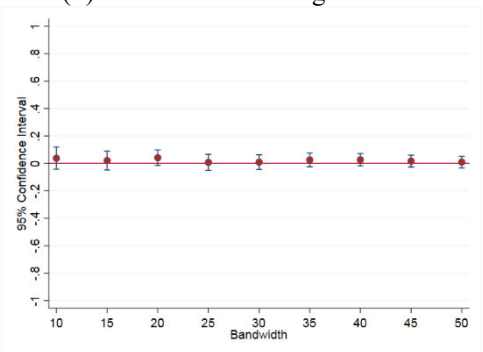
(D) Advanced Math in Middle School



(E) Advanced Science in Middle School

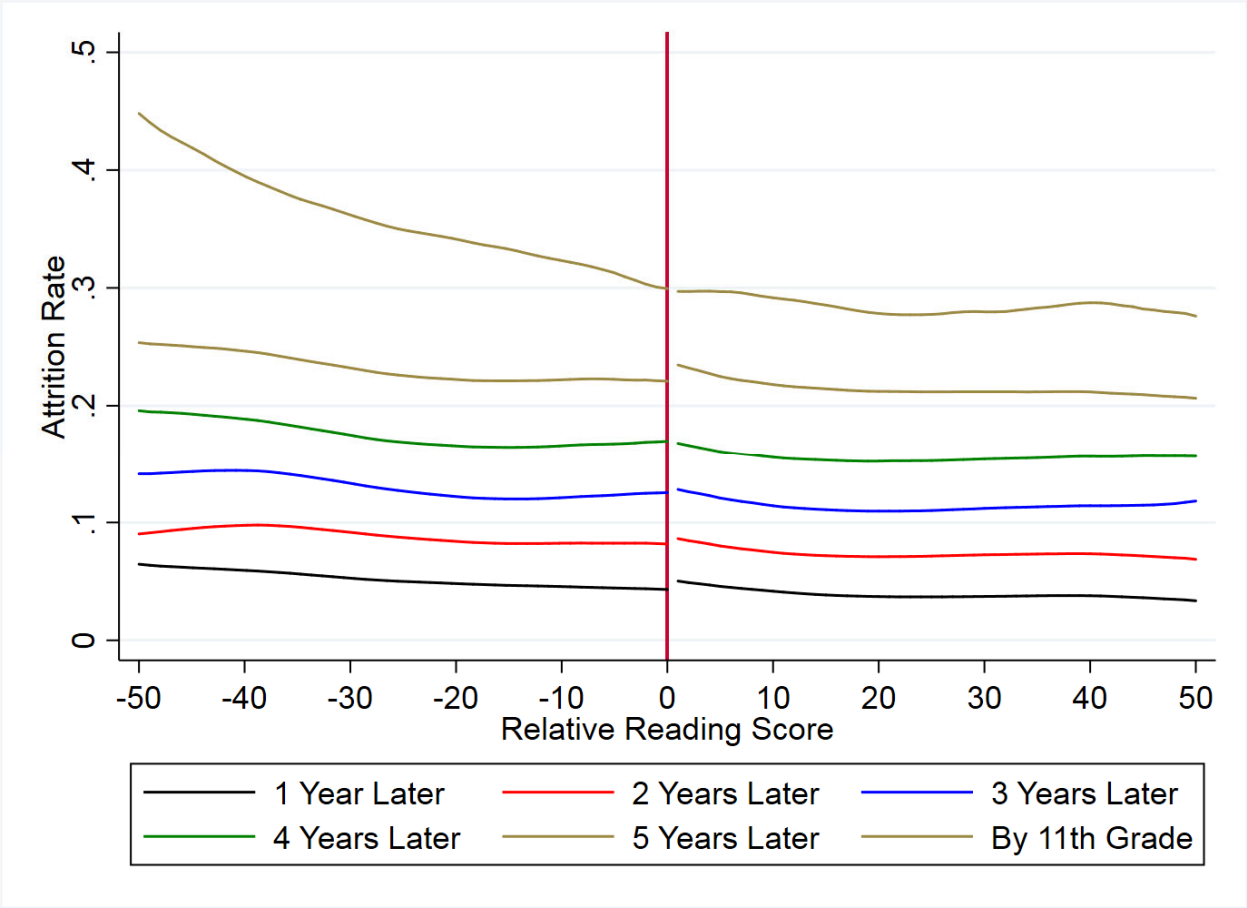


(F) AP/IB Course in High School



Notes: The figures present the pseudo-treatment effect and the 95% confidence interval estimated using the bandwidth shown and linear specification in the two years before the policy took effect.

Figure 10. Sample Attrition Around the Retention Cutoff



Notes: The figures present the local linear smoothing of the attrition rate in the years following the third grade on relative reading score of the student separately for the left of the cutoff date and the right.