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

David Figlio

Umut Özek

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

David Figlio
Northwestern University and NBER/CALDER

Umut Özek
American Institutes for Research/CALDER

Contents

Acknowledgments

Abstract

In this study, we use microdata from 12 Florida county-level school districts to examine the effects of early grade retention on the short-, medium-, and long-term outcomes of English learners in a regression discontinuity design. We find that retention in the third grade coupled with instructional support 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 triples the likelihood of taking college credit-bearing courses in high school for English learners. We do not find any adverse effects of the policy on disciplinary problems or absences among English learners.

1. Introduction

The United States is in the midst of the second largest wave of immigration in its history, and the share of English learners—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— in public schools has grown significantly over the last two decades, especially in urban school districts. It is well documented that English learners face significant achievement gaps: According to the 2019 wave of the National Assessment of Educational Progress (NAEP), only 16 percent of English learners in fourth grade perform at or above the proficient level in mathematics nationwide, compared to 45 percent of the non-English learners. Therefore, how to better serve English learners in public schools and how different interventions affect these students are vital education policy questions.

There is emerging literature on the effects of education policies and programs targeting English learners[1] (e.g., different modes of English learner instruction, classroom structure), yet relatively little is known about the effects of broader education policies on English learners. In this study, we address this question in the context of an increasingly popular intervention targeting low-performing students – early grade retention policies that require students to be retained and receive instructional support if they score below a predetermined threshold on standardized reading tests in third grade. As of 2018, about half of U.S. states and the District of Columbia require or encourage school districts to retain third grade students who lag behind in reading (ECS, 2018).

Examining the effects of these policies on English learners is particularly important for several reasons. First, while early grade retention policies do not specifically target English

---

[1] See, for example, Umansky and Reardon (2014); Valentino and Reardon (2015); and Steele et al. (2017).

learners, these students are disproportionately affected. For example, under Florida's third grade

retention policy, on which we focus in this study, 45 percent of English learners score below the

retention cutoff on the third grade reading test compared to 17 percent of non-English learners.[2]

Second, an overarching conclusion of the existing literature on English learners 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 these students (Figlio

and Ozek forthcoming; Halle et al. 2012). Early grade retention policies could help English

learners accomplish this goal not only by providing additional time for these students to acquire

the necessary English skills but also by providing instructional support in reading. Additional

time could be particularly beneficial for English learners who recently relocated to the U.S. and

might simply need more time to learn English, in contrast to low-performing non-English

learners whose academic struggles might be driven by factors beyond English deficiencies.

In particular, we investigate the causal effects of Florida's early grade retention policy 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 passing

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. Grade retention policies have many

potential downsides as well, and we also therefore examine the possible adverse effects of

---

[2] These figures are based on authors' calculations using student-level administrative data from twelve anonymous
school districts from Florida for school years between 2002–03 and 2005–06.

Florida's policy on student disengagement from schooling as proxied by disciplinary incidents and absences.

There exists extensive literature examining the effects of grade retention on student outcomes yielding mixed results, with several studies examining Florida's third grade retention policy. For instance, Schwerdt et al. (2017) find that although retained students outperform their same-age peers in the short term (one to three years), these achievement gains fade out entirely after 6 years in Florida. 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, results from other states and districts are less positive[3]: In Chicago, for example, Jacob and Lefgren (2009) finds that grade retention increases high school dropout for eighth graders, but not for sixth graders. Similarly, Eren et al. (2017) and Larsen and Valant (2018) show that grade retention increases high school dropout for both fourth and eighth graders. In New York City, Mariano et al. (2018) finds that grade retention reduces high school credit accumulation and increases high school dropout for middle school students.

To the best of our knowledge, there is no study to date that takes a thorough look at the causal effects of test-based early grade retention policies on English learners.[4] A deep dive into

---

[3] The discrepancy between the findings in Florida and other states could be driven by differences in instructional support provided to the retained students and (2) differences in grade of retention.

[4] Schwerdt et al. (2017) do present one analysis of the English learner subgroup in their online appendix, but they do not substantially investigate whether and why English learners benefit from the program. In particular, in Table A-10 of their online appendix, they examine the effects of being retained for different student subgroups (including English learners) on reading and math scores 1-3 and 4-6 years after third grade, retention in the years following third grade, high school graduation, and college enrollment. They find similar effects for English learners compared to the general population. We build on this earlier work by focusing on English learners, examining the effects of the policy on other important academic outcomes including time to proficiency and course-taking in middle and high school, investigating heterogeneous treatment effects among English learners, and exploring the mechanisms (e.g., retention versus summer school) through which the policy might be affecting student outcomes.

the effects of the policy for English learners is necessary because there is reason to believe that, especially for English learners, there might be substantial heterogeneity in the effects of the policy depending on background differences[5]; the unique data that we bring to bear to this question allow us to investigate the heterogeneous impacts of the policy on different groups of English learners. Further, unlike previous studies on Florida's retention policy, we observe several components of instructional support under the policy, including summer reading program enrollment and reading instruction time in the following year, which allows us to better disentangle the effects of grade retention from the effects of instructional support on English learners.

We find significant benefits of Florida's early grade retention policy on English learners. In particular, we find that English learners whose third grade reading scores fell right below the retention cutoff (and hence were subject to the policy) significantly outperform their peers on the other side of the cutoff on reading tests, attain English proficiency sooner, are less likely to require English remediation in middle school, and are more likely to take advanced courses in middle and high school. These benefits are larger in many cases for foreign born English learners, and are in line with the effect sizes estimated for non-English learners with the exception of college credit-bearing courses in high school whereby we observe large and statistically significant benefits of the retention policy for English learners in stark contrast to the small and insignificant effects on non-English learners. This latter finding could be driven by recent immigrant optimism - differences in expectations for the future between non-English learners and English learners most of who are recent immigrants. In other words, if immigrant

---

[5] For example, a number of studies have shown that it takes Spanish-speaking English learners and English learners from disadvantaged backgrounds much longer to attain proficiency compared to other English learners groups (Conger, 2010; Grissom, 2004; Hakuta, Butler, & Witt, 2000; Parrish et al., 2006; Salazar, 2007; Slama, 2014; Thompson, 2012).

optimism and motivation fade across generations as evidenced in the literature (e.g., Kao and Tienda 1995; Figlio and Ozek, forthcoming), a similar improvement in human capital in elementary and middle school (driven by the retention policy) could be leading to a much larger change in high school course-taking behavior among English learners who have higher aspirations to attend college.[6]

An important concern with grade retention policies is the potential adverse effects on student disengagement from schooling. Indeed, similar to the overall effects presented in Ozek (2015), we find that the grade retention policy leads to increases in the likelihood of being involved in severe disciplinary incidents (that lead to suspensions) among English learners in the first year. However, unlike non-English learners, the findings suggest that English learners whose third grade reading scores fell right below the retention cutoff were significantly less likely to have disciplinary problems in middle school compared to English learners on the other side of the cutoff, especially among foreign-born English learners, although these results are less robust to different bandwidth choices. We did not find any significant adverse effects of the policy on student absences.

Finally, we examine the potential mechanisms behind the observed benefits of the retention policy. We find significant effects of failing the third grade reading test on the likelihood of retention (equal to 30 percentage points) and much larger effects on summer school enrollment (roughly 55 percentage points) at the retention cutoff. This indicates that failing the test could improve student outcomes through mechanisms other than its effect on retained students if summer school alone has a significant positive effect on English learners who were

---

[6] According to the 2015 wave of the Programme for International Student Assessment (PISA), students with an immigrant background express greater motivation to achieve, and are significantly more likely to expect to complete tertiary education (OECD, 2018).

exempt from retention, yet still eligible for summer school under Florida's policy if they failed the third grade reading test. We find no significant effect of attending summer school on exempt English learners, providing evidence that the observed benefits are primarily driven by the effect of the policy on retained students. With this finding in mind, we then estimate the effects of grade retention coupled with instructional support using a fuzzy regression discontinuity design. We find that retention reduces the time to proficiency by about a year, decreases the likelihood of taking a remedial English course in middle school by one-third, almost doubles the likelihood of taking an advanced course in math and science in middle school, and triples the likelihood of taking college credit-bearing courses in high school.

While our findings suggest that early grade retention could be an effective policy lever to improve the outcomes of English learners, it is important to note that these results capture the combined effect of grade retention and the instructional support provided to retained students under Florida's retention policy, and hence might not be generalizable to other retention policies that lack such support systems. That said, Florida's experience could be valuable for other jurisdictions to improve the well-being of recent immigrants as they face increasing financial burden with the recent influx of English learners into public schools.[7] Our findings suggest that 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 strategy that incorporates instructional support might be worth it if it leads to accelerated English language proficiency and heightened human capital accumulation, especially if policies aimed at counteracting increased dropout risk are introduced as well.

---

[7] 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). Therefore, reducing the time to proficiency even by one year translates into significant cost savings for many states and school districts.

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." Similar to many other retention policies examined in the literature (e.g., Chicago Public Schools, New York City, Louisiana), students who score below this cutoff are eligible to participate in summer school at the end of the year.[8] An important difference between Florida's policy and many other test-based retention policies, however, is that the legislation requires that schools provide substantial instructional support for retained students in the following school year. These include instruction using proven-effective teaching strategies, assigning retained students to high-performing teachers, and at least 90 minutes of reading instruction each day.[9] 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.

---

[8] Under Florida's policy, schools and districts are required to create a reading camp schedule that facilitates intensive reading intervention lasting between six to eight weeks, four days per week, and six hours per day. For these sessions, schools and districts are encouraged to (1) choose qualified teachers and reading coaches with reading certification or endorsement and reading coaches and (2) provide reading instruction utilizing a research based sequence of reading instruction and small group differentiated instruction in order to meet individual student needs.

[9] In particular, schools are required to develop academic improvement plans for retained students that specifically address their needs, to assign these students to high-performing teachers (based on student performance and performance appraisals), and to provide an additional ninety minutes of daily reading instruction.

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 "longer term" English learners who have not been able to attain proficiency in at least two years. Students are also eligible for a good-cause exemption if they have certain disabilities and have been already retained once until third grade, 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 statewide reading test by performing at an acceptable level on an alternative standardized reading assessment approved by the State Board of Education (e.g., scoring at or above the 51st percentile on the third grade FCAT-NRT reading examination during the time frame we examine), or by demonstrating proficiency through a teacher-developed portfolio.

While some low-performing students are exempt from retention under Florida's policy, all students who score in the lowest achievement level on the third grade reading test are eligible to enroll in a non-mandatory summer school.[10] In the cohorts we examine in our study (first-time third graders in between 2002 and 2005 in the twelve anonymous districts), we observe that 48 percent of low-performing third graders who were exempt from retention enrolled in summer school (44 percent among English learners). Individual districts have flexibility over different aspects of this summer program, but Florida Department of Education recommends six to eight

---

[10] For example, see the memorandum from the Commissioner of Education to district superintendents in April 2013: https://info.fldoe.org/docushare/dsweb/Get/Document-1078/Summer%20Reading%20Instruction.pdf, accessed on March 7, 2019.

weeks of instruction, 4 days per week, and 5.5 hours of reading instruction per day including two 90-minute blocks of reading instruction, which corresponds to essentially two days of reading instruction during the regular school year in one day. We further discuss the implications of this summer school component for our analysis in Section 6.

3. Data

In our analysis, we make use of detailed longitudinal, student-level administrative data that cover the school years between 2000–01 and 2011–12 from twelve anonymous, county-level school districts in Florida. In particular, the student-level administrative data contain Florida Comprehensive Assessment Test (FCAT) reading and math scores for all students between grades three and ten in these districts, information about disciplinary problems and absences during that time frame, and 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, and birth records for all children born in Florida between 1992 and 2002 and who subsequently attended a public school in the state. These birth records contain maternal characteristics such as educational attainment, marital status, and age at birth, which we use to examine the differential effects of the policy on different student groups.

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,

9

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 two years when they first entered the third grade and hence were not eligible for the good-cause exemption. The first two columns of Table 1 compare 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 are 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 the grade retention policy on student outcomes is that the students who are affected by the policy are not selected randomly. Therefore, regression-adjusted differences based on observable student attributes between treated students and others are likely to yield biased inferences. The last three columns of Table 1 illustrate these differences between English learner students who were retained, English learners who enrolled in summer school but were not retained, and English learner students who neither were retained nor enrolled in summer school, and shows substantial differences between the three student groups. For example, compared to other students, retained English learners 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

10

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

To estimate the causal effect of the grade retention policy on English learners, we rely on an RD design using the student-level treatment cutoffs based on the third grade reading scores of students. In our main analysis, we are interested in the overall effect of the policy on English learners, and we discuss different treatments and their effects in Section 6. Let S_i denote the difference between the FCAT scale score of student i on the third grade reading test and the retention cutoff—with negative values indicating scores below cutoff—and B_i denote an indicator for students below the cutoff. In this setting, the effect of failing the third grade test on student outcomes is given by:

$$\beta = \lim_{S_i \uparrow 0} E[Y_i|S_i] - \lim_{S_i \downarrow 0} E[Y_i|S_i]. \tag{1}$$

where Y_i is the outcome of interest (e.g., test scores, years to proficiency after third grade, advanced or remedial course-taking, disciplinary incidents, or absences). In our core specification, we estimate β using the following equation and OLS:

$$Y_i = \gamma + \beta B_i + k(S_i) + k(S_i) * B_i + v_i \tag{2}$$

where $k(S_i)$ is a polynomial function of the relative test score. We estimate this model using the linear polynomial specification, and cluster the standard errors at the third grade reading score level as suggested by Lee and Card (2009).[11] In our main analysis, we use a

---

[11] This clustering approach has been recently questioned (Kolesár and Rothe 2018), and we check the robustness of our findings and compare the 95% confidence intervals obtained using standard errors clustered at the running variable level with confidence intervals obtained using Eicker-Huber-White heteroskedasticity-robust standard errors in Appendix Figure 3. These two approaches yield very similar conclusions using different bandwidths for our main outcomes of interest.

bandwidth of 20 points based on the range of bandwidths suggested for various outcomes by the bandwidth selection procedure in Calonico et al. (2017), and check the robustness of our findings to different bandwidths in Appendix Figure 3.

In this empirical framework, $\hat{\beta}$ will yield unbiased estimates of the causal effect of failing the third grade reading test if all other student attributes are smooth around the cutoff. While this 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:

$$X_i = \alpha + \varphi B_i + k(S_i) + k(S_i) * B_i + \varepsilon_i \qquad (3)$$

where X_i represents baseline student characteristics including third grade math scores. In Table 2, we present falsification tests where each row represents a separate regression using the identified variable as the dependent variable estimated using the bandwidth of 20 points and the linear polynomial specification, and the estimated coefficient ($\hat{\varphi}$) indicates the size of the discontinuity. Online Appendix Figure 1 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 concerning 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. Online Appendix Figure 2 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 the retention policy on English learners' English skills as proxied by their reading test scores in elementary and middle school.[12] 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, and provide the effect sizes in standard deviation units in brackets. We present the results of the same-age analysis in the top panel of Table 3 (first column), whereas panels (A) and (B) of Figure 1 provides a visual inspection of the discontinuity in reading scores in the years following third grade at the retention cutoff using local linear smoothing.

The results suggest large benefits of the retention policy on the reading scores of English learners in the short run. In particular, English learners who just failed the third grade reading test outperform their peers who just passed the test by nine percent of the standard deviation (σ) in the first year, 0.17σ in the second year, and 0.12σ in the third year. These benefits dissipate

---

[12] While the FCAT reading test does not specifically target English learners, and does not measure the listening, speaking, or writing skills of English learners, FCAT reading scores for English learners are highly correlated with the English proficiency test scores, which are used to determine whether English learners are ready to exit the program. For example, in a Florida district where we observe both FCAT reading and the overall score on the English proficiency test for English learners, we observe a correlation of 0.7. Therefore, we believe that FCAT reading scores provide a good proxy for the student's overall English skills.

quickly in middle school, with fourth-year effects of 0.06σ, and fifth-year effects of 0.04σ, and the estimated effect in year 5 is no longer statistically different than zero at conventional levels. The second column in the top panel of Table 3 compares these effects with the same-age effects for non-English learners and suggest that while the effects are comparable in the first two years, the effects of the retention policy on English learners in years 3 to 5 are about 10 to 50 percent larger than the effects on non-English learners. [13]

The bottom panels of Table 3 and panels (A) and (B) of Appendix Figure 1 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 the reading scores of English learners that persist through eighth grade, with effect sizes of roughly 0.15σ 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. Once again, these same-grade effects are considerably larger in middle school (by about 20 percent) for English learners compared to non-English learners.

5.2. Estimated Effects on Time to Proficiency

There are two important implications of improved English skills for low-performing English learners. First, these students could attain English proficiency and leave the English learner program earlier than their peers who just passed the third grade reading test.[14] The top

---

[13] We also examine the effects of the retention policy on the math scores of English learners. The results, which are available upon request, are very similar to the reading effects presented in Table 3.

[14] During the time frame we examine, English learners in third through twelfth grades in Florida were required to demonstrate proficiency in speaking, listening, reading, and writing on tests approved by Florida Department of Education (FLDOE) including the Comprehensive English Language Learning Assessment (CELLA), FCAT reading and writing, or another nationally norm referenced test approved by FLDOE. School districts had flexibility over which test would be utilized to assess English learner proficiency, yet they were required to use the proficiency thresholds set by the state. Almost all districts in our sample used CELLA to assess English learner proficiency in speaking, listening, and writing, and a combination of CELLA and FCAT scores to determine proficiency in

panel of Table 4 presents the 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 panel (C) of Figure 1 presents the local linear smoothing of the years-to-proficiency variable on the relative reading 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 English learners right below the retention cutoff spend 0.2 fewer years in the English learner program compared to their peers right above the cutoff. Given that English learners whose third grade reading scores fell right above the cutoff spend an additional two years in the English learner program, this effect size implies that the grade retention policy reduces additional time spent in English learner program by 10 percent.

The second panel of Table 4 repeats 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, compared to their peers who just passed the third grade test, just-failing English learners are two percentage points (or three percent of the control mean at the cutoff) less likely to be identified as English learners in the first year, which increases to ten percent of the control mean at the cutoff in the second year, and 18 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. The results of the same-grade analysis, reported in the third panel of Table 4, reveal similar findings and show that the

---

reading. CELLA had different proficiency clusters based on grade level (K-2, 3-5, 6-8, and 9-12), but the proficiency thresholds were identical across grades within these levels. Therefore, retained and promoted English learners faced the same proficiency thresholds in many cases.

retention policy 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. Therefore, the early grade retention policy could reduce the likelihood of taking remedial reading courses in middle school due to the aforementioned positive effects on the reading scores of English learners. In Table 5, we address this question for English learners in the first column, and for non-English learners in the second column. The first row of Table 5 presents the estimated effects of failing the third grade test on the likelihood of taking at least one remedial reading course in middle school (we present graphical evidence in panel (C) of Appendix Figure 1).[15] The results indicate that just-failing English learners are seven percentage points (or ten percent of the control mean) less likely to take a remedial English course.

The reduced need for English remediation could also free up time and allow these students to take more advanced courses in middle school.[16] We explore this possibility in rows (II) to (IV) of Table 5, panels (D) and (E) of Figure 1, and panel (D) of Appendix Figure 1, where we examine the effects of the early grade retention policy on whether the student (I) took

---

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

[16] 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.

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 failing the third grade reading test on the middle school course-taking behavior among English learners. In particular, we find that just-failing English learners are about five percentage points (or 25 percent of the control mean at the cutoff) more likely to take an advanced English course, eight percentage points (or 25 percent of the control mean at the cutoff) more likely to take an advanced math course, and six percentage points (or 20 percent of the control mean at the cutoff) more likely to take an advanced science course in middle school.

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 or International Baccalaureate (IB) courses, potentially better preparing them for college.[17] 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).[18] Once again, our results point to a sizable effect of the early grade retention policy on course-taking among low-performing English learners: The estimates, presented in the last row of Table 5 (graphical inspection provided in the panel (F) of Figure 1) reveal that failing the third grade test increases the likelihood of taking

[17] There is less evidence on the causal link between taking college credit-bearing courses in high school and later outcomes (Kolluri 2018). Smith et al. (2017) examine the effect of receiving credit-granting scores on AP exams using a regression discontinuity design, and find that attaining a passing score increases the probability that a student will receive a bachelor's degree within four years by 1 to 2 percentage points. While we are unable to examine the effects of the retention policy on passing the AP/IB exams, it is plausible that the policy could improve this likelihood (1) through its effect on student human capital when they reach high school as shown in Table 4 and (2) through its effect on the likelihood that English learners take these courses in high school as shown in Table 6.
[18] 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 data (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 eleventh grade took one of these courses in 12th grade.

college credit-bearing courses in high school for English learners by 10 percentage points, increasing the probability of taking these courses by roughly 80 percent of the control mean at the cutoff.[19]

In the second column of Table 5, we present the effects of retention on course-taking for non-English learners. The benefits of the retention policy for English learners are slightly larger than for non-English learners in many cases, but the differences are small. The most striking difference between the effects sizes for the two groups is in the effect of the policy on advanced course taking in high school wherein we find no significant effect on non-English learners both statistically and economically, yet sizable effects on English learners.

5.4. Estimated Effects on Disciplinary Incidents and Absences

An important concern with grade retention policies is that retention could impose 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. We examine these possible adverse effects using (1) severe disciplinary incidents that lead to in-school or out-of-school suspensions, and (2) student absences as two proxies.

Table 6 and Appendix Table 1 present the estimated effects of failing the third grade reading test on disciplinary incidents and absence rate respectively (graphical inspection is given in Appendix Figure 2). Once again, we examine the same-age (top panels) and same-grade (bottom panels) effects separately for English learners and non-English learners. In the same-age

---

[19] We check the robustness of the findings presented in Tables 3, 4, and 5 using the oldest cohort in our sample, which we use in the high school course-taking analysis. The results, which are available upon request, are very similar to those obtained using the full sample.

comparisons, we control for peer incident rate at the school-by-grade level to account for differences in disciplinary enforcement across schools and grade levels.

Similar to the overall effects presented in Ozek (2015), we find that just-failing English learners are more likely to have disciplinary problems in the first year compared to their peers who just pass the test, yet these adverse effects vanish in middle school. Further, among English learners, the results suggest that the policy reduces the likelihood of being involved in disciplinary incidents in middle school considerably. In particular, we find that English learners right below the retention cutoff are seven percentage points (or roughly 20 percent of the dependent variable mean at the cutoff) less likely to be involved in a disciplinary incident during middle school. We find no such effect for non-English learners. Appendix Table 1 repeats the same analysis for absence rates and reveals no significant effect of retention on absence rates.

5.5. Extensions

Do certain English learner groups benefit more from the grade retention policy compared to others? While we do not have sufficient power to statistically distinguish the estimated effects across different English learner groups, in the online appendix, we present the results of an exploratory analysis providing the estimated effects of failing the third grade test (along with their 95% confidence intervals) for different student groups of interest. Perhaps the most interesting takeaway from this analysis is the discrepancies between foreign-born and native English learners. For example, while the short-term effects of the policy on reading scores are comparable for U.S. born and immigrant English learners, the longer-term benefits are larger for the latter group. 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

larger for native-born English learners for all other course-taking outcomes. Further, we find evidence suggesting that the benefits of grade retention on disciplinary problems in middle school are more pronounced for foreign born English learners. For these students, we also find evidence that retention reduces absence rates in middle school.

An important limitation of our RD design is that the estimated effects are only applicable to students around the retention cutoff. As such, it is hard to generalize our results to English learners with lower reading scores. To examine the broader effects of Florida's grade retention policy on low-performing English learners, we employ a difference-in-difference-in-differences (DDD) approach and the sample of first time third and fourth graders in 2000-01, 2001-02, and 2002-03 school years who score in the lowest two achievement levels in reading.[20] The findings, which we present in the online appendix, reveal that the broader effects of the retention policy on low-performing English learners are smaller than the effects on the English learners around the cutoff, yet provides evidence that the retention policy has a positive effect on all low-performing English learners.

5.6. Robustness Checks

Appendix Figure 3 checks the sensitivity of our main findings to bandwidth selection and provides the effects (and the 95% confidence intervals) of failing the third grade reading test estimated using different bandwidths and the linear polynomial specification. The estimated discontinuities at the cutoff are extremely robust to bandwidth selection with the exception of the effects on disciplinary problems. Even in that case, the signs of the point estimates are consistent

---

[20] In particular, in this analysis, we compare the lowest-scorers to the next-lowest-scorers, in third grade versus next grade, after the retention policy versus before the retention policy.

even though the effect sizes vary across different bandwidths, and hence 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 the grade retention policy. To test this hypothesis, we construct "pseudo" cutoffs using earlier cohorts who were not subject to the retention policy (English learners who first entered third grade in 2001–02) and check the discontinuities in outcomes. Appendix Figure 4 presents these findings where the top panel presents these pseudo effects and the bottom panel presents the actual effects as a comparison. The findings presented in the top panel show that there are no statistically significant discontinuities in our main outcome variables at these pseudo-cutoffs, and that the magnitudes of these pseudo effects are much smaller than the real effects presented in the bottom panel. This exercise provides 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, Appendix Figure 5 checks the discontinuities in attrition rates around the cutoff in the five years following the first time students enter the third grade, and by the time students enter eleventh 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.

21

6. Interpretation of Results Through Instrumental Variables

We have thus far focused on reduced-form effects, which is useful to assess the overall effects of the policy on low-performing English learners. However, from a policy perspective, it is also important to understand the extent to which the treatment effects are driven by the effect of the policy on retained students through retention and instructional support (e.g., summer school, support in the following school year). An important challenge in this context is that Florida's retention policy allows all third graders whose reading scores fall in the lowest achievement level to attend summer school including those who are exempt from retention, which could imply that the exclusion restriction is violated in a two-stage least squares (2SLS) framework where we instrument for grade retention (R_i) using B_i as follows:

$$R_i = \gamma + \delta B_i + k(S_i) + k(S_i) * B_i + \omega_i \tag{5}$$

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

$$Y_i = \rho + \mu \widehat{R}_i + k(S_i) + k(S_i) * B_i + \sigma_i. \tag{6}$$

Figure 2 presents the estimated effect of failing the third grade reading test on retention (first row) and summer school enrollment (second row) for English learners after the policy (left column) and in the year before the policy (right column). The results reveal a significant discontinuity in retention probability (roughly equal to 30 percentage points) and a much larger discontinuity in summer school enrollment (equal to 55 percentage points) at the retention cutoff after the policy took effect, yet no such discontinuities in the year before the policy. This implies that failing the third grade reading test indeed increases the likelihood of summer school at the cutoff for English learners who are exempt from retention under Florida's policy.

We then examine the extent to which summer school improves the educational outcomes of English learners who were exempt from retention. While it is not feasible to identify English learners who were eligible for certain exemptions (e.g., promotion through teacher portfolio), we focus on three student groups who are exempt from retention under Florida's policy: (1) English learners who were in the program for less than two years; (2) English learners who scored at or above the 51st percentile on the FCAT-NRT reading test; and (3) English learners who were identified as special education students and had been retained at least once until third grade. These students constitute 86 percent of all English learners who were promoted to fourth grade with exemptions despite failing the third grade reading test during the time frame we examine. Using this sample, we estimate the effect of failing the third grade test on summer school enrollment, retention, and our outcomes of interest.[21]

The results presented in Table 7 reveal that exempt English learners who scored right below the retention cutoff were 29 percentage points more likely to attend summer school at the end of third grade, yet we find very small effects on retention. We find no significant effect of failing the third grade test on student outcomes, with significantly smaller and opposite signed point estimates in many cases compared to the estimates presented using all English learners. These findings provide evidence that the reduced-form estimates presented earlier are primarily driven by the effect of the policy on retained students through instructional support (summer school and support in the following year) and the retention itself.[22]

---

[21] In this analysis, due to sample size issues, we are unable to examine the effect on AP/IB course-taking in high school.

[22] It is important to note that this finding is not sufficient to rule out summer school as a mediator for retained students as the effect of summer school could be different for exempt English learners compared to retained English learners.

With this finding in mind, Table 8 presents the 2SLS estimates (μ) of the effects of grade retention with instructional support. The first row presents the first stage, indicating that failing the third grade test increases retention probability by 29 percentage points at the retention cutoff. We find significant effects of retention on the reading scores of English learners that persist through middle school. The results suggest large benefits of retention on the reading scores of English learners in the short run, with 0.49 percent of the standard deviation in years 1 to 3, which decline to 0.18σ in years 4 and 5, yet remain statistically and economically significant distinct from zero. Same-grade comparisons reveal even larger and more persistent effects with effect sizes of 0.85σ in elementary grades (grades 4 and 5) and 0.65σ in middle school grades (grades 6 to 8).

The results also suggest that retention coupled with instructional support reduces the time to proficiency by about 0.8 years (or by 40 percent of the control mean at the cutoff), decreases 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 tripling the likelihood of taking college credit-bearing courses in high school. Grade retention also reduces disciplinary incidents in middle school considerably although the estimated coefficients are not precisely estimated using lower bandwidths.[23]

---

[23] In the online appendix, we also examine the extent to which retained English learners received instructional support during summer and in the following year along three dimensions: (1) enrollment in summer school; (2) reading and ELA instruction time per week in the following year; and (3) reading teacher experience in the following year. We find that the summer school take-up rate was high among English learners, while we did not find any significant discontinuities in (2) or (3) at the third grade retention cutoff.

7. Conclusion

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 coupled with instructional support 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.

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

25

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 such as Florida's policy 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.

It is very important to underscore that, while grade retention is a major component of the policy we evaluate, it is not the only component. Florida's policy couples grade retention with additional reading instruction, summer school, placement with a teacher deemed to be effective, and an academic improvement plan. And while our results illustrate the potential benefits of giving English learners additional time to learn English, there are still ample risks associated with implementing a program such as Florida's, and we are only able to address a few of these potential risks in this study. Mandatory grade retention is far from a "silver bullet", and any

policy that is implemented, especially pertaining to vulnerable populations such as English

learners, should be carefully monitored to gauge both the potential areas of success as well as the

potential areas of risk.

References

Calonico, S., Cattaneo, M., Farrell, M., & Titiunik R. (2017). "rdrobust: Software for regression-discontinuity designs," The Stata Journal, 17 (2), 372–404.

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

ECS (2018), State Kindergarten-Through-Third gradePolicies: Is there a third grade retention policy? http://ecs.force.com/mbdata/MBQuest2RTanw?rep=KK3Q1818.

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

Eren, O., Depew, B., & Barnes, S. (2017). Test-based promotion policies, dropping out, and juvenile crime. Journal of Public Economics, 153, 9–31.

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.

28

Halle, T., Hair, E., Wandner, L., Mcnamara, M., & Chien, N. (2012). Predictors and outcomes of early versus later English language proficiency among English language learners. Early Childhood Research Quarterly, 27(1), 1–20.

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

Kao, G. & Tienda, M. (1995). Optimism and Achievement: The Educational Performance of Immigrant Youth. Social Science Quarterly 76, 1–19.

Kolluri, S. (2018). Advanced Placement: The Dual Challenge of Equal Access and Effectiveness. *Review of Educational Research*, *88*(5), 671–711.

Larsen, M. F., & Valant, J. (2018). Socially Promoted, Academically Retained: The Effects of Assorted Grade Retention Policies on High School and College Persistence. Unpublished Manuscript.

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

Mariano, L. T., Martorell P., & Berglund, T. (2018). The Effects of Grade Retention on High School Outcomes: Evidence from New York City Schools. Santa Monica, CA: RAND Corporation.

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

McFarland, J., Hussar, B., Zhang, J., Wang, X., Wang, K., Hein, S., Diliberti, M., Forrest Cataldi, E., Bullock Mann, F., and Barmer, A. (2019). The Condition of Education 2019 (NCES 2019-144). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved 1/3/2020 from https://nces.ed.gov/ pubsearch/pubsinfo.asp?pubid=2019144.

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.

OECD (2018), The Resilience of Students with an Immigrant Background: Factors that Shape

Well-being, OECD Reviews of Migrant Education, OECD Publishing, Paris

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.

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, G., West, M. R., & Winters, M. A. (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.

Steele, J., Slater, R., Zamarro, G., Miller, T., Li, J., Burkhauser, S., & Bacon, M. (2017). Effects of Dual-Language Immersion Programs on Student Achievement: Evidence From Lottery Data. American Educational Research Journal, 54(1), 282S–306S.

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

Umansky, I., & Reardon, S. (2014). Reclassification Patterns Among Latino English Learner Students in Bilingual, Dual Immersion, and English Immersion Classrooms. American Educational Research Journal, 51(5), 879–912.

Valentino, R., & Reardon, S. (2015). Effectiveness of Four Instructional Programs Designed to Serve English Learners: Variation by Ethnicity and Initial English Proficiency. Educational Evaluation and Policy Analysis, 37(4), 612–637.

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.

Tables and Figures

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

| | Non-English Learner Students | English Learners | | | |
		All	Retained	Promoted, Enrolled in Summer School	Promoted, No Summer School
Third grade reading score	0.020	-0.819	-1.708	-1.402	-0.445
	(0.974)	(1.009)	(0.704)	(0.823)	(0.894)
Third grade math score	-0.0803	-0.744	-1.533	-1.149	-0.442
	(0.972)	(1.020)	(0.843)	(0.900)	(0.928)
Below the retention cutoff	0.165	0.451	0.987	0.833	0.220
	(0.371)	(0.498)	(0.115)	(0.373)	(0.414)
Years in EL program		2.238	2.764	2.069	2.193
		(1.434)	(1.152)	(1.613)	(1.419)
Retained	0.0912	0.185			
	(0.288)	(0.388)			
Eligible for subsidized meals	0.520	0.801	0.891	0.870	0.773
	(0.500)	(0.399)	(0.311)	(0.336)	(0.419)
Disciplinary incident	0.042	0.030	0.051	0.038	0.022
	(0.200)	(0.169)	(0.220)	(0.192)	(0.147)
Percent absent days	0.045	0.041	0.045	0.038	0.039
	(0.046)	(0.042)	(0.045)	(0.040)	(0.041)
White	0.433	0.064	0.039	0.051	0.070
	(0.496)	(0.244)	(0.194)	(0.220)	(0.255)
Black	0.275	0.159	0.216	0.190	0.143
	(0.447)	(0.366)	(0.412)	(0.392)	(0.350)
Hispanic	0.226	0.727	0.716	0.711	0.733
	(0.418)	(0.445)	(0.451)	(0.453)	(0.442)
Male	0.508	0.525	0.581	0.544	0.506
	(0.500)	(0.499)	(0.493)	(0.498)	(0.500)
Special education	0.139	0.129	0.226	0.193	0.0911
	(0.346)	(0.335)	(0.418)	(0.395)	(0.288)
Foreign born	0.0658	0.491	0.357	0.521	0.510
	(0.248)	(0.500)	(0.479)	(0.500)	(0.500)
Age (in months)	104.2	104.9	104.7	106.5	104.6
	(5.640)	(7.135)	(6.554)	(8.103)	(6.936)
Spanish speaker	0.181	0.714	0.704	0.689	0.721
	(0.385)	(0.452)	(0.457)	(0.463)	(0.449)
Number of unique students	392,121	40,418	7,141	5,441	25,972

Notes: Standard deviations are given in parentheses. Reading and math scores are standardized at the grade-year level to zero mean and unit variance. Note that the last three columns include English learners who were observed in the year after their first time in third grade. Therefore, the number of observations in these columns do not add up to the total number of English learners in the third column.

Table 2. Baseline Equivalency of Student Characteristics Around the Retention Cutoff

Third grade math score		0.037
		(0.025)
Years in English learner program		-0.027
		(0.049)
Eligible for subsidized meals		0.006
		(0.012)
Third grade disciplinary incident		-0.012*
		(0.007)
Third grade percent absent days		-0.001
		(0.001)
White		0.005
		(0.011)
Black		0.001
		(0.012)
Hispanic		-0.004
		(0.017)
Male		-0.013
		(0.022)
Special education		-0.019
		(0.011)
Foreign born		0.012
		(0.016)
Age (in months)		-0.409
		(0.333)
Joint test for balance on observables		
	F-stat	1.21
	p-value	0.30
Number of unique students		9,921

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

Table 3. Estimated Effects of Failing Third Grade Reading Test on Future Reading Test Scores for English Learners

	Same-Age Analysis	
	English Learners	Non-English Learners
1 year later (SD of Y = 352)	29.969***	26.057***
	(8.397)	(3.429)
Effect size as % of SD	[0.085σ]	[0.074σ]
2 years later (SD of Y = 316)	53.634***	60.495***
	(7.547)	(4.411)
Effect size as % of SD	[0.17σ]	[0.191σ]
3 years later (SD of Y = 311)	37.936***	28.807***
	(7.978)	(3.682)
Effect size as % of SD	[0.122σ]	[0.093σ]
4 years later (SD of Y = 302)	16.775**	15.365***
	(8.257)	(2.914)
Effect size as % of SD	[0.056σ]	[0.051σ]
5 years later (SD of Y = 253)	8.825	3.948
	(7.750)	(4.096)
Effect size as % of SD	[0.035σ]	[0.016σ]
	Same-Grade Analysis	
Fourth grade (SD of Y = 318)	80.930***	82.588***
	(8.457)	(3.771)
Effect size as % of SD	[0.255σ]	[0.260σ]
Fifth grade (SD of Y = 302)	64.822***	66.705***
	(8.258)	(4.048)
Effect size as % of SD	[0.215σ]	[0.221σ]
Sixth grade (SD of Y = 305)	57.179***	46.987***
	(10.134)	(3.803)
Effect size as % of SD	[0.187σ]	[0.154σ]
Seventh grade (SD of Y = 274)	44.938***	39.453***
	(6.673)	(3.906)
Effect size as % of SD	[0.164σ]	[0.144σ]
Eighth grade (SD of Y = 221)	32.853***	25.752***
	(6.915)	(3.057)
Effect size as % of SD	[0.149σ]	[0.117σ]

Notes: All regressions control for the baseline student characteristics listed in Table 2, and standard errors that are clustered at the third grade reading score level are given in parentheses. The estimates represent the discontinuities in student test scores in the years (upper panel) or the grades (lower panel) following third

grade at the retention cutoff, obtained using linear polynomial specification and a bandwidth of 20 points. [*], [**], and [***] represent statistical significance at 10, 5, and 1 percent, respectively. Regressions in the upper panel include 9,446; 9,134; 8,789; 8,382; and 7,854 English learners, and 56,156; 54,423; 52,119; 49,654; and 46,064 non-English learners 1, 2, 3, 4, and 5 years after third grade respectively. Regressions in the lower panel include 9,182; 8,908; 8,524; 8,219; and 7,991 English learners, and 54,026; 52,150; 49,964; 48,343; and 47,122 non-English learners in third, fourth, fifth, sixth, seventh, and eighth grades, respectively. The numbers in brackets represent the estimated effect sizes as a percentage of the dependent variable standard deviation.

Table 4. Estimated Effects of Failing Third Grade Reading Test on Time to Proficiency for English Learners

Years to proficiency after third grade	-0.214***
	(0.054)
Control mean at cutoff	2.08

	Same-Age Analysis
Classified as English learner…	
1 year later	-0.020
	(0.020)
Control mean at cutoff	0.738
2 years later	-0.046***
	(0.016)
Control mean at cutoff	0.479
3 years later	-0.061***
	(0.020)
Control mean at cutoff	0.327
4 years later	-0.044***
	(0.014)
Control mean at cutoff	0.225
5 years later	-0.023*
	(0.012)
Control mean at cutoff	0.155

	Same-Grade Analysis
Classified as English learner in…	
Fourth grade	-0.113***
	(0.019)
Control mean at cutoff	0.733
Fifth grade	-0.092***
	(0.016)
Control mean at cutoff	0.470
Sixth grade	-0.078***
	(0.016)
Control mean at cutoff	0.319
Seventh grade	-0.048***
	(0.014)
Control mean at cutoff	0.220
Eighth grade	-0.020
	(0.013)
Control mean at cutoff	0.142

Notes: All regressions control for the baseline student characteristics listed in Table 2, and standard errors that are clustered at the third grade reading score level are given in parentheses. The estimates represent the discontinuities in time (in years) to proficiency after third grade (top panel), and the likelihood of being classified as an English learner in the years (second panel) or the grades (third panel) following third grade at the retention cutoff, obtained using linear polynomial specification and a bandwidth of 20 points. The years to proficiency variable captures the number of years between the first time the student enters third grade and the first year the student is no longer

classified as an English learner. *, **, and *** represent statistical significance at 10, 5, and 1 percent, respectively. Regressions in the top panel include 9,861 English learners. Regressions in the second panel include 9,446; 9,134; 8,789; 8,382; and 7,854 English learners 1, 2, 3, 4, and 5 years after third grade respectively, and the third panel include 9,182; 8,908; 8,524; 8,219; and 7,991 English learners in third, fourth, fifth, sixth, seventh, and eighth grades, respectively.

Table 5. Estimated Effects of Failing Third Grade Reading Test on Middle School and High School Course-Taking Among English Learners

	English Learners	Non-English Learners
Remedial language arts course in MS	-0.071***	-0.067***
	(0.017)	(0.006)
Control mean at cutoff	0.721	0.793
Advanced language arts course in MS	0.050**	0.071***
	(0.019)	(0.008)
Control mean at cutoff	0.308	0.250
Advanced math course in MS	0.075***	0.061***
	(0.019)	(0.008)
Control mean at cutoff	0.299	0.260
Advanced science course in MS	0.058***	0.063***
	(0.016)	(0.007)
Control mean at cutoff	0.294	0.239
College-credit bearing course by 12th grade	0.102***	0.014
	(0.027)	(0.009)
Control mean at cutoff	0.154	0.165

Notes: All regressions control for the baseline student characteristics listed in Table 2, and standard errors that are clustered at the third grade reading score level are given in parentheses. The estimates represent the discontinuities in the likelihood of taking at least one course in the corresponding category, obtained using linear polynomial specification and a bandwidth of 20 points. *, **, and *** represent statistical significance at 10, 5, and 1 percent, respectively. Regressions for middle school course-taking include 7,991 English learners and 47,122 non-English learners, and regressions for high school course-taking include 2,023 English learners and 11,992 non-English learners.

Table 6. Estimated Effects of Failing Third Grade Reading Test on Disciplinary Incidents for English Learners

	Same-Age Analysis	
	English Learners	Non-English Learners
1 year later	0.013*	-0.003
	(0.006)	(0.004)
Control mean at cutoff	0.037	0.096
2 years later	-0.007	0.008
	(0.010)	(0.005)
Control mean at cutoff	0.050	0.098
3 years later	-0.043**	-0.016**
	(0.017)	(0.007)
Control mean at cutoff	0.216	0.309
4 years later	-0.057***	0.005
	(0.018)	(0.008)
Control mean at cutoff	0.256	0.370
5 years later	-0.068***	-0.020*
	(0.017)	(0.010)
Control mean at cutoff	0.314	0.374
	Same-Grade Analysis	
Fourth grade	0.015**	0.001
	(0.007)	(0.004)
Control mean at cutoff	0.042	0.097
Fifth grade	-0.013	0.005
	(0.011)	(0.005)
Control mean at cutoff	0.052	0.096
Sixth grade	-0.029	0.017*
	(0.018)	(0.009)
Control mean at cutoff	0.226	0.334
Seventh grade	-0.054***	0.010
	(0.015)	(0.009)
Control mean at cutoff	0.266	0.369
Eighth grade	-0.063***	-0.009
	(0.019)	(0.009)
Control mean at cutoff	0.297	0.362

Notes: All regressions control for the baseline student characteristics listed in Table 2, and standard errors that are clustered at the third grade reading score level are given in parentheses. The estimates represent the discontinuities in the likelihood of being involved in a disciplinary incident in the years (upper panel) and the grades (lower panel) after the third grade, obtained using linear polynomial specification and a bandwidth of 20 points. *, **, and *** represent statistical significance at 10, 5, and 1 percent, respectively. Regressions in the upper panel include 9,446; 9,134; 8,789; 8,382; and 7,854 English learners, and 56,156; 54,423; 52,119; 49,654; and 46,064 non-English learners 1, 2, 3, 4, and 5 years after third grade respectively. Regressions in the lower panel include 9,182; 8,908; 8,524; 8,219; and 7,991 English learners, and 54,026; 52,150; 49,964; 48,343; and 47,122 non-English learners in third, fourth, fifth, sixth, seventh, and eighth grades, respectively.

Table 7. Estimated Effects of Failing the Third Grade Reading Test, English Learners Exempt from Retention

Enrolled in summer school	0.291***
	(0.031)
Retained	0.042**
	(0.018)
Reading scores -	
Grades 4 and 5	-3.220
	(12.536)
Effect size as % of SD	[0.01σ]
Grades 6 to 8	15.407
	(12.485)
Effect size as % of SD	[0.057σ]
Time to proficiency	0.126
	(0.155)
Remedial English in middle school	0.036
	(0.037)
Advanced English in middle school	-0.042
	(0.034)
Advanced Math in middle school	-0.004
	(0.043)
Advanced Science in middle school	-0.004
	(0.039)

Notes: All regressions control for the baseline student characteristics listed in Table 2, and standard errors that are clustered at the third grade reading score level are given in parentheses. The estimates represent the discontinuities in the likelihood of being enrolled in summer school right after third grade, the likelihood of being retained after third grade, and the outcomes of interest obtained using linear polynomial specification and a bandwidth of 20 points. *, **, and *** represent statistical significance at 10, 5, and 1 percent, respectively. Regressions include 2,697 English learners exempt from retention in the first two rows, 2,510 and 2,355 English learners in rows 3 and 4, 2,697 English learners in row 5, and 2,135 English learners in rows 6 to 9. The numbers in brackets represent the estimated effect sizes as a percentage of the dependent variable standard deviation.

Table 8. Interpretation of Results through Instrumental Variables – Estimated Effects of Grade Retention on English Learners

|	
First stage *(N=9,464)*	
Retained	0.293***
	(0.019)
Reading scores -	
Years 1 to 3 *(N=9,464)*	131.291***
	(26.390)
	[0.43σ]
Years 4 and 5 *(N=8,397)*	47.596*
	(25.810)
	[0.18σ]
Grades 4 and 5 *(N=9,330)*	250.636***
	(29.630)
	[0.85σ]
Grades 6 to 8 *(N=8,797)*	152.269***
	(20.121)
	[0.52σ]
Time to proficiency *(N=9,464)*	-0.805***
	(0.183)
Remedial English in middle school *(N=7,991)*	-0.245***
	(0.051)
Advanced English in middle school *(N=7,991)*	0.173***
	(0.066)
Advanced Math in middle school *(N=7,991)*	0.260***
	(0.069)
Advanced Science in middle school *(N=7,991)*	0.202***
	(0.058)
AP/IB in high school *(N=2,015)*	0.323***
	(0.098)
Disciplinary incidents -	
Years 1 and 2 *(N=9,385)*	-0.021
	(0.034)
Years 3 to 5 *(N=8,549)*	-0.739***
	(0.158)
Grades 4 and 5 *(N=9,415)*	-0.005
	(0.033)
Grades 6 to 8 *(N=9,037)*	-0.231***
	(0.055)
Absence rate -	
Years 1 and 2 *(N=9,385)*	-0.000
	(0.004)
Years 3 to 5 *(N=8,549)*	-0.005
	(0.008)
Grades 4 and 5 *(N=9,415)*	0.005
	(0.004)
Grades 6 to 8 *(N=9,037)*	-0.004
	(0.009)

Notes: All regressions control for the baseline student characteristics listed in Table 2, and standard errors that are clustered at the third grade reading score level are given in parentheses. The estimates represent the discontinuity in

the likelihood of being retained at the cutoff after third grade, and the effect of grade retention on the outcomes of interest using the 2SLS approach described in equations (5) and (6), linear polynomial specification and a bandwidth of 20 points. *, **, and *** represent statistical significance at 10, 5, and 1 percent, respectively.

Figure 1. Reading Achievement, Time to Proficiency, and Advanced Course-Taking in Middle and High School around the Retention Cutoff

(A) Average Developmental Reading Scores: Years 1 to 3
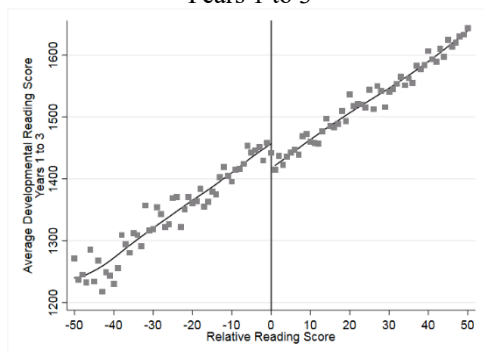
(B) Average Developmental Reading Scores: Years 4 and 5
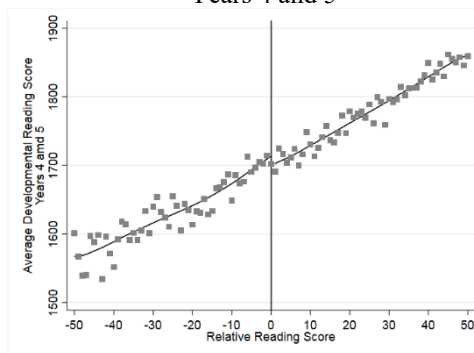
(C) Years to Proficiency
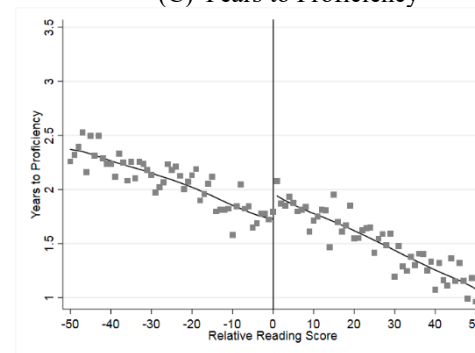
(D) Advanced Language Arts in Middle School
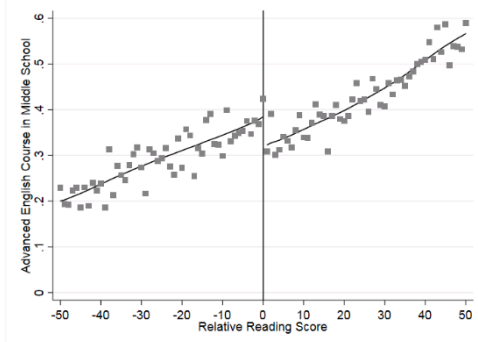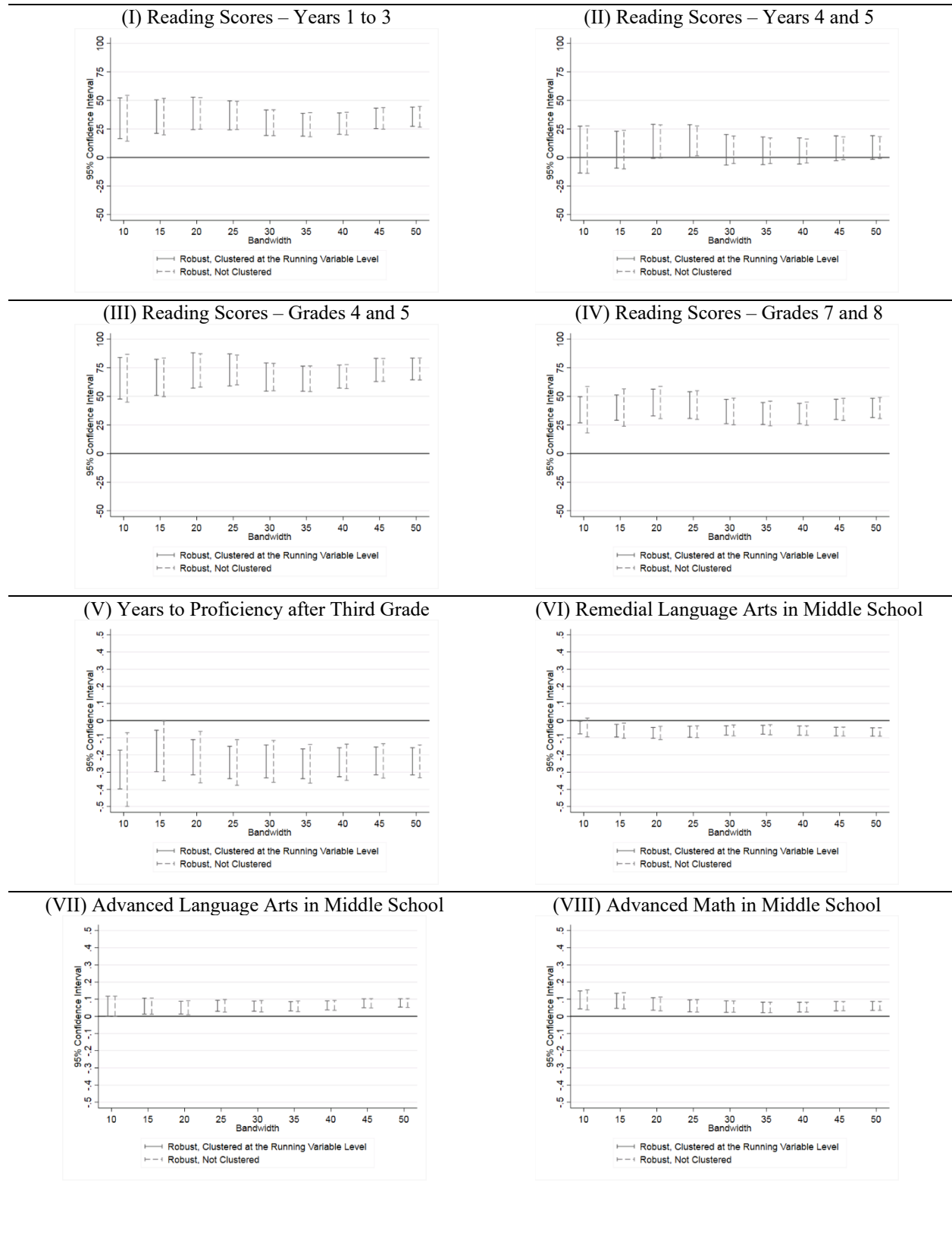
(E) Advanced Math in Middle School
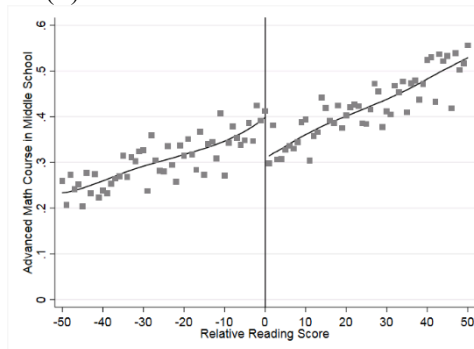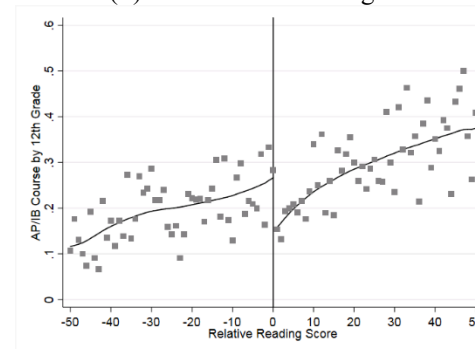
(F) AP-IB Course in High School

Notes: The figures present the local linear smoothing of averaged development reading scores in the following years, time to proficiency (in years), and 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.
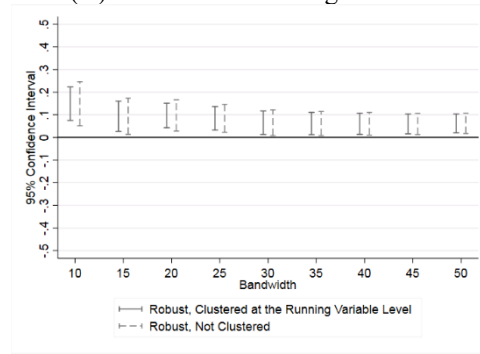
43

Figure 2. Grade Retention and Summer School Among English Learners Around Retention Cutoff

(A) Retention: 2002-05 cohorts	(B) Retention: 2001 cohort

(C) Summer School: 2002-05 cohorts	(D) Summer School: 2001 cohort
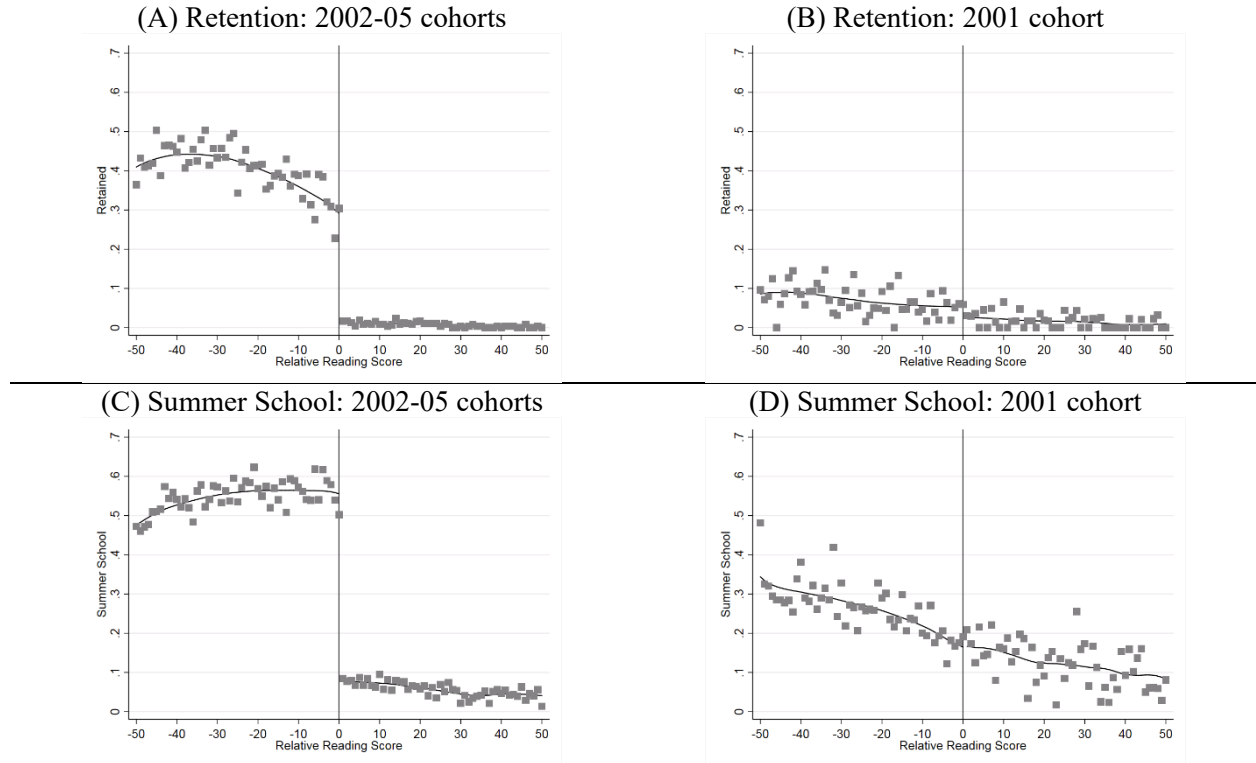
Notes: The figure presents the raw cell means of the corresponding treatment for each reading score between 50 points below and 50 points above the retention cutoff. The solid lines represent the local linear smoothing of the treatment variable separately for the left of the cutoff date and the right. The triangle kernel and a bandwidth of 20 points are used in the estimation.

Appendix Table 1. Estimated Effects of Failing Third Grade Reading Test on Absences for English Learners

		Same-Age Analysis	
		English Learners	Non-English Learners
1 year later		-0.000	0.001
		(0.001)	(0.001)
	Control mean at cutoff	0.046	0.051
2 years later		-0.000	0.000
		(0.001)	(0.001)
	Control mean at cutoff	0.047	0.054
3 years later		-0.001	-0.002
		(0.002)	(0.001)
	Control mean at cutoff	0.051	0.059
4 years later		-0.004	-0.002
		(0.003)	(0.001)
	Control mean at cutoff	0.062	0.065
5 years later		-0.002	0.001
		(0.003)	(0.001)
	Control mean at cutoff	0.066	0.073
		Same-Grade Analysis	
Fourth grade		0.001	0.001
		(0.002)	(0.001)
	Control mean at cutoff	0.046	0.051
Fifth grade		0.001	0.000
		(0.002)	(0.001)
	Control mean at cutoff	0.046	0.053
Sixth grade		-0.003	0.001
		(0.002)	(0.001)
	Control mean at cutoff	0.054	0.060
Seventh grade		-0.000	0.002**
		(0.003)	(0.001)
	Control mean at cutoff	0.057	0.065
Eighth grade		-0.001	0.001
		(0.004)	(0.001)
	Control mean at cutoff	0.068	0.069

Notes: All regressions control for the baseline student characteristics listed in Table 2, and standard errors that are clustered at the third grade reading score level are given in parentheses. The estimates represent the discontinuities in absence rates in the years (upper panel) and the grades (lower panel) after the third grade, obtained using linear polynomial specification and a bandwidth of 20 points. *, **, and *** represent statistical significance at 10, 5, and 1 percent, respectively.

Appendix Figure 1. Reading Achievement Among English Learners around Retention Cutoff
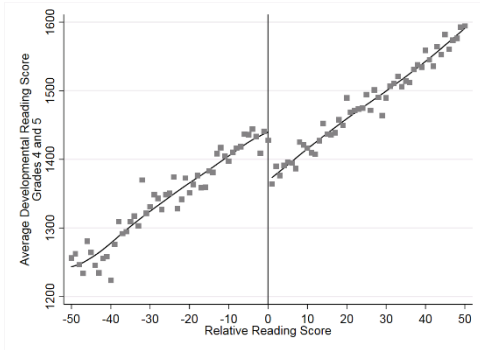
(A) Average Developmental Reading Scores: Grades 4 and 5	(B) Average Developmental Reading Scores: Grades 7 and 8
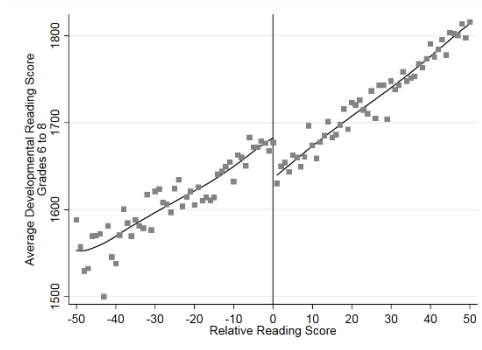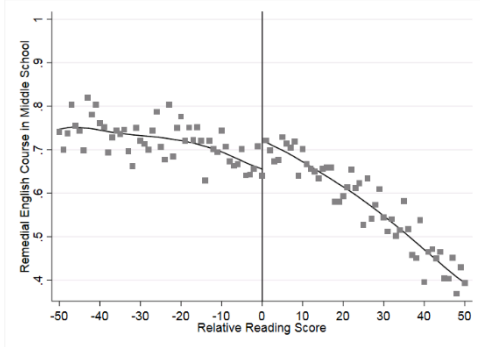
(C) Remedial Language Arts in Middle School	(D) Advanced Science in Middle School
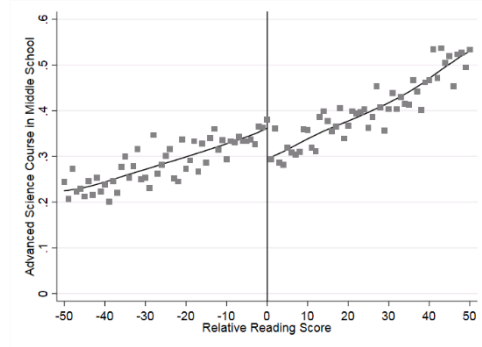
Notes: The figures present the local linear smoothing of averaged development reading scores in the following grades and student course-taking indicators on relative reading score of the student separately for the left of the cutoff date and the right. The triangle kernel and a bandwidth of 20 points are used in the estimation. The solid circles represent raw cell means.

Appendix Figure 2. Disciplinary Incidents and Absences around the Retention Cutoff

(A) Incidents – Years 1 and 2   (B) Incidents – Years 3 to 5

(C) Incidents – Grades 4 and 5   (D) Incidents – Grades 6 to 8

(E) Absence Rate – Years 1 and 2   (F) Absence Rate – Years 3 to 5

(G) Absence Rate – Grades 4 and 5   (H) Absence Rate – Grades 6 to 8
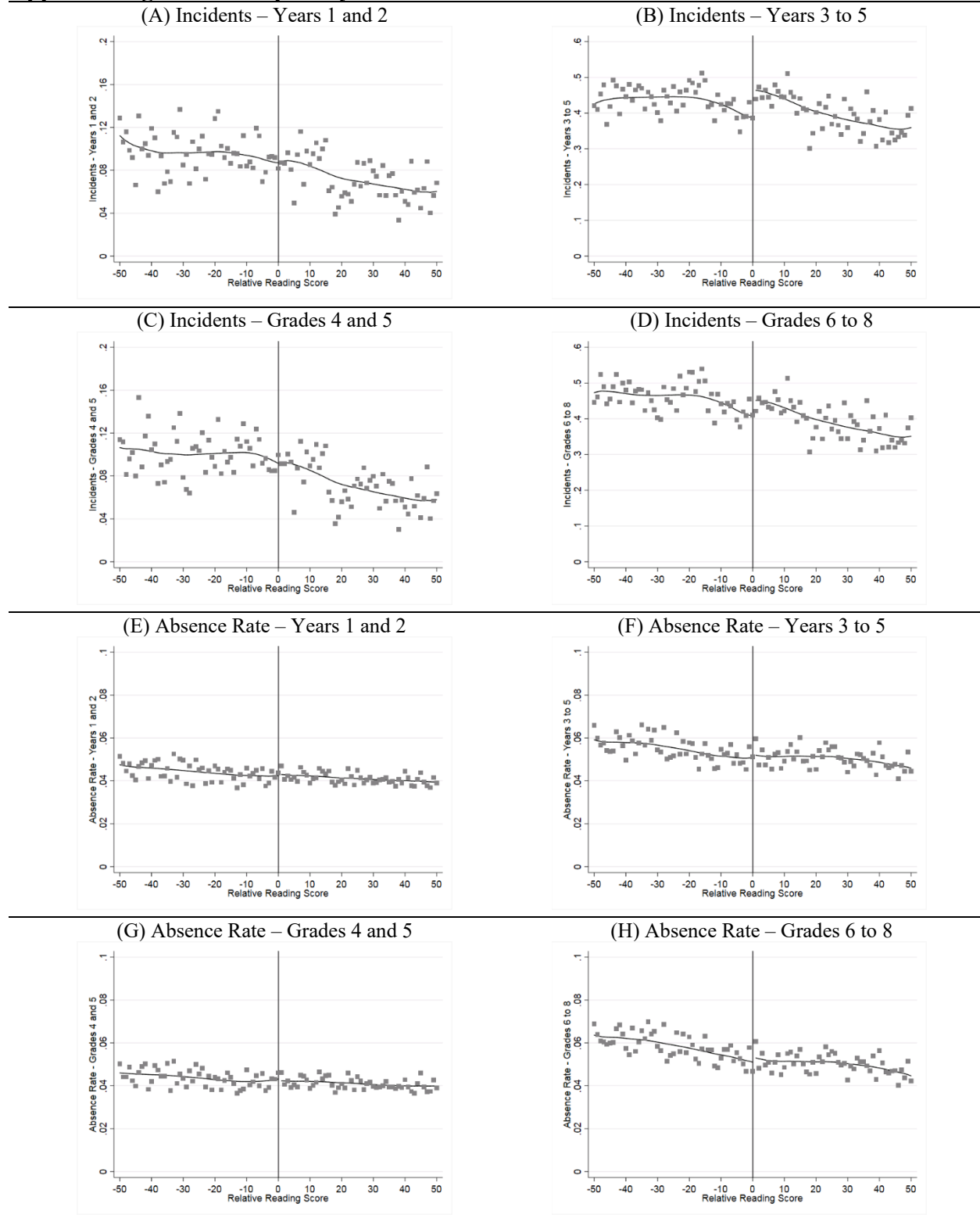
Notes: The figures present the local linear smoothing of disciplinary incidents and absence rates indicators on relative reading score of the student separately for the left of the cutoff date and the right. The triangle kernel and a bandwidth of 20 points are used in the estimation. The solid circles represent raw cell means.

Appendix Figure 3. Robustness to Bandwidth Selection and Standard Error Clustering

(I) Reading Scores – Years 1 to 3	(II) Reading Scores – Years 4 and 5

(III) Reading Scores – Grades 4 and 5	(IV) Reading Scores – Grades 7 and 8

(V) Years to Proficiency after Third Grade	(VI) Remedial Language Arts in Middle School

(VII) Advanced Language Arts in Middle School	(VIII) Advanced Math in Middle School

(IX) Advanced Science in Middle School
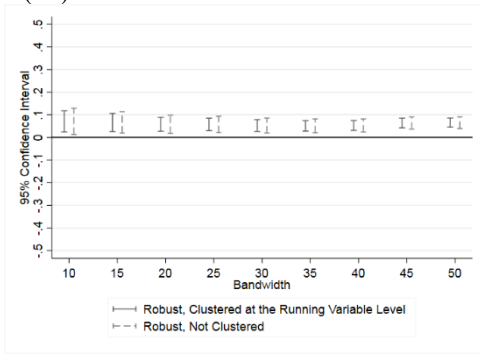
(X) AP/IB Course in High School

(XI) Ever involved in a disciplinary incident
Years 1 and 2
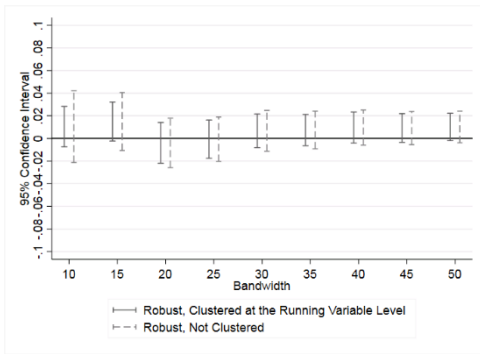
(XII) Ever involved in a disciplinary incident
Years 3 to 5
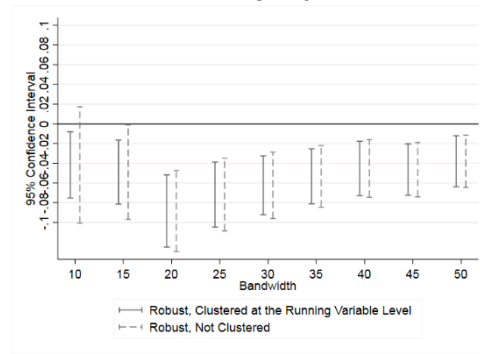
(XIII) Ever involved in a disciplinary incident
Grades 4 and 5
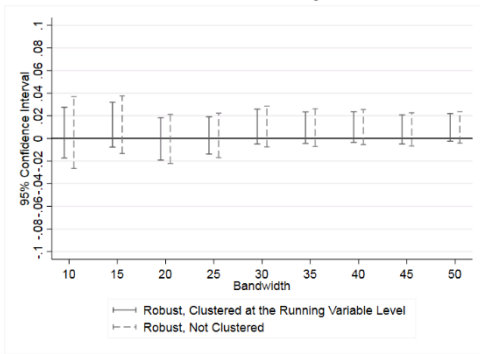
(XIV) Ever involved in a disciplinary incident
Grades 6 to 8
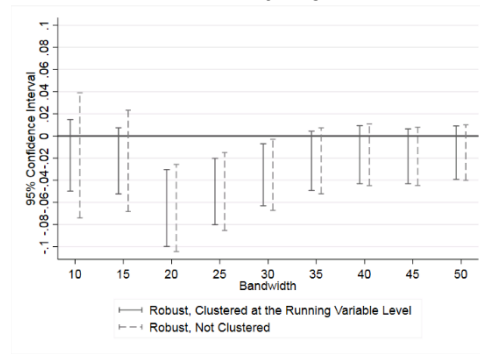
(XV) Average absence rate
Years 1 and 2
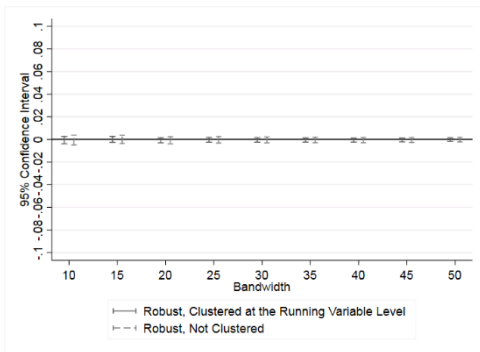
(XVI) Average absence rate
Years 3 to 5
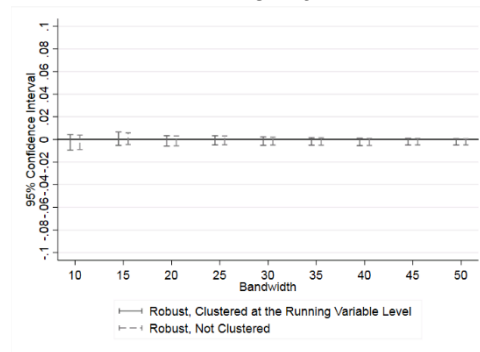
(XVII) Average absence rate
Grades 4 and 5
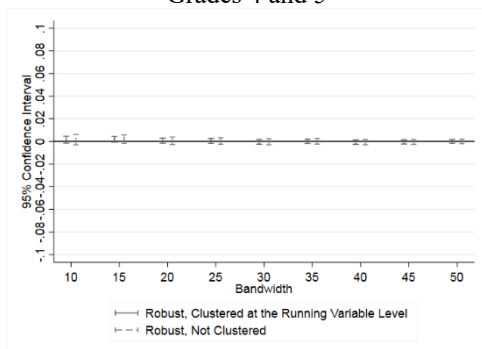
(XVIII) Average absence rate
Grades 6 to 8
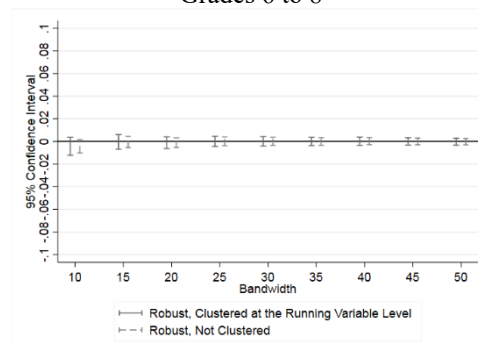
Notes: The figures present the treatment effect (β) and the 95% confidence interval estimated using the bandwidth shown and linear specification, with robust standard errors clustered at the third grade reading score level (solid line) and not clustered (dashed line).

Appendix Figure 4. Falsification Test—Pseudo-Effects

(I) Pseudo-Cutoff using Third Graders in 2001-02

Reading scores	Time to Proficiency and Course-Taking	Disciplinary Incidents and Absences
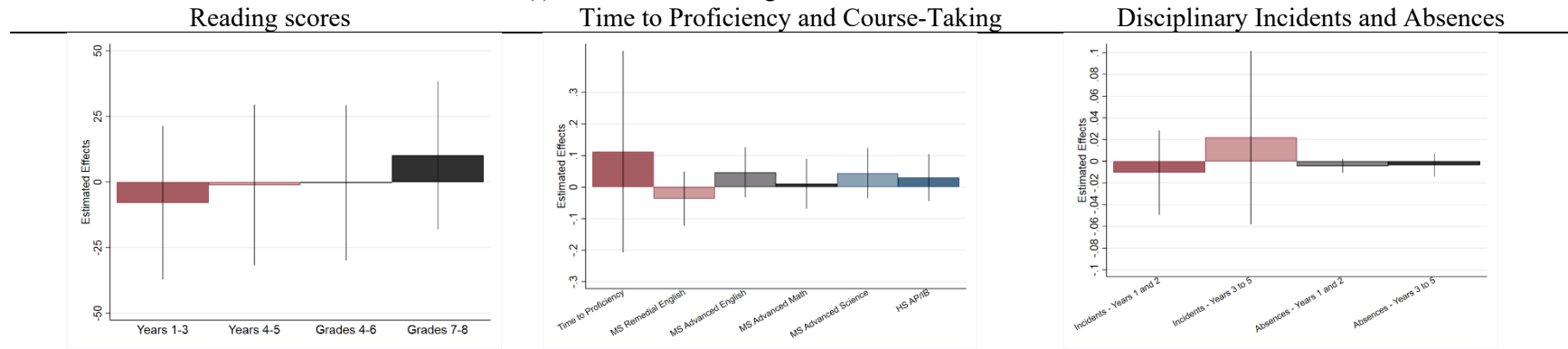
(II) Actual Cutoff in Years between 2002-03 and 2005-06
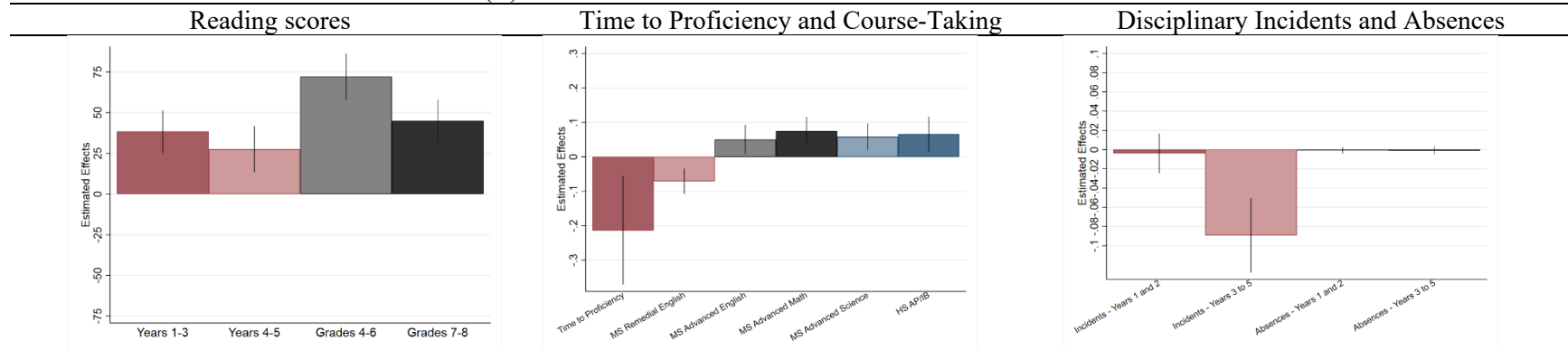
Reading scores	Time to Proficiency and Course-Taking	Disciplinary Incidents and Absences

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 2, and standard errors are clustered at the third grade reading score levels.

Appendix Figure 5. Sample Attrition Around the Retention Cutoff
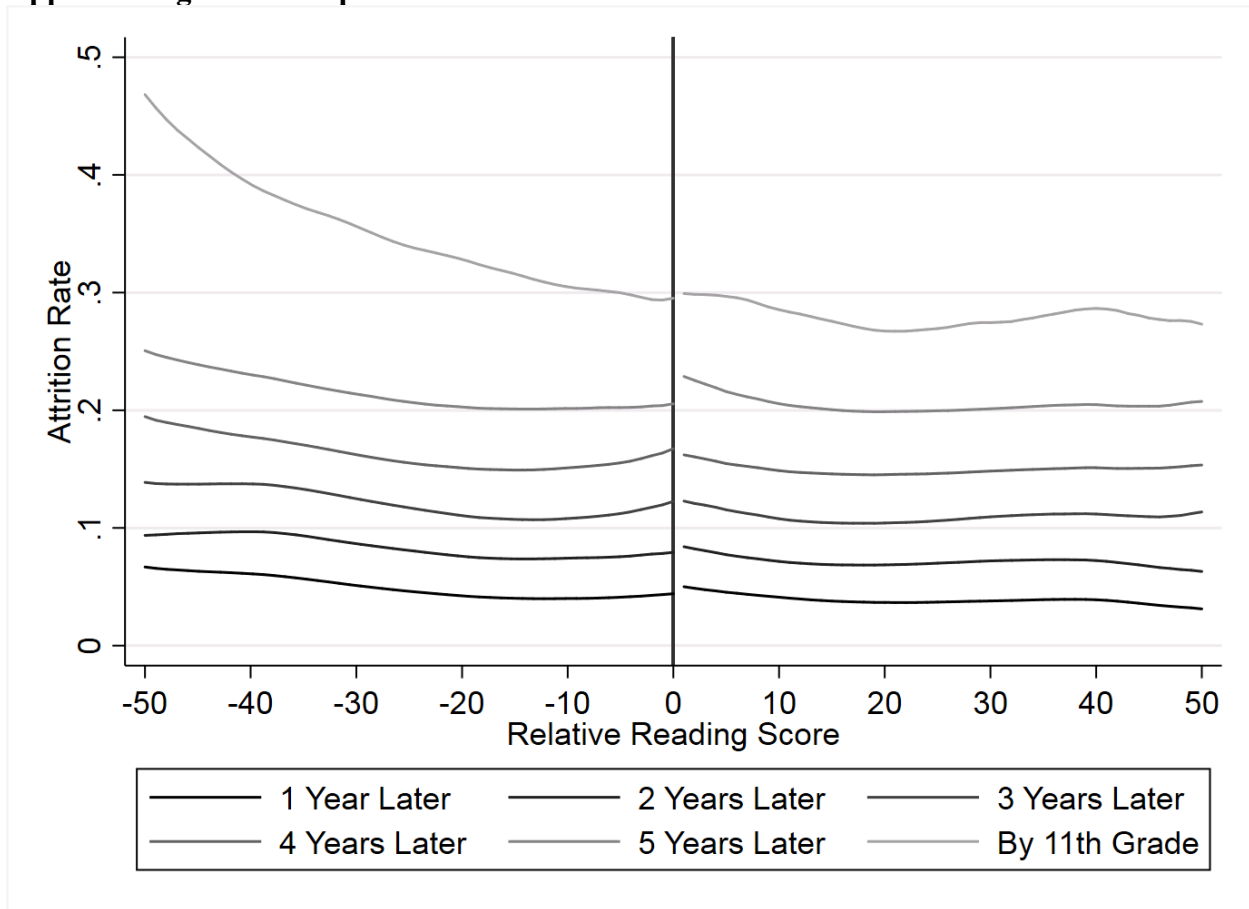
Notes: The figures present the local linear smoothing of the attrition rate in the years following the third grade on relative reading score of the student separately for the left of the cutoff date and the right.

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

David Figlio
Northwestern University and NBER

Umut Özek
American Institutes for Research

For online publication only

1. Validity of the Regression Discontinuity Design

In this section, we present further evidence on the validity of the regression discontinuity design used in our main analysis. In particular, in Online Appendix Figure 1, we examine the continuity of student baseline characteristics around the retention cutoff and plot 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 concerning discontinuity at the cutoff.

Online Appendix Figure 1. Relationship Between Reading Scores in Third Grade and English Learner Student Characteristics
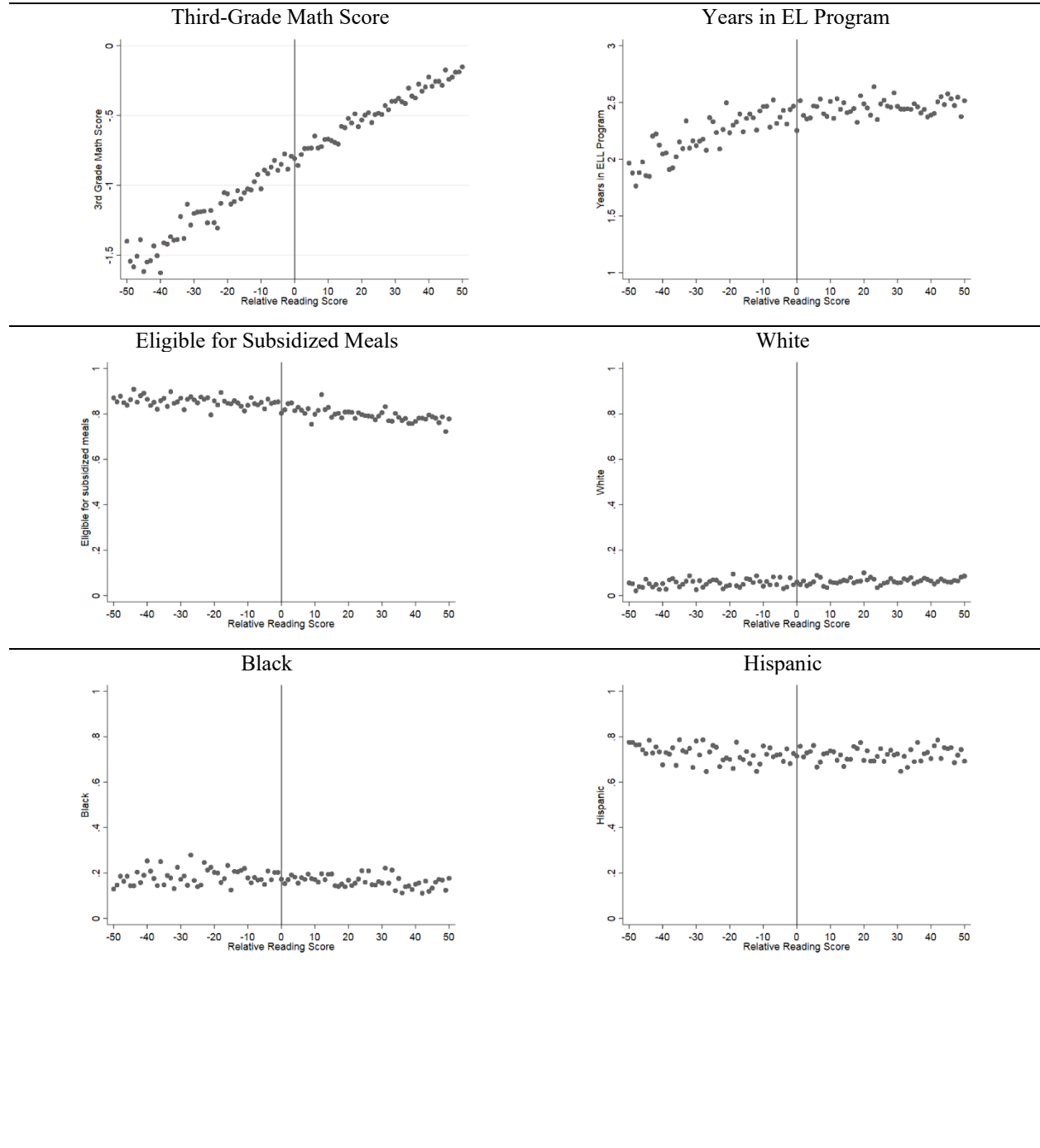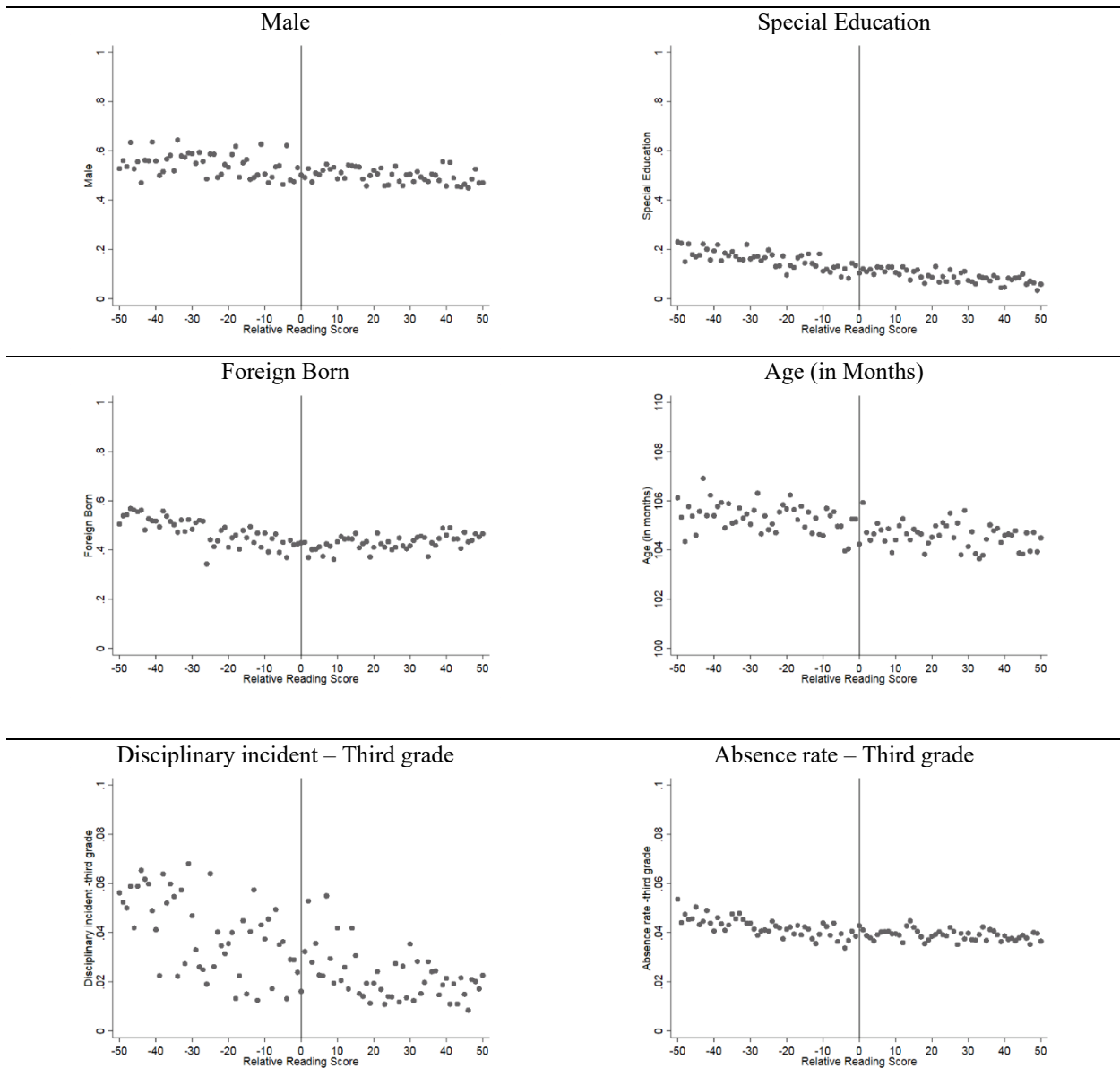
Notes: The figure presents the raw cell means of the given student attribute for each reading score between 50 points below and 50 points above the retention cutoff.

Online Appendix Figure 2 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.

Online Appendix Figure 2. Distribution of Third-Grade Reading Scores, English Learners
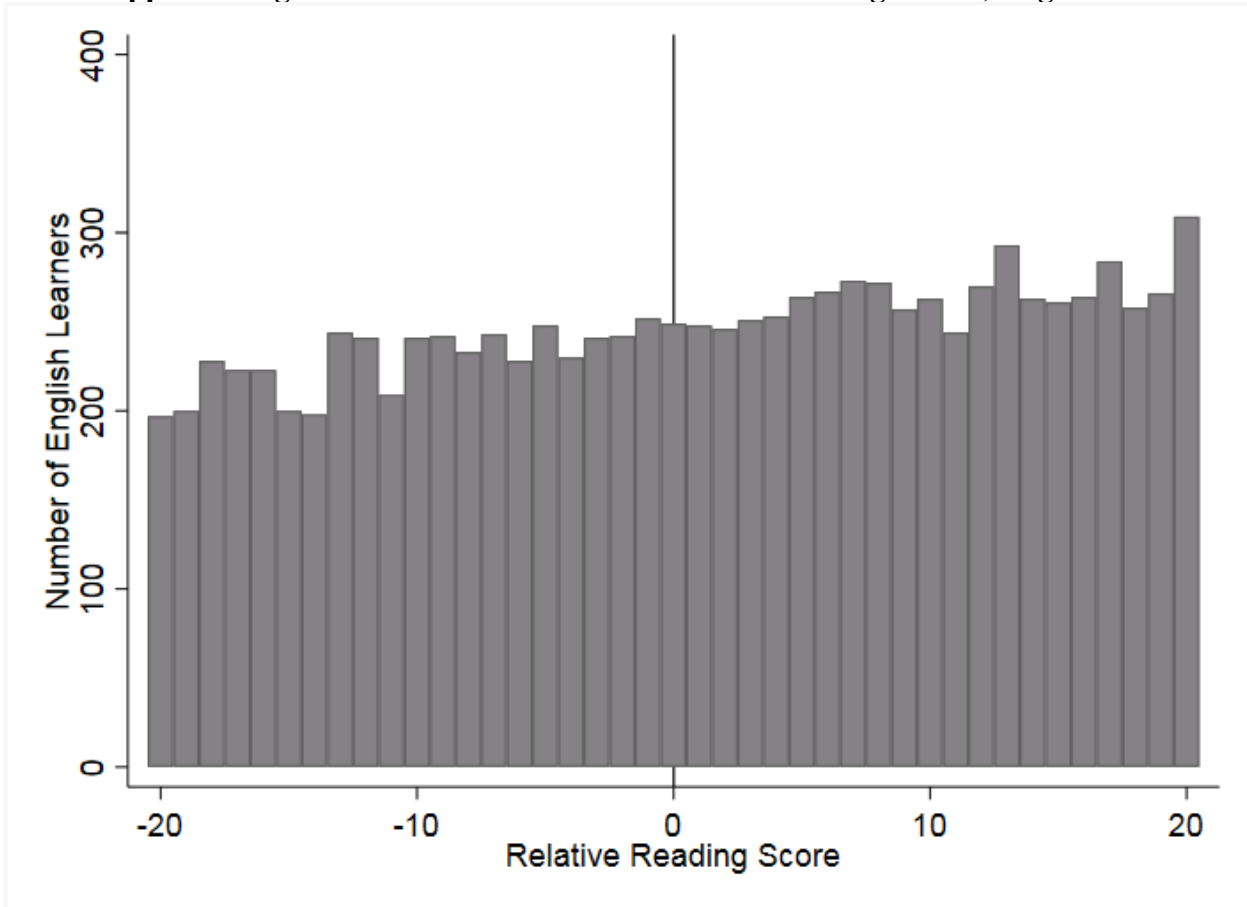
Notes: The figure presents the number of students in each reading score bin between 20 points below and above the retention cutoff, which is shown by the vertical line.

2. Effects of the Retention Policy away from the Cutoff

An important limitation of our RD design is that the estimated effects represent local average treatment effects and hence are only applicable to students around the retention cutoff. As such, it is hard to generalize our results to English learners with lower reading scores. To examine the broader effects of Florida's grade retention policy on low-performing English learners, we use the following difference-in-difference-in-differences (DDD) approach and the sample of first time third and fourth graders in 2000-01, 2001-02, and 2002-03 school years who score in the lowest two achievement levels in reading:

$$Y = \alpha + \beta_1(L1 * Post * G3) + \beta_2(L1 * G3) + \beta_3(Post * G3) + \beta_4(G3) + \beta_5(Post) +$$

$$\beta_6(L1) + \beta_7 X + \delta_s + \varepsilon. \tag{OA-1}$$

where $L1$ is an indicator for students who score in the lowest achievement level on the third grade reading test, $Post$ is an indicator for 2002-03 school year, $G3$ is an indicator for third graders, X is the vector of student characteristics listed in Table 2, and δ_s are school fixed effects. That is, we are comparing the lowest-scorers to the next-lowest-scorers, in third grade versus next grade, after the retention policy versus before the retention policy. In this setting, β_1 represents the estimated effects of the policy on all low-performing English learners, and we present these estimates in Online Appendix Table 1 for different outcomes of interest. This exercise reveals that the broader effects of the retention policy on low-performing English learners are smaller than the effects on the English learners around the cutoff, yet provides evidence that the retention policy has a positive effect on all low-performing English learners.

Online Appendix Table 1. Estimated Effects of Failing the Third Grade Reading Test, DDD Estimates

	DDD Estimates
Reading scores -	
Years 1 and 2	35.016***
	(10.281)
Effect size as % of SD	[0.121σ]
Years 3 and 4	5.852
	(10.345)
Effect size as % of SD	[0.021σ]
Grades 5 and 6	58.231***
	(10.931)
Effect size as % of SD	[0.207σ]
Grades 7 and 8	19.146**
	(9.338)
Effect size as % of SD	[0.079σ]
Time to proficiency	-0.171*
	(0.103)
Remedial English in middle school	-0.022
	(0.028)
Advanced English in middle school	0.051**
	(0.025)
Advanced Math in middle school	0.049*
	(0.026)
Advanced Science in middle school	0.063***
	(0.024)
AP/IB in high school	0.044*
	(0.024)

Notes: All regressions control for the baseline student characteristics listed in Table 2, and standard errors that are clustered at school level are given in parentheses. *, **, and *** represent statistical significance at 10, 5, and 1 percent, respectively. The numbers in brackets represent the estimated effect sizes as a percentage of the dependent variable standard deviation. The numbers in brackets represent the estimated effect sizes as a percentage of the dependent variable standard deviation.

3. Subgroup Analysis

Do certain English learner groups benefit more from the grade retention policy compared to others? While we do not have sufficient power to statistically distinguish the estimated effects across different English learner groups, Online Appendix Figures 3 to 6 present the results of an exploratory analysis providing the estimated effects of failing the third-grade test (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 the policy on reading scores are comparable for U.S. born and immigrant English learners, the longer-term benefits are larger for the latter group. 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 larger for native-born English learners for all other course-taking outcomes. Finally, Online Appendix Figure 5 reveals that the benefits of grade retention on disciplinary problems in middle school are more pronounced for foreign born English learners. For these students, we also find evidence that retention reduces absence rates in middle school.

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 the policy are slightly larger for the former group in many cases, with the exception of the effects on 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. The effects of the policy on disciplinary problems in middle school are also larger for Spanish-speaking English learners.

The effects of Florida's early grade retention policy could also 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 failing the third-grade test 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. We find mixed results. The effects of failing the third-grade reading test are larger for students who were not proficient in math or for students enrolled in higher-poverty elementary schools on future reading scores, remedial English course taking in middle school, and time to proficiency. Further, the reduction in disciplinary problems in middle school are more pronounced for these students. On the other hand, the effects on advanced course taking in middle school are larger for students with larger latent human capital or for English learners in lower-poverty elementary schools.

We further explore the heterogeneous effects of retention by socioeconomic status using an approach similar to Figlio et al. (2014), and construct an index of student socioeconomic status for Florida-born students by using FRPL eligibility, maternal education, mother's marital status and age at birth, and maternal immigrant status to predict student's third grade math

scores. Online Appendix Figure 7 presents the estimated effects of failing the third-grade test

separately for students above and below the median of this index for reading scores in panel (A),

time to proficiency and course-taking in panel (B), disciplinary incidents in panel (C), and

absences in panel (D). The effects on reading scores, disciplinary incidents, and absences are

comparable between lower- and higher-SES students, yet we find significant differences in time

to proficiency and high school course-taking favoring higher-SES English learners.

Online Appendix Figure 3. Estimated Effects of Failing the Third Grade Reading Test on Reading Scores, by Subgroup

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 2, and standard errors are clustered at the third-grade reading score levels. High (low) poverty schools are defined as those with the percentage of free or reduced priced lunch recipients higher (lower) than the school-level median.

Online Appendix Figure 4. Estimated Effects of Failing the Third Grade Reading Test on Time to Proficiency and Course-Taking, by Subgroup
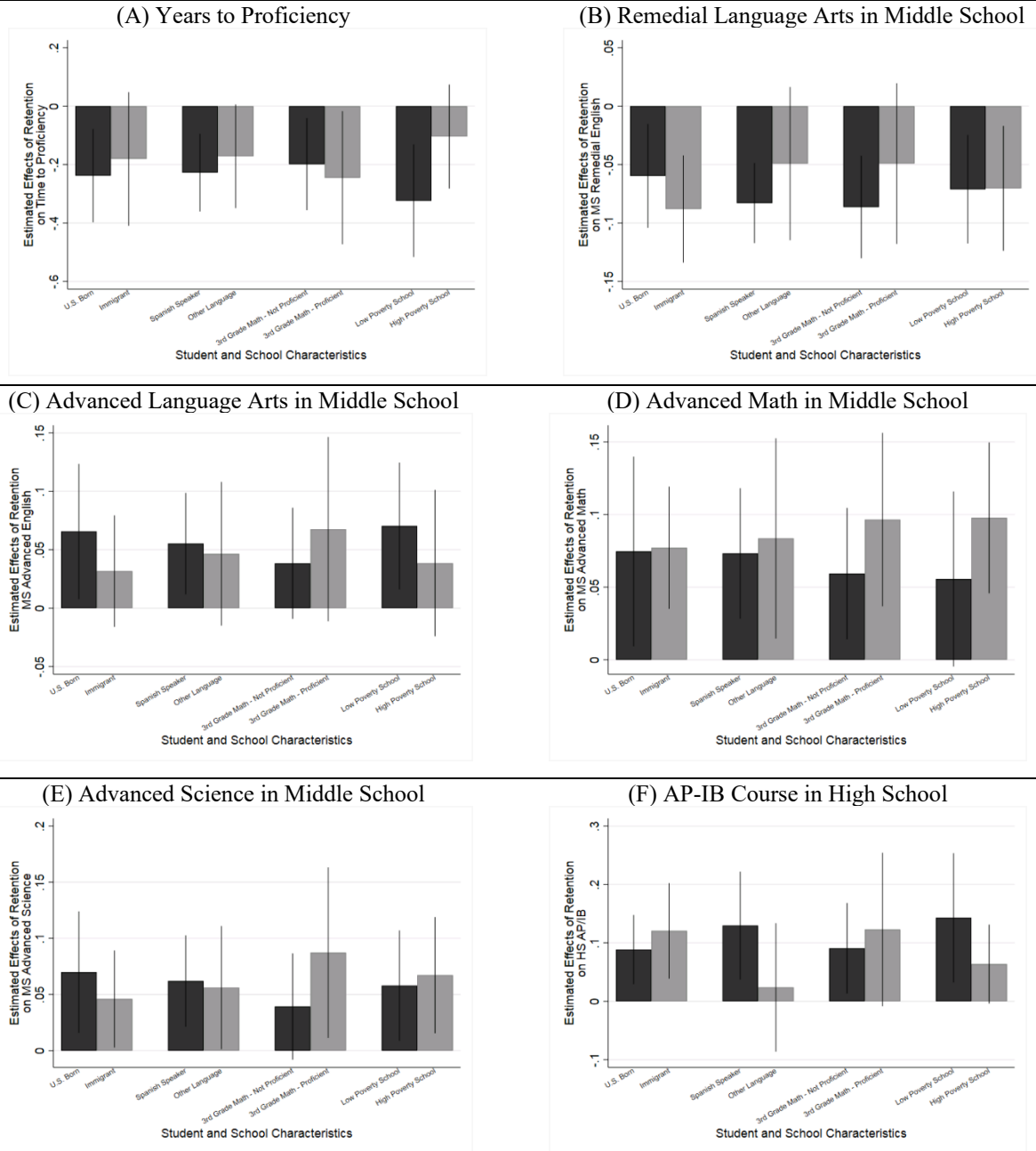
(A) Years to Proficiency

(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: Bars in each figure represent the effect of failing the third grade reading test 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 2, and standard errors are clustered at the third-grade reading score levels. High (low) poverty schools are defined as those with the percentage of free or reduced priced lunch recipients higher (lower) than the school-level median.

Online Appendix Figure 5. Estimated Effects of Failing the Third Grade Reading Test on Disciplinary Incidents and Absences, by Subgroup
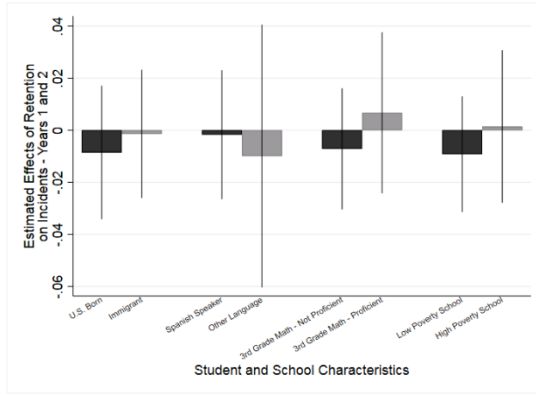
(A) Ever Involved in a Disciplinary Incident Years 1 and 2	(B) Ever Involved in a Disciplinary Incident Years 3 to 5
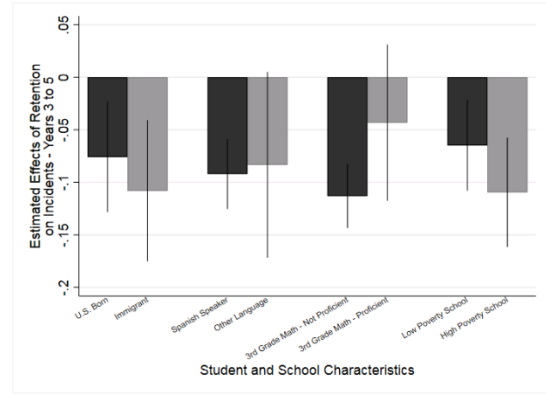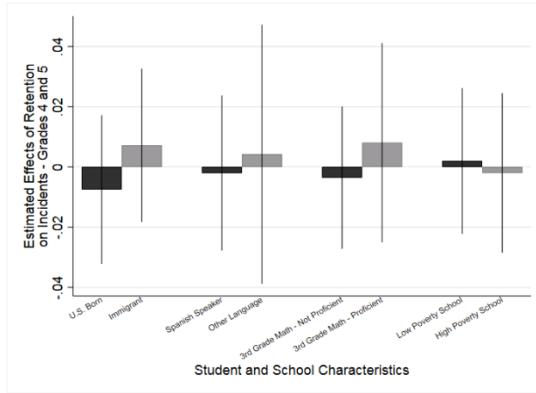
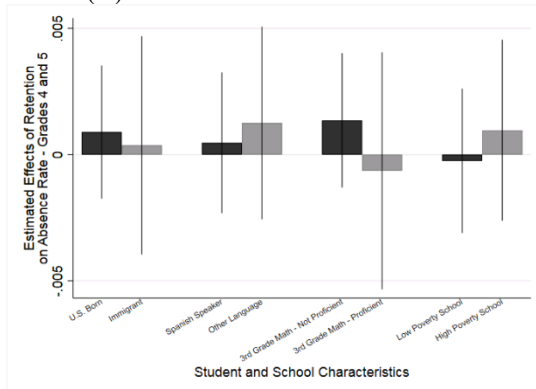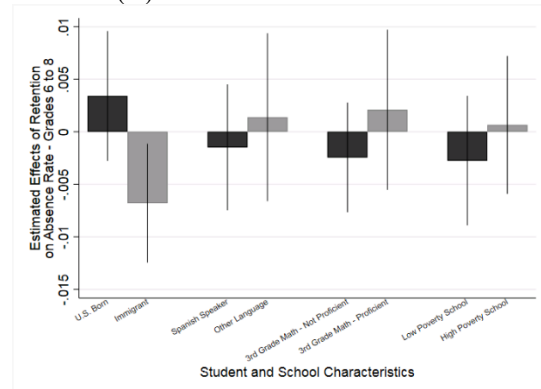| (C) Ever Involved in a Disciplinary Incident Grades 4 and 5	(D) Ever Involved in a Disciplinary Incident Grades 6 to 8
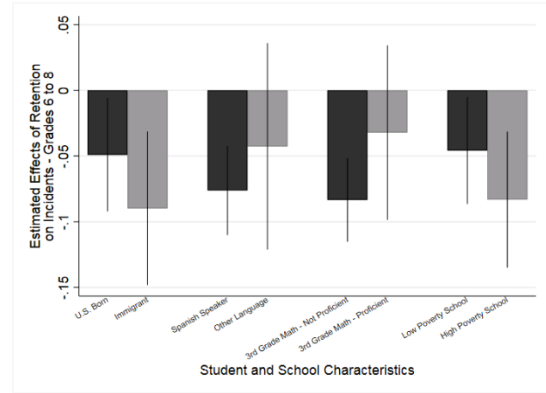
(E) Absence Rate – Years 1 and 2	(F) Absence Rate – Years 3 to 5
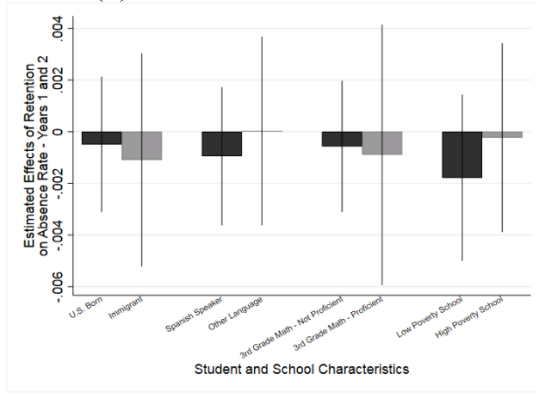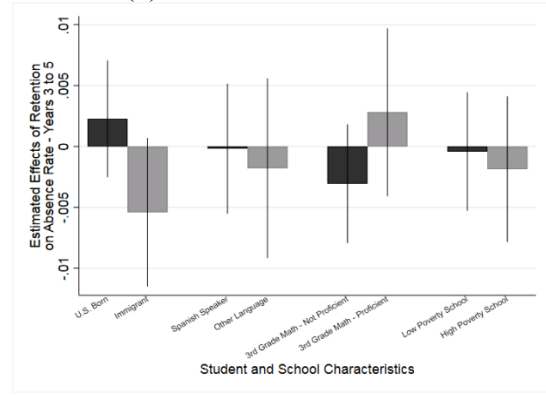
(G) Absence Rate – Grades 4 and 5

(H) Absence Rate – Grades 6 to 8

Notes: Bars in each figure represent the effect of failing the third grade reading test 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 2, and standard errors are clustered at the third-grade reading score levels. High (low) poverty schools are defined as those with the percentage of free or reduced priced lunch recipients higher (lower) than the school-level median.

Online Appendix Figure 6. Estimating Effects of Failing the Third Grade Reading Test on English Learners, by Socioeconomic Status, for Florida Born Students
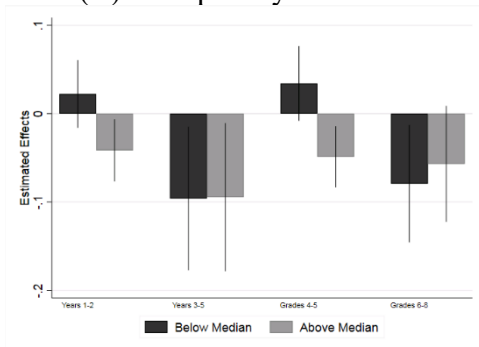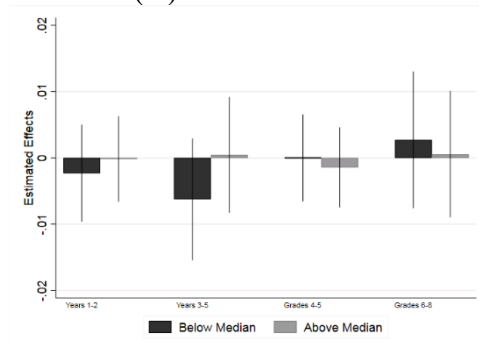

(A) Reading scores
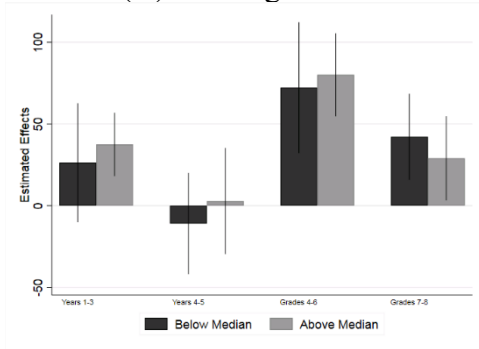
(B) Time to Proficiency and Course-Taking
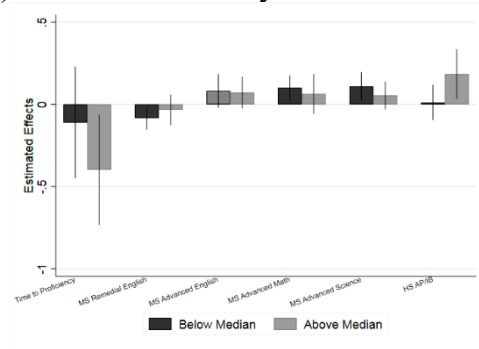
(C) Disciplinary incidents

(D) Absence rates

Notes: Bars in each figure represent the effect of failing the third grade reading test 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 2, and standard errors are clustered at the third-grade reading score level.

4. Instructional Support in the Following Year

In this section, we examine the extent to which retained English learners received instructional support during summer and in the following year along three dimensions: (1) enrollment in summer school; (2) reading and ELA instruction time per week in the following year; and (3) reading teacher experience in the following year. Summer school take up rate was high among retained English learners, with 67 percent enrolling in the summer program at the end of third grade. To examine (2) and (3), we use the sample of English learners who entered third-grade for the first time between 2006 and 2008, when the course enrollment data started breaking down third grade courses by subject. Using these data, we identify the time (in minutes) each student spent per week in reading courses and the experience level of their reading teacher in the following year.

Online Appendix Figure 7 presents the results, comparing the time spent in all English language arts (ELA) courses in panel (A), reading courses in panel (B), reading teacher experience in panel (C), and the likelihood of being assigned to a novice reading teacher in panel (D) in the following year around the retention cutoff. We do not observe any statistically and economically significant discontinuity in instructional time or teacher experience in the following year at the retention cutoff. While we are unable to completely rule out the role of instructional support on the observed benefits of the policy on English learners, these findings combined with the findings on the effects of summer school on exempt students in Table 8 provide strong evidence that repeating third grade is an important driver of these beneficial outcomes.

Online Appendix Figure 7. Estimated Effects of Failing Third Grade Reading Test on Reading Instruction Time and Teacher Experience in the Following Year
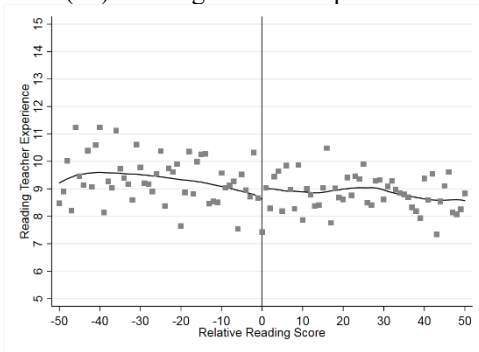

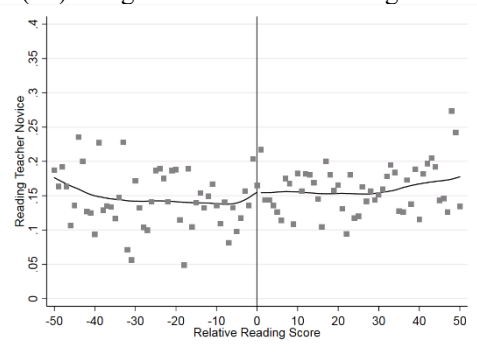(I) Minutes in ELA Courses per Week
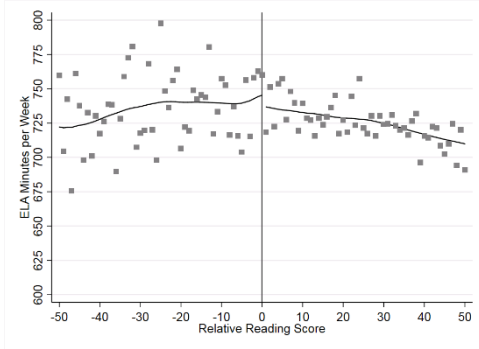
(II) Minutes in Reading Courses per Week
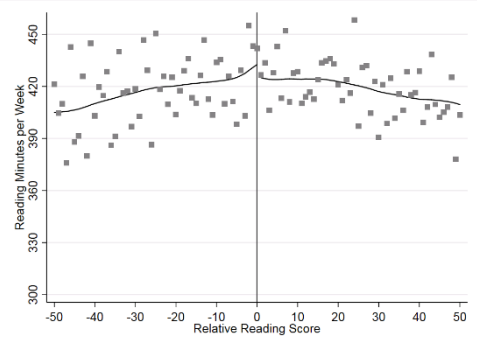
(III) Reading Teacher Experience

(IV) Assigned to a Novice Reading Teacher

Notes: The figures present the local linear smoothing of instructional time and teacher experience variables 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.