

The Effect of Bilingual Education: Long-Run Student Outcomes and Spillover Effects*

Kevin Hunt[†]

Hyunkyeong Lim[‡]

Congressional Budget Office

University of Wisconsin–Madison

August 9, 2026

Abstract

A large and growing share of students arrive at school as English language learners (ELL). In Texas, schools provide either a bilingual education or an English as a second language (ESL) program for these students. Using a regression discontinuity design based on a quasi-experimental policy variation in Texas, this paper examines the effect of exposure to bilingual education in 1st grade on students' test scores and college enrollment. Using the 1999–2008 first grade cohorts as our main sample, we find little effect of bilingual education on elementary school standardized test scores but positive effects on middle school test scores: math scores rise by 0.23 standard deviations and reading scores rise by 0.19 standard deviations. In the long run, bilingual education raises four-year university enrollment by 6.9 percentage points (a 46% increase) for all students, with the gains driven by a 6.6 percentage point (70%) increase for Spanish-speaking ELL students. The positive middle school test score effects for both ELL and non-ELL students indicate that the benefits of bilingual education for ELL students are not well captured by early test scores and generate meaningful spillovers to their non-ELL peers.

*The conclusions of this research do not necessarily reflect the opinions or official position of the Texas Education Agency, the Texas Higher Education Coordinating Board, the Texas Workforce Commission or the State of Texas.

[†]kghunt@wisc.edu

[‡]hyunkyeong.lim@wisc.edu

1 Introduction

A large and growing share of students in the United States arrive at school as English language learners (ELL), typically because they speak another language at home. In Fall 2020, 15% of public school first-graders were classified as ELL, and 76% of these students spoke Spanish as a first language (NCES (2023)). Large achievement gaps between ELL and non-ELL students make it crucial for policy-makers to understand how to best educate these students (NCES (2018)). These policies directly impact the outcomes of participating ELL students, and non-ELL students may also be affected by changing classroom compositions and peer effects. While previous papers have considered the causal effect of ELL education programs in elementary school on contemporaneous test scores (Chin et al. (2013), Pope (2016)), their effects on outcomes past elementary school, such as middle school test scores, high school graduation, and college enrollment, are not well understood.

This paper aims to fill the gap in the literature by estimating the causal effect of bilingual education compared to English as a Second Language (ESL) programs in Texas elementary schools on the medium- and long-run outcomes for both ELL and non-ELL students. Bilingual education programs place ELL students with the same home language into a separate classroom where the core curriculum is taught in their native language alongside English instruction. Conversely, ESL programs keep ELL and non-ELL students in the same classroom for core curriculum instruction in English, and ELL students learn English during a designated pull-out time. While ELL students may learn the core curriculum more effectively via their native language in a bilingual education program, this may also slow their English acquisition because of reduced exposure compared to ESL programs. Both ELL and non-ELL students may also be affected by the different composition of their classmates under each program.

Using administrative records from the Texas Education Research Center (ERC), we are able to observe which elementary schools provide bilingual education and link elementary school students to future education and unemployment insurance records. Following Chin et al. (2013), we leverage quasi-experimental policy variation in Texas for whether an elementary school offers a bilingual education or ESL program. Texas law requires school districts to offer bilingual education in at least one school for an elementary grade if there are more than 20 ELL students with the same home language enrolled in that grade. School districts near the cut-off are likely to be similar in unobservable characteristics but have a differential probability of providing bilingual education. By comparing the outcomes of students just above and below the cut-off in a fuzzy regression discontinuity design, we can causally estimate the effect of bilingual education compared to ESL.

To begin, we verify that the 20 student cut-off is a valid instrument for the provision of bilingual education in elementary school. We focus on the number of ELL students with Spanish as a home language enrolled in a district in first grade as the running variable and whether a school district provides a bilingual education program in first grade as the treatment variable. Additionally, we restrict the sample to smaller districts with less than 200 total students because the policy cut-off is more likely to bind for these districts; larger districts with more resources are likely to provide bilingual education programs even when they have relatively few ELL students. Although this means that our results are only for a selected sample of schools, they are still policy-relevant. Our sample districts cover a wide geographic range in Texas, and changing migration patterns have led to an increase in ELL students in smaller, more rural districts similar to our sample. Our main analysis uses first grade cohorts from academic years 1998–1999 to 2007–08 linked to outcomes through 2021. Within this main sample, districts with 20 or more Spanish-speaking ELL students in first grade are about 20 percentage points more likely to offer bilingual education in first grade, with a first-stage F -statistic of roughly around 10.

Utilizing the 20 student policy cut-off, we first examine how bilingual education exposure in 1st grade affects students’ elementary and middle school test scores. In elementary school, our IV estimates are close to zero and not statistically significant. In middle school, however, we find sizable positive effects that were not detectable in elementary school. For all students, bilingual education in first grade raises middle school math scores by 0.23 standard deviations ($p = 0.089$) and reading scores by 0.19 standard deviations ($p = 0.049$). These gains are present for both Spanish-speaking ELL students (0.24 SD in math, 0.20 SD in reading) and non-Spanish, non-ELL students (0.22 SD in math, 0.18 SD in reading). Consistent with previous work (e.g., [Bailey et al. \(2021\)](#)), test-score effects that are muted in the year of the intervention appear to grow with time.

Our rich administrative data also allows us to consider longer-run outcomes including high school graduation and college enrollment. In the long-run, we find large positive impacts of bilingual education on college-going. Provision of bilingual education in first grade raises the future probability that students enroll in a four-year public university in Texas by 6.9 percentage points (46%). This is true for both Spanish-speaking ELL first-graders, whose four-year enrollment increases by 6.6 percentage points (70%), and for non-ELL students, whose four-year enrollment increases by 6.2 percentage points (38%). The estimate for all students and for Spanish-speaking ELL students is statistically significant at the 10% level. The existence of a positive effect for both ELL and non-ELL students suggests that bilingual education for ELL students generates positive spillovers to their non-ELL peers that persist into post-secondary education. We do not find any evidence of an effect of

bilingual education on high school graduation rates or on community college enrollment. These results are consistent with previous work that finds early-life interventions can have large long-run effects on student outcomes in adulthood ([Chetty et al. \(2014\)](#), [Chetty et al. \(2011\)](#), [Jackson \(2018\)](#), [Petek and Pope \(2016\)](#)).

Mechanisms that could generate delayed impacts of bilingual education on ELL students include improvements in students' beliefs about their own abilities or in their expectations of the return to education, through interaction with teachers who share their linguistic and cultural background. Additionally, bilingual education may provide skills and knowledge that are not well measured by early standardized tests but improve performance in later grades and post-secondary outcomes.

Our research contributes to two strands of literature. First, our project contributes to the literature on the causal impact of education programs for ELL students by considering the future impacts of bilingual education on outcomes such as middle and high school test scores, high school graduation, and college enrollment. Previous research has focused on short-run outcomes such as elementary school test scores and found mixed results. Studies show that ELL students perform worse when enrolled in bilingual education ([Pope \(2016\)](#)), that bilingual education has negligible impacts on ELL student outcomes ([Matsudaira \(2005\)](#); [Chin et al. \(2013\)](#)), or that bilingual education improves ELL student outcomes ([Collier and Thomas \(2017\)](#)). [Chin et al. \(2013\)](#) also highlights the importance of considering the impact of ELL student education programs on non-ELL students, finding that non-ELL students perform better on standardized tests in elementary school when ELL students are taught in separate classrooms. Our results extend this finding by showing that these spillovers to non-ELL students grow with time and persist into college enrollment.

Second, our research contributes to the literature on the impacts of early-life interventions such as kindergarten and first grade teachers. This literature has highlighted the importance of considering long-run outcomes to evaluate education programs that happen at young ages. For example, the literature on the impact of pre-school enrollment has often found that test-score gains from pre-school enrollment do not persist into future grades, but that these benefits reappear in future outcomes such as high school graduation, college enrollment, and labor income ([Deming \(2009\)](#); [Pages et al. \(2020\)](#); [Bailey et al. \(2021\)](#); [Gray-Lobe et al. \(2021\)](#)). Many programs for ELL students are targeted at earlier grades because ELL student numbers are largest in earlier grades. Our research shows the importance of considering long-run outcomes for evaluating ELL student programs. While we find little effect of first-grade bilingual education on early elementary test scores, students who attended a first-grade district that provided bilingual education have higher middle school test scores and higher four-year college enrollment rates.

2 Policy Background

The education of ELL students and use of non-English languages in public schools have been the subject of legal debate in the United States dating back to the 19th century.¹ Schools must strike a delicate balance between providing an adequate education for ELL students and not segregating these students from their non-ELL peers. While courts and federal policy have avoided specifying an exact education model, schools are required to provide a curriculum for ELL students that both teaches English as well as general academic content.

Schools’ instruction of ELL students typically falls into one of two categories. English as a second language (ESL) is the most common form, accounting for the education of 77.6% of ELL public school students in 2019–20 ([Williams and Zabala \(2023\)](#)). In these programs, ELL students and non-ELL students are taught in the same classroom via English for the core curriculum, and teachers for these classrooms do not necessarily have any additional skills beyond standard requirements. During a pull-out time, ELL students are taught English by a certified ESL teacher ([TEA, 2022](#)). In addition to standard teacher requirements, this certification in Texas requires 120–150 hours of coursework as well as passing an exam on topics including basic linguistics and ESL teaching methodologies. ESL certifications need to be renewed each five years, and renewal requires at least 150 hours of professional development ([Anantha \(2024\)](#)). ELL students enrolled in ESL will be exposed to a substantial amount of English during the core curriculum as well as with their ESL teacher, who may not even speak their home language. This may improve the speed at which they learn English but also hamper their understanding of material in the core curriculum. Non-ELL students will also interact with ELL students during the core curriculum, and this may impact how the core curriculum is taught.

In contrast to ESL, a bilingual education program separates ELL students into a different classroom taught by a certified bilingual educator; the core curriculum for ELL students is taught in their native language while students also learn English. Programs may either be classified as early-exit, meaning the students are expected to be reclassified as non-ESL in 2–5 years, or late-exit, where reclassification is expected in 6–7 years ([TEA, 2022](#)). In addition to standard teaching requirements, bilingual education teachers in Texas must pass exams on teaching methodologies as well as proficiency in the language of instruction. After passing exams, a one-year internship of teaching with a field supervisor must be completed for full certification ([TTOT, 2023](#)). Hiring and retaining a bilingual certified teacher is the main hurdle for schools to implement a bilingual education program. While the number of ELL students in Texas has grown dramatically, the number of bilingual certified educators

¹See [Wright \(2010\)](#) for a summary of court rulings and laws affecting ELL student education in the USA.

has actually declined (Lopez (2022)). Relative to ESL, bilingual education exposes ELL students to less English; this may slow their English acquisition but also improve their understanding of the core curriculum. Non-ELL students will also interact less with ELL students, and teachers in the standard classroom may adjust their teaching in response to changing classroom composition.

Texas state law has several stipulations about when a school district must provide bilingual education.² In elementary school (grades 1 to 5), a school district with 20 or more ELL students with the same home language must provide a bilingual education program in that grade. School districts with less than 20 ELL students with the same home language are free to choose between providing a bilingual education or ESL program. Not all school districts above the 20 student threshold are able to provide bilingual education, largely because they do not have a bilingual education certified teacher. These districts must apply for and receive a waiver from the state board of Texas to not face sanctions. As part of the waiver, these districts must outline their plan for providing bilingual education in the future (for example, by hiring or training a bilingual teacher). After elementary school, districts can choose to provide either a bilingual education or ESL program regardless of the number of ELL students. The requirement that elementary school districts with 20 or more ELL students in a grade-home language cell must provide a bilingual education program forms the basis of our empirical strategy discussed in section 3.

In Texas, students are assigned an ELL status based on their home language and an English proficiency test.³ Students who report English as their only home language are classified as English proficient. If a student speaks a language other than English at home, they are required to take an English proficiency test specific to their grade level. The Texas Education Agency determines the passing threshold for this test, and students who score below the threshold are classified as ELL. If the ELL student is enrolled in a district that provides bilingual education, the student will be offered enrollment in the bilingual education course. Parents, though, may request that their child is not enrolled in the bilingual education program. Students whose parents request to not be enrolled in bilingual education or who are enrolled in districts that do not provide bilingual education are offered enrollment in the ESL program. Again, parents can request that their child does not participate in the ESL program. Students lose their ELL status once they are able to pass the English proficiency test, which is administered each year to all ELL students. After passing the proficiency test,

²Detailed instructions for requirements for education of ELL students can be found on the Texas Education Agency's website (<https://tea.texas.gov/academics/special-student-populations/english-learner-support/bilingual-education-exception-and-esl-waiver-resources>).

³A detailed flow chart for ELL student classification and progression can be found at <https://tea.texas.gov/sites/default/files/eb-el-decision-chart-for-lpac-accessible-version.pdf>.

students who were previously classified as ELL are monitored for an additional two years to confirm English proficiency.

3 Empirical Strategy

To estimate the effect of exposure to bilingual education on student outcomes, we need to compare students who did and did not attend elementary school at a district that provided bilingual education. This exposure to bilingual education is not random. For example, school districts may decide to provide bilingual education if their ELL students are performing particularly poorly in a standard classroom setting. In this scenario, we may estimate that ELL students do worse when a bilingual education program is provided, while in fact bilingual education is provided because these ELL students are doing poorly. Moreover, school districts with larger numbers of ELL students may be more likely to provide bilingual education, and these school districts are perhaps different than school districts that do not provide bilingual education (e.g. larger, more urban, have more economically disadvantaged students). If the likelihood of bilingual education provision increases with the number of ELL students, this can also coincide with negative peer effects from academically less-prepared economically disadvantaged students, which can create negative bias to the causal effect of bilingual education.

Thus, we leverage policy variation in Texas that generates quasi-random exposure to bilingual education in elementary school to measure the causal impact of bilingual education. As discussed in section 2, Texas requires that school districts with 20 or more ELL students in an elementary grade (1–5) with the same home language must provide bilingual education. School district grades with 19 and 20 ELL students are likely to be very similar on average; the only difference will be that the school district is required to provide bilingual education if there are 20 ELL students. By comparing school districts just above and just below the cut-off, we compare similar sets of students except for the exposure to bilingual education. Any differences in outcomes between these two sets of students is likely to be caused by the bilingual education program. This policy variation suggests a regression discontinuity design as discussed in [Imbens and Lemieux \(2008\)](#) and [Lee and Lemieux \(2010\)](#).

Previous research by [Chin et al. \(2013\)](#) leverages this 20 ELL student cut-off in Texas to estimate the impact of bilingual education programs on elementary school test scores. We extend this strategy to document the impact of bilingual education on non-test-score and longer-run outcomes including middle school test scores, high school graduation, and college enrollment. We estimate the effect of bilingual education at the individual level as our main interest is to measure the impact of bilingual education on each *student's* longer-

run outcomes. The goal is to understand how bilingual education helps or hurts students in developing language and academic skills.

Under Texas law, the 20 ELL student cut-off is not always strictly binding as discussed in section 2. Thus, we use a fuzzy regression discontinuity design where bilingual education provision is instrumented by the 20 ELL student cutoff. In order for the policy variation to be a valid instrument, the 20 ELL student threshold must be a strong predictor of bilingual education programs. We first verify that the cutoff is a valid instrument through the following first-stage regression:

$$BE_{idc} = \delta^{FS} \mathbf{1}\{\#ELL_{dc} \geq 20\} + f^{FS}(\#ELL_{dc}) + X'_{dc}\gamma^{FS} + \epsilon_{idc}^{FS}. \quad (1)$$

BE_{idc} is an indicator for whether school district d of individual i provided bilingual education for cohort c in the first grade. BE_{idc} takes 1 if there was any student who received bilingual education in school district d in the first grade for cohort c based on each individual's bilingual education status. $\#ELL_{dc}$ is the number of ELL students in this district-cohort where f^{FS} is a continuous function of $\#ELL_{dc}$.⁴ We additionally control for demographic characteristics of the first grade cohort c in school district d to alleviate the concern of ELL students sorting into school districts with certain demographic composition.

In addition to being a strong predictor of bilingual education provision, the instrument (i.e., the number of ELL students being above the threshold) must not be correlated with other variables that impact student achievement. In the regression discontinuity framework, it is also important to check for bunching below the threshold, which would suggest that schools actively try to reduce ELL counts to avoid providing bilingual education. We show below that the density of school districts is smooth below and above the 20 ELL student cutoff. Finally, we only use schools with 8–39 ELL students in the first grade cohort to accurately measure the local average treatment effect of bilingual education following the previous work by [Chin et al. \(2013\)](#).

After confirming that the cut-off is a valid instrument for bilingual education provision, we compare student outcomes at school districts just above and just below the threshold using the following regression:

$$y_{idc} = \delta^{SS} \mathbf{1}\{\#ELL_{dc} \geq 20\} + f^{SS}(\#ELL_{dc}) + X'_{dc}\gamma^{SS} + \epsilon_{idc}^{SS}. \quad (2)$$

Here, δ^{SS} captures the effect of a change in student outcomes y above the policy threshold. As mentioned previously, students just below and above this threshold will be very similar

⁴We use a linear function of $\#ELL_{dc}$ where the slopes are allowed to be different below and above the 20 ELL cutoff for f^{FS} .

on average except that students above the threshold are much more likely to have been at a school district that provided bilingual education. Thus, δ^{SS} provides evidence of the causal impact of bilingual education on student outcomes. Furthermore, re-scaling δ^{SS} by the first-stage coefficient will give an estimate that can be interpreted as the causal effect of bilingual education provision.

4 Data

Our project leverages detailed longitudinal administrative data from the Texas Education Research Center (Texas ERC) about Texas K–12 students. The sample contains Texas public school students enrolled for the academic years 1993–94 to 2020–21.

The data contains information about a student’s home language and ELL status. Therefore, we are able to see which students are eligible for additional help from schools based on their English proficiency.⁵ We focus on ELL students whose home language is Spanish. Additionally, the data identifies which program the ELL students are enrolled in (bilingual education or ESL). Using this enrollment information, we can classify school districts as providing bilingual education or not.⁶ The longitudinal structure of the data allows us to analyze a rich set of student outcomes.

First, we have standardized test scores in math and reading for students in grades 3 through 8. Our sample cohorts entered first grade between 1998–99 and 2007–08, and their elementary- and middle-school testing years span three Texas assessment regimes: the Texas Assessment of Academic Skills (TAAS, in place through spring 2002), the Texas Assessment of Knowledge and Skills (TAKS, spring 2003 through spring 2011 for grades 3–8), and the State of Texas Assessments of Academic Readiness (STAAR, spring 2012 onward). Because Texas replaced the primary regime for grades 3–8 at clean breaks (spring 2003 and spring 2012), no student in our sample is tested under two regimes in the same grade-year. For each student-grade-year we use the score from the primary spring administration; if a student appears with more than one score (for example, from a retake), we take the earliest one. To make raw scale scores comparable across regimes and time, we standardize each score to a z-score within its year $\times$ grade $\times$ subject $\times$ regime cell using only the primary-month administration.

We do not use high school test scores. Beginning with 9th graders in spring 2012, TAKS was replaced by subject-specific STAAR end-of-course (EOC) exams (Algebra I, English I,

⁵We do not observe the test scores that were used to assign students as ELL or not.

⁶We do not observe directly whether a school district provides bilingual education or not. We identify a school district as providing bilingual education if there is any student who is enrolled in bilingual education.

Biology, and so on) that students take at different grade levels depending on their course sequences, so EOC scores are not comparable across cohorts or to the grade-based math and reading scores. Instead, we capture high school achievement through the graduation outcome. Because a Texas high school diploma required passing the state exit exam during our sample period, HS graduation implicitly captures whether a student met the state’s high-school-level achievement standard. A limitation is that when we see a student who did not graduate because they left the high school, we cannot disentangle whether he or she did so because they dropped out or because they transferred to a school outside the Texas public system. Because Texas is large and most students who begin schooling in Texas public schools also finish there, we are able to link a large fraction of graduates back to their first-grade district and BE exposure. Furthermore, if we assume that the bilingual education program has little or no influence on the decision to transfer to schools outside the Texas public school system, then any high school graduation differences can be attributed to changing drop out rates.

Third, we have college enrollment information for Texas public two- and four-year institutions, linked to individual student records through the Texas Higher Education Coordinating Board (THECB) post-secondary enrollment files. Similar to high school graduation, many Texas students enroll in Texas public colleges for their post-secondary education, so we are able to match a large number of college students to their first-grade district and BE exposure. If we assume that BE exposure does not influence the decision to enroll in colleges outside the Texas public system, then we can estimate the impact of BE on college enrollment decisions in Texas. Because our data end with the 2020–21 academic year, the youngest cohort in our sample (first graders in 2007–08) can be observed at most 13 years after first grade. We therefore restrict our college-enrollment outcome to a common 13-year post-first-grade observation window for every cohort, so that estimates are not driven by longer follow-up for older cohorts. We do not report college graduation outcomes: 13 years is too short to reliably observe two-year associate or four-year bachelor’s degrees for the youngest cohorts, and reporting graduation only for older cohorts would reintroduce the cohort-selection bias that the common enrollment window is designed to avoid. Degree-completion effects will become estimable as additional years of data accrue.

Finally, the data contain individual labor income for Texas workers covered by unemployment insurance. We do not report labor-income results here because even our oldest cohort would be only in their late 20s at the end of our sample window, before career earnings stabilize. We expect to include labor-income outcomes as additional cohorts age.

Following [Chin et al. \(2013\)](#), we focus our analysis on first-graders whose district had between 8 and 39 ELL students because the regression discontinuity is identified in the

neighborhood of the 20 student cut-off. Additionally, we drop school districts with more than 200 total students in the first grade cohort in 2004–2005 academic year; larger school districts have more financial resources and are more likely to provide a bilingual education program regardless of whether or not they are above the cut-off. Table 1 summarizes district characteristics in the 2004–05 first grade cohort by district size, and Table 2 summarizes the analysis sample of small districts within the 8–39 Spanish LEP bandwidth by cohort window. Restricting our sample to school districts with relatively small numbers of ELL students and small total number of students is a limitation of our empirical strategy as these school districts are substantially different than the average school district. Despite this limitation, the clean identification of bilingual education vs. ESL yields interesting insights.

Our main sample uses first grade cohorts entering in the 1998–1999 through 2007–08 academic years (hereafter, “1999–2008” cohorts). We use the earlier 1994–95 through 1997–98 cohorts (“1994–1998”) and the more recent 2008–09 through 2018–19 cohorts (“2009–2019”) only in the first-stage validity checks discussed below, where they illustrate how compliance with the 20-student policy has evolved over time. All main-analysis test-score and long-run estimates use the 1999–2008 sample.

5 Results

5.1 Validity of the Empirical Strategy

As discussed in section 3, our main empirical setting is the policy variation in Texas that requires school districts with 20 or more ELL students in an elementary grade (1–5) with the same home language to provide bilingual education. However, the 20 ELL student cutoff is not always strictly enforced, and some schools receive waivers from the requirement to provide a separate bilingual classroom (see section 2). Thus, for the policy variation to be a valid instrument, there must be a meaningful jump in the likelihood of providing bilingual education at the 20 ELL student cutoff.

Discontinuity Figure 2 illustrates the fraction of districts which provide bilingual education as a function of the number of Spanish ELL first graders in the district for our 1999–2008 main sample. Visually, it confirms that there is a discontinuous jump in the probability at the 20 ELL student cutoff. The relationship between the probability of providing bilingual education and the number of Spanish ELL students in the district becomes noisier above the cutoff. To be more statistically precise, we run equation 1 and check whether the probability of bilingual education provision changes statistically significantly at the 20 ELL student cutoff. Table 3 presents the results. For our 1999–2008 main sample, the estimated jump in

the probability of providing bilingual education at the cutoff is approximately 20 percentage points, with a first-stage F -statistic of about 14, confirming that the policy variation provides us with a valid empirical setting to evaluate the effect of bilingual education. The table also reports first-stage estimates for the 1994–1998 and 2009–2019 cohort windows for context; these estimates are used only to characterize the evolution of the policy environment and are not used in our main test-score or long-run analyses.

Bunching and Balance Test We do not see any bunching of the districts or school campuses around the 20 ELL student cutoff. A potential threat to the empirical strategy is that districts might strategically adjust the number of ELL students so they may or may not have to provide bilingual education. However, Figure 3 shows no evidence of such strategic behavior at either the district or school campus level. The distribution of districts is smooth along the number of ELL students in each district.

Moreover, we test whether predetermined student characteristics are smooth across the cutoff — a necessary condition for the fuzzy RD assumption that students above and below the cutoff are otherwise comparable. We run Equation 2 with student characteristic variables on the left-hand side, examining students’ sex, ethnicity, free-or-reduced-lunch status, special education status, and gifted status, all measured in first grade. Table 4–Table 6 show that student characteristics are largely balanced across the cutoff. Among non-Spanish non-ELL students (Table 6), the female share is 1.0 percentage points higher above the cutoff, significant at the 10% level but small relative to the sample mean of 47.9%. To check for changes in student composition at the school level, we collapse the data to the school-cohort level and re-estimate Equation 2 on school-cohort shares of female, ethnic, free-or-reduced-lunch, and special education students. Table 7 shows two coefficients significant at the 10% level: the female share is 1.2 percentage points higher and the special-education share is 2.0 percentage points lower above the cutoff. Both magnitudes are modest. For example, the special-education difference is a 2 percentage-point change on a base of 10.7%. Moreover, all three characteristics enter the fuzzy RD specification as controls, so they should not bias our estimates. Overall, the balance evidence is consistent with the assumptions of the fuzzy RD design.

Discussion Compliance with the 20-student policy has changed over time. Figure 1 plots the point estimate of δ^{FS} in equation 1 by first grade cohort along with its 90% confidence interval. The jump in the probability of providing bilingual education at the cutoff is around 0.20–0.30 for the 1994–1998 and 1999–2008 cohorts, but falls sharply for the 2009–2019 cohorts to below 0.10 on average, and in many years the point estimate is not statistically distinguishable from zero. In some years the point estimate is even below zero, which would

imply that a district is less likely to provide bilingual education when it has 20 or more Spanish-speaking ELL students. For this reason, our 2009–2019 estimates are noisier and should be interpreted with caution: the first-stage F -statistic pooled over these cohorts falls well below conventional weak-instrument thresholds. Our main analysis therefore focuses on the 1999–2008 cohorts, where the instrument is strong, first-grade students can be tracked through college, and outcomes are observed cleanly.

A potential explanation for the decrease in compliance with the policy lies in funding deficiency. In Texas, the source of state funding for all public school districts is the Texas Foundation School Program (FSP). FSP allocates the state funding primarily based on the average daily attendance of students. Moreover, FSP adjusts the allocation based on the district and student characteristics to meet the needs of districts. For instance, having more ELL students will allocate more funding to the district. However, [Rolle and Jimenez-Castellanos \(2014\)](#) concludes that the FSP formula was not enough to balance out another big source of revenue to the school districts: the revenue generated from local property values. According to their analyses, the districts with the highest percentage of ELL students generated on average \$1,300 less per pupil than districts with the lowest percentage of ELL students⁷. In fact, Texas recently passed House Bill 3 (HB 3) by the 86th Texas Legislature in 2019, in which the State upward-adjusted the weight for ELL students in the FSP formula—originally set in 1984 and never changed—to better fund bilingual education.

5.2 Test Scores

In this section, we examine the effect of exposure to bilingual education in the first grade on elementary and middle school test scores using the 1999–2008 main sample. Tables 11–13 report the reduced form and IV effects of bilingual education exposure on test scores, along with the first-stage results for IV. Throughout Section 5, we focus on the IV estimates (δ^{SS}), as they are the most relevant from the policy-maker’s perspective. Different columns correspond to different schooling levels (elementary or middle school) and different subjects. Tables 12–13 dissect the sample into different groups of students (Spanish-speaking ELL and non-Spanish, non-ELL) to see if there are heterogeneous effects across the groups. In a subsection below, we compare our test score estimates to [Chin et al. \(2013\)](#) and discuss how sensitive elementary versus middle school estimates are to the cohort window chosen.

In elementary school, we do not find any statistically significant effect of bilingual education on test scores for our main 1999–2008 sample. Based on the IV estimates, bilingual education on average changes elementary math test scores by -0.01 standard deviations

⁷The total difference amounts to approximately \$5.9 million, which translates to funding the salaries and benefits for 90 Texas teachers annually ([Rolle and Jimenez-Castellanos \(2014\)](#)).

and elementary reading test scores by 0.05 standard deviations. Both effects are small and statistically indistinguishable from zero. The estimates for Spanish-speaking ELL students (-0.03 SD in math, -0.01 SD in reading) and non-Spanish, non-ELL students (-0.01 SD in math, 0.05 SD in reading) are similarly small and statistically insignificant at the 10% level.

In middle school, however, the estimated effects of first-grade bilingual education become substantially larger and, for several groups, statistically significant. For all students, bilingual education raises middle school math scores by 0.23 standard deviations ($p = 0.089$) and reading scores by 0.19 standard deviations ($p = 0.049$). Broken down by subgroup, the effect on Spanish-speaking ELL students is 0.24 SD in math ($p = 0.077$) and 0.20 SD in reading ($p = 0.072$), while the effect on non-Spanish, non-ELL students is 0.22 SD in math ($p = 0.129$) and 0.18 SD in reading ($p = 0.083$). Middle school gains are therefore roughly of the same magnitude across the two groups, though the point estimate is slightly larger for Spanish-speaking ELL students. The fact that gains are muted in elementary school but appear in middle school is consistent with previous work finding that some early-education interventions have delayed rather than immediate effects on measured achievement (e.g., [Bailey et al. \(2021\)](#)).

5.2.1 Comparison to [Chin et al. \(2013\)](#)

Because our main sample and empirical implementation differ from [Chin et al. \(2013\)](#), it is useful to explicitly compare our test score results to theirs. [Chin et al. \(2013\)](#) pool 3rd through 5th grade test scores from the 2002–03 through 2009–10 school years using publicly available school-grade-year cell data. This outcome window corresponds to first-grade cohorts entering school in 1998–99 through 2007–08, which is essentially the same as our main 1999–2008 sample. Under this same cohort window (i.e. [Chin et al. \(2013\)](#)’s implied first grade cohorts) our IV estimates for elementary test scores (Tables 11–13, elementary columns) are close to zero and statistically insignificant, in contrast to [Chin et al. \(2013\)](#)’s reported spillover effect of roughly 0.2 SD on non-ELL students.

When we shift the cohort window forward to 2003–2010 first grade cohorts, however, our elementary estimates line up much more closely with [Chin et al. \(2013\)](#)’s baseline 2SLS results. Table 14 presents this side-by-side comparison. For all students, we estimate that bilingual education raises elementary math scores by 0.19 SD and reading scores by 0.16 SD, compared to [Chin et al. \(2013\)](#)’s 0.21 SD and 0.14 SD. For non-Spanish, non-ELL students, we estimate 0.22 SD in math ($p = 0.083$) and 0.18 SD in reading ($p = 0.094$), very close to their 0.21 SD and 0.17 SD. Our estimates for Spanish-speaking ELL students are smaller (0.02 SD in math, 0.01 SD in reading) than their 0.11 SD and 0.04 SD, though both sets of estimates are statistically indistinguishable from zero and share the pattern that

ELL students’ own elementary test scores are little affected. The 2003–2010 comparison suggests that individual-level data can approximately replicate [Chin et al. \(2013\)](#)’s findings when applied to a somewhat later cohort window, but the near-zero estimates in our main 1999–2008 sample indicate that elementary school effects are sensitive to which cohorts are analyzed.

In contrast, middle school test score effects are more stable across cohort windows. Table 15 reports middle school IV estimates for the same 2003–2010 first grade cohorts. Middle school math estimates in the 2003–2010 window (0.19–0.22 SD across the three samples) are essentially the same as the corresponding 1999–2008 estimates in Tables 11–13 (0.22–0.24 SD). Middle school reading estimates are positive and statistically significant in both windows, though somewhat larger in the 2003–2010 window (0.25–0.31 SD) than in the 1999–2008 window (0.18–0.20 SD). Taken together, this pattern suggests that the effects of bilingual education on cognitive achievement are more consistently measured at later grades than in the years immediately following treatment.

5.3 Long-Run Outcomes

We now turn to how exposure to bilingual education in the first grade affects longer-run educational outcomes, such as high school graduation and college enrollment. Although bilingual education may have limited effect on early elementary test scores, the effect may manifest later either through academic skills or through less tangible forms of social capital. We discuss the potential mechanisms of bilingual education which may differ for the short-run and long-run outcomes in Section 6. Tables 8–10 present the reduced form and IV estimates of the effect of bilingual education exposure in the 1st grade on high school graduation and college enrollment. Again, each column corresponds to a different outcome variable and we focus on the IV estimates. Tables 9–10 report separate results for Spanish-speaking ELL and non-Spanish, non-ELL students. Discussion in this section focuses on the 1999–2008 main sample.

To begin with, there is little to no effect of bilingual education on high school graduation. Column (1) of Table 8 reports an IV estimate of -1.2 percentage points, which is statistically insignificant. The unconditional mean of high school graduation for the sample is around 74%, and even for the subgroup of Spanish-speaking ELL students the graduation rate is over 75%. The small gap in high school graduation rates between ELL and non-ELL students implies that the ability to communicate in English is perhaps not the crucial factor that determines high school graduation, so it is not surprising that bilingual education has negligible effect on this outcome.

On the other hand, bilingual education has a sizable effect on college enrollment, especially on four-year university enrollment. The IV estimate in column (3) of Table 8 indicates that first-grade exposure to bilingual education raises the probability of enrolling in a four-year Texas public university by 6.9 percentage points, a 46% increase relative to the sample mean of 15.1% ($p = 0.071$). Community college enrollment (column (4)) rises by 5.2 percentage points but is not statistically significant, and the combined “any college” effect (column (2)) is 7.8 percentage points (about 18%; $p = 0.169$). The overall long-run college-going result is therefore driven by four-year enrollment.

The effect on four-year enrollment is present for both Spanish-speaking ELL and non-Spanish, non-ELL students, but the pattern differs across the two groups. For Spanish-speaking ELL students, bilingual education in first grade raises four-year university enrollment by 6.6 percentage points (Table 9, column (3); $p = 0.059$), which is a 70% increase relative to the sample mean of 9.4%. The corresponding effect on any-college enrollment is essentially zero (+0.3 percentage points, not significant), and the point estimate on community college enrollment is slightly negative. In other words, the entirety of the college-going effect for Spanish-speaking ELL students is concentrated on four-year universities.

For non-Spanish, non-ELL students, bilingual education raises four-year enrollment by 6.2 percentage points (Table 10, column (3); $p = 0.154$), a 38% increase relative to the sample mean of 16.3%. Community college enrollment for these students rises by 6.5 percentage points (about 16%, not significant), so their any-college estimate is 8.5 percentage points but the estimates are noisy. The point estimates on four-year enrollment for the two groups are therefore of similar magnitude in percentage points, but the implied percent increase relative to baseline enrollment is much larger for Spanish-speaking ELL students because their baseline four-year enrollment rate is lower.

Taken together, these results show that bilingual education has meaningful positive effects on four-year college enrollment for both ELL and non-ELL students, and that the effect on non-ELL students is quantitatively important, although the estimate is slightly noisy. This confirms that there is a positive spillover from bilingual education to non-ELL students that persists into post-secondary education.

6 Mechanisms: Effectiveness of Bilingual Education

In this section, we discuss mechanisms by which a bilingual education program would impact outcomes for ELL and non-ELL students both in the short-run and long-run. An obvious mechanism by which bilingual education impacts ELL students is that they receive instruction in their native language. This should improve the student’s ability to understand

classroom content, especially in subjects such as math and science. Conversely, bilingual education may slow an ELL student’s English acquisition because the student interacts with the teacher in the student’s native language rather than English.

Bilingual education teachers may improve ELL student outcomes because they have a similar race or background to the students compared to a standard teacher. Previous research has found that having a teacher with a similar background can improve student short-run outcomes such as test scores and behavior (Dee (2004), Fairlie et al. (2014), Egalite et al. (2015)). This may be because these teachers better understand how to educate students who share their background or because teachers act as role models. Having a teacher with a similar background has also been shown to improve student long-run outcomes such as high school graduation and college enrollment (Delhommer (2022), Gershenson et al. (2018)). Long-run outcomes could be improved by bilingual education teachers for ELL students if these teachers improve skills that are not well measured by contemporaneous academic achievement or because teachers increase students’ expectations about the return to education or their own ability.

Furthermore, exposure to bilingual education vastly changes the classroom composition that students experience. Under an ESL program, ELL and non-ELL students are mixed in the same classroom. Under a bilingual education program, ELL students are placed in a separate classroom from non-ELL students. ELL students on average come from a lower socioeconomic background and have lower test scores compared to their non-ELL peers, so bilingual education reduces the average ability of an ELL student’s classmates while increasing the average ability of a non-ELL student’s classmates. Previous research has explored how the quality of a student’s peers can impact students’ outcomes finding mixed results (Sacerdote (2011)). In addition to test scores, interacting with English proficient peers may be an important channel by which ELL students learn English. When separated into the bilingual education classroom, ELL students will have few classmates who are fully proficient in English, which potentially hampers their acquisition of English. Our finding of a substantial spillover to non-ELL students’ middle school test scores and college enrollment is consistent with this classroom-composition mechanism.

The long-run impacts of bilingual education in first grade may differ from its short-run impacts. In our data, the pattern of effects (small effects on elementary test scores, larger and often significant effects on middle school test scores, and sizable effects on four-year college enrollment) is consistent with benefits that accumulate or become measurable only after the treatment ends. The goal of bilingual education is for ELL students to return to the mainstream classroom, and this transition can be an abrupt change in learning environment.⁸

⁸Previous research has also found mixed results on the effects of changing school environments on student

Several mechanisms can generate the delayed pattern we observe. For ELL students, bilingual instruction may build content knowledge in math, science, and reading comprehension in the native language during elementary school, but this knowledge is difficult to demonstrate on English-language assessments while English proficiency is still developing; once these students transition to the mainstream classroom and gain sufficient English fluency, typically by middle school, the underlying conceptual foundation can surface on standardized tests. For non-ELL students, the classroom-composition changes during elementary school (i.e. fewer ELL classmates and correspondingly higher average peer achievement) may compound over subsequent grades and yield measurable middle school achievement differences even when the immediate elementary effects are small. Bilingual education may also improve non-cognitive skills, motivation, or engagement in both groups that do not appear in contemporaneous test scores but affect learning trajectories in later grades. This delayed-benefit interpretation mirrors findings from previous work on early-education interventions ([Bailey et al. \(2021\)](#)).

7 Conclusion

In this paper, we have analyzed how school districts’ provision of bilingual education vs. ESL in first grade impacts student outcomes both in the short-run (elementary school test scores) and longer-run (middle school test scores, high school graduation, and college enrollment). Our empirical strategy leverages policy variation in Texas that requires school districts with 20 or more ELL students with the same home language in the same elementary grade to provide bilingual education; schools below this cut-off are free to choose bilingual education or ESL programs for their ELL students. Students above and below this cut-off should be similar except that students above the cut-off are more likely to have been at a school district that provided bilingual education.

Using the 1999–2008 first grade cohorts as our main sample, we find that bilingual education has small and statistically insignificant impacts on elementary school test scores, but produces sizable positive effects on middle school test scores (approximately 0.23 SD in math and 0.19 SD in reading) and on four-year university enrollment (about 7 percentage points, a 46% increase). Positive four-year enrollment effects are found for both Spanish-speaking ELL and non-Spanish, non-ELL students. Peer effects, facing different instructors, and changes in classroom composition could be important mechanisms by which bilingual education impacts these longer-run student outcomes. We also show that when we restrict

outcomes in the context of school closings ([Brummet \(2014\)](#)), grade promotion ([Rockoff and Lockwood \(2010\)](#)), and voluntary student moves ([Hanushek et al. \(2004\)](#)).

attention to a cohort window comparable to Chin et al. (2013) (first grade cohorts entering 2003–04 through 2009–10), our elementary test-score estimates are similar in magnitude and pattern to theirs, so our qualitative departure from their conclusions comes from extending the analysis to middle school test scores and long-run college outcomes rather than from a difference in the elementary-school results themselves.

This study and most previous studies have focused on the effectiveness of ESL compared to bilingual education, but schools may choose a variety of bilingual education programs to implement. Bilingual education programs differ in the speed with which ELL students are transitioned back to the standard classroom (early vs. late exit programs). They also differ in the extent to which they promote proficiency in the native language (transitional vs. dual language programs). Some bilingual education programs even combine native and non-native English students into the same classroom (two-way immersion programs). Understanding the effectiveness of these different modes of bilingual education is an important next step for future research, especially for large school districts with many ELL students.

References

- Anantha, S. (2024). How to become a certified esl teacher in texas. <https://texestest.org/texas-teacher-certification/texas-esl-teachers/>. Accessed: 2024-02-16.
- Bailey, M. J., Sun, S., and Timpe, B. (2021). Prep school for poor kids: The long-run impacts of head start on human capital and economic self-sufficiency. *American Economic Review*, 111(12):3963–4001.
- Brummet, Q. (2014). The effect of school closings on student achievement. *Journal of Public Economics*, 119:108–124.
- Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., and Yagan, D. (2011). How does your kindergarten classroom affect your earnings? evidence from project star. *The Quarterly journal of economics*, 126(4):1593–1660.
- Chetty, R., Friedman, J. N., and Rockoff, J. E. (2014). Measuring the impacts of teachers ii: Teacher value-added and student outcomes in adulthood. *American economic review*, 104(9):2633–79.
- Chin, A., Daysal, N. M., and Imberman, S. A. (2013). Impact of bilingual education programs on limited english proficient students and their peers: Regression discontinuity evidence from texas. *Journal of Public Economics*, 107:63–78.
- Collier, V. P. and Thomas, W. P. (2017). Validating the power of bilingual schooling: Thirty-two years of large-scale, longitudinal research. *Annual review of applied linguistics*, 37:203–217.

- Dee, T. S. (2004). Teachers, race, and student achievement in a randomized experiment. *Review of economics and statistics*, 86(1):195–210.
- Delhommer, S. (2022). High school role models and minority college achievement. *Economics of Education Review*, 87:102222.
- Deming, D. (2009). Early childhood intervention and life-cycle skill development: Evidence from head start. *American Economic Journal: Applied Economics*, 1(3):111–34.
- Egalite, A. J., Kisida, B., and Winters, M. A. (2015). Representation in the classroom: The effect of own-race teachers on student achievement. *Economics of Education Review*, 45:44–52.
- Fairlie, R. W., Hoffmann, F., and Oreopoulos, P. (2014). A community college instructor like me: Race and ethnicity interactions in the classroom. *American Economic Review*, 104(8):2567–91.
- Gershenson, S., Hart, C. M., Hyman, J., Lindsay, C., and Papageorge, N. W. (2018). The long-run impacts of same-race teachers. Technical report, National Bureau of Economic Research.
- Gray-Lobe, G., Pathak, P. A., and Walters, C. R. (2021). The long-term effects of universal preschool in boston. Technical report, National Bureau of Economic Research.
- Hanushek, E. A., Kain, J. F., and Rivkin, S. G. (2004). Disruption versus tiebout improvement: The costs and benefits of switching schools. *Journal of public Economics*, 88(9-10):1721–1746.
- Imbens, G. W. and Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of econometrics*, 142(2):615–635.
- Jackson, C. K. (2018). What do test scores miss? the importance of teacher effects on non-test score outcomes. *Journal of Political Economy*, 126(5):2072–2107.
- Lee, D. S. and Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of economic literature*, 48(2):281–355.
- Lopez, B. (2022). Texas schools are majority hispanic. there’s been a shortage of bilingual teachers since 1990, and the pandemic made it worse. *The Texas Tribune*.
- Matsudaira, J. D. (2005). Sinking or swimming? *Evaluating the impact of English immersion versus bilingual education*. Berkeley: University of California.
- NCES (2018). The nation’s report card. https://www.nationsreportcard.gov/math_2017/#/nation/achievement?grade=4 and https://www.nationsreportcard.gov/reading_2017/#/nation/achievement?grade=4. Accessed: 2024-01-30.
- NCES (2023). *The Condition of Education 2023*, chapter English Learners in Public Schools. U.S. Department of Education, Institute of Education Sciences, <https://nces.ed.gov/programs/coe/indicator/cgf>.

- Pages, R., Lukes, D. J., Bailey, D. H., and Duncan, G. J. (2020). Elusive longer-run impacts of head start: Replications within and across cohorts. *Educational Evaluation and Policy Analysis*, 42(4):471–492.
- Petek, N. and Pope, N. (2016). The multidimensional impact of teachers on students. Technical report, University of Chicago Working Paper.
- Pope, N. G. (2016). The marginal effect of k-12 english language development programs: Evidence from los angeles schools. *Economics of Education Review*, 53:311–328.
- Rockoff, J. E. and Lockwood, B. B. (2010). Stuck in the middle: Impacts of grade configuration in public schools. *Journal of public economics*, 94(11-12):1051–1061.
- Rolle, R. A. and Jimenez-Castellanos, O. (2014). An efficacy analysis of the texas school funding formula with particular attention to english language learners. *Journal of Education Finance*, 39(3):203–221.
- Sacerdote, B. (2011). Peer effects in education: How might they work, how big are they and how much do we know thus far? In *Handbook of the Economics of Education*, volume 3, pages 249–277. Elsevier.
- Texas Education Agency (2022a). English as a second language (esl) program implementation rubric. <https://tea.texas.gov/academics/special-student-populations/english-learner-support/esl-program-implementation-rubric.pdf>. Accessed: 2024-02-16.
- Texas Education Agency (2022b). Transitional bilingual education (tbe) program implementation rubric. <https://tea.texas.gov/academics/special-student-populations/english-learner-support/tbe-program-implementation-rubric.pdf>. Accessed: 2024-02-16.
- Texas Teachers of Tomorrow (TTOT) (2023). How to become a bilingual teacher in texas. <https://www.teachersoftomorrow.org/texas/teacher-resources/high-need-areas/bilingual/>. Accessed: 2024-02-16.
- Williams, C. and Zabala, J. (2023). How to grow bilingual teacher pathways: Making the most of u.s. linguistic and cultural diversity. <https://tcf.org/content/report/how-to-grow-bilingual-teacher-pathways-making-the-most-of-u-s-linguistic-and-cultural>. Accessed: 2024-02-16.
- Wright, W. E. (2010). Foundations for teaching english language learners: Research, theory. *Policy, and Practice*. Philadelphia: Caslon Publishing, pages 70–76.

A Tables and Figures

List of Figures in the Appendix

1	First Stage by First Grade Cohort	38
2	Discontinuous Change in Probability of Providing Bilingual Education at the Cut-off	38
3	McCrary Density Test	38

List of Tables in the Appendix

1	School District Information in 2005 by District Size and Number of Spanish ELL Students	23
2	School District Information by Year	24
3	First Stage Coefficients	25
4	Balance Test (All)	26
5	Balance Test (Spanish, ELL)	27
6	Balance Test (Non-Spanish, Non-ELL)	28
7	Balance Test (School level)	29
8	Effect of Bilingual Education on Long-Run Outcomes (All)	30
9	Effect of Bilingual Education on Long-Run Outcomes (Spanish, ELL)	31
10	Effect of Bilingual Education on Long-Run Outcomes (Non-Spanish, Non-ELL)	32
11	Effect of Bilingual Education on Test Score Outcomes (All)	33
12	Effect of Bilingual Education on Test Score Outcomes (Spanish, ELL)	34
13	Effect of Bilingual Education on Test Score Outcomes (Non-Spanish, Non-ELL)	35
14	Elementary Test Score Effects, 2003–2010 Cohort: Comparison to Chin, Daysal, and Imberman (2013)	36
15	Middle School Test Score Effects, 2003–2010 First Grade Cohorts	37

Table 1: School District Information in 2005
by District Size and Number of Spanish ELL Students

	≤ 200		> 200	
	8-39	> 39	8-39	> 39
Avg. Number of Students	98.215	122.545	323.177	1566.541
S.D. Number of Students	48.881	38.914	113.576	2130.730
Avg. Number of Campuses	1.099	1.091	2.899	15.894
S.D. Number of Campus	0.351	0.302	1.972	22.269
Pct. District Provides Bilingual Education	0.271	0.727	0.443	0.953
Pct. Spanish LEP Students	0.225	0.592	0.068	0.267
Pct. Female	0.477	0.459	0.481	0.481
Pct. Black	0.078	0.039	0.115	0.128
Pct. Hispanic	0.511	0.805	0.271	0.526
Pct. White	0.403	0.153	0.592	0.321
Districts	181	11	79	170
First Grader Obs.	17,777	1,348	25,531	266,312

*Note: *=10% Significance Level, **=5% Significance Level, ***=1% Significance Level*

Table 2: School District Information by Year

	1994-1998	1999-2008	2009-2019
Avg. Number of Students	103.060	99.823	103.515
Avg. Number of Campuses	1.128	1.106	1.188
Pct. District Provides Bilingual Education	0.205	0.260	0.270
Pct. Spanish LEP Students	0.195	0.213	0.212
Pct. Female	0.483	0.483	0.487
Pct. Black	0.074	0.084	0.082
Pct. Hispanic	0.499	0.496	0.498
Pct. White	0.423	0.411	0.391
Districts	596	1,662	2,168
First Graders Obs.	61,424	165,905	224,420

*Note: *=10% Significance Level, **=5% Significance Level, ***=1% Significance Level*

Table 3: First Stage Coefficients

	1994-1998	1999-2008	2009-2019
Coef., No Controls	0.501	0.505	0.448
St.E., No Controls	0.074	0.046	0.046
Coef., With Controls	0.256	0.199	0.079
St.E., With Controls	0.075	0.054	0.050
Districts	192	302	337
Campuses	222	379	464
Students	63,214	172,158	230,883
<i>Note: * = 10% Significance Level, ** = 5% Significance Level, *** = 1% Significance Level</i>			

Table 4: Balance Test (All)

	(1)	(2)	(3)	(4)	(5)	(6)
Coefficient	-0.017	0.007	-0.010	0.002	-0.005	-0.002
Standard Error	(0.026)	(0.005)	(0.016)	(0.032)	(0.005)	(0.004)
Outcome Mean	0.594	0.479	0.092	0.451	0.099	0.030
Observations	172,158	172,158	172,158	172,158	172,158	172,158
Outcome Sample	FRL Status All	Female All	Black All	Hispanic All	Special Ed. All	Gifted All

*Note: *=10% Significance Level, **=5% Significance Level, ***=1% Significance Level*

Table 5: Balance Test (Spanish, ELL)

	(1)	(2)	(3)	(4)	(5)	(6)
Coefficient	-0.035	-0.006	0.001	-0.004	0.008	0.001
Standard Error	(0.035)	(0.013)	(0.001)	(0.003)	(0.009)	(0.003)
Outcome Mean	0.836	0.483	0.001	0.992	0.072	0.011
Observations	28,018	28,018	28,018	28,018	28,018	28,018
Outcome Sample	FRL Status Spanish, ELL	Female Spanish, ELL	Black Spanish, ELL	Hispanic Spanish, ELL	Special Ed. Spanish, ELL	Gifted Spanish, ELL

*Note: *=10% Significance Level, **=5% Significance Level, ***=1% Significance Level*

Table 6: Balance Test (Non-Spanish, Non-ELL)

	(1)	(2)	(3)	(4)	(5)	(6)
Coefficient	-0.013	0.010*	-0.011	0.000	-0.008	-0.003
Standard Error	(0.027)	(0.006)	(0.019)	(0.037)	(0.005)	(0.004)
Outcome Mean	0.543	0.479	0.111	0.344	0.104	0.034
Observations	142,351	142,351	142,351	142,351	142,351	142,351
Outcome	FRL Status	Female	Black	Hispanic	Special Ed.	Gifted
Sample	Non-Spanish, Non-ELL	Non-Spanish, Non-ELL	Non-Spanish, Non-ELL	Non-Spanish, Non-ELL	Non-Spanish, Non-ELL	Non-Spanish, Non-ELL

*Note: *=10% Significance Level, **=5% Significance Level, ***=1% Significance Level*

Table 7: Balance Test (School level)

	(1)	(2)	(3)	(4)	(5)	(6)
Coefficient	-0.009	0.012*	-0.015	0.017	-0.020*	-0.002
Standard Error	(0.031)	(0.006)	(0.018)	(0.035)	(0.011)	(0.005)
Outcome Mean	0.613	0.479	0.088	0.500	0.107	0.029
Observations	1,838	1,838	1,838	1,838	1,838	1,838
Outcome Sample	FRL Status School level	Female School level	Black School level	Hispanic School level	Special Ed. School level	Gifted School level

*Note: *=10% Significance Level, **=5% Significance Level, ***=1% Significance Level*

Table 8: Effect of Bilingual Education on Long-Run Outcomes (All)

	(1)	(2)	(3)	(4)
IV Coefficient	-0.012	0.078	0.069*	0.052
IV Standard Error	0.039	0.056	0.038	0.058
FS Coefficient	0.199***	0.199***	0.199***	0.199***
FS Standard Error	0.054	0.054	0.054	0.054
RF Coefficient	-0.002	0.015	0.014**	0.010
RF Standard Error	0.008	0.010	0.006	0.011
Outcome Mean	0.740	0.439	0.151	0.392
Observations	172,158	172,158	172,158	172,158
Outcome Sample	HS Graduation All	Any College Enroll. All	4-year Univ. Enrollment All	CC Enrollment All

*Note: *=10% Significance Level, **=5% Significance Level, ***=1% Significance Level*

Table 9: Effect of Bilingual Education on Long-Run Outcomes (Spanish, ELL)

	(1)	(2)	(3)	(4)
IV Coefficient	-0.039	0.003	0.066*	-0.032
IV Standard Error	0.057	0.071	0.035	0.071
FS Coefficient	0.262***	0.262***	0.262***	0.262***
FS Standard Error	0.049	0.049	0.049	0.049
RF Coefficient	-0.010	0.001	0.017**	-0.008
RF Standard Error	0.015	0.019	0.009	0.019
Outcome Mean	0.752	0.362	0.094	0.327
Observations	28,018	28,018	28,018	28,018
Outcome Sample	HS Graduation Spanish, ELL	Any College Enroll. Spanish, ELL	4-year Univ. Enrollment Spanish, ELL	CC Enrollment Spanish, ELL

*Note: *=10% Significance Level, **=5% Significance Level, ***=1% Significance Level*

Table 10: Effect of Bilingual Education on Long-Run Outcomes (Non-Spanish, Non-ELL)

	(1)	(2)	(3)	(4)
IV Coefficient	-0.020	0.085	0.062	0.065
IV Standard Error	0.042	0.062	0.043	0.065
FS Coefficient	0.185***	0.185***	0.185***	0.185***
FS Standard Error	0.056	0.056	0.056	0.056
RF Coefficient	-0.004	0.016	0.011*	0.012
RF Standard Error	0.008	0.011	0.007	0.012
Outcome Mean	0.739	0.455	0.163	0.406
Observations	142,351	142,351	142,351	142,351
Outcome Sample	HS Graduation Non-Spanish, Non-ELL	Any College Enroll. Non-Spanish, Non-ELL	4-year Univ. Enrollment Non-Spanish, Non-ELL	CC Enrollment Non-Spanish, Non-ELL

*Note: *=10% Significance Level, **=5% Significance Level, ***=1% Significance Level*

Table 11: Effect of Bilingual Education on Test Score Outcomes (All)

	(1)	(2)	(3)	(4)
IV Coefficient	-0.009	0.049	0.231*	0.185**
IV Standard Error	0.103	0.093	0.136	0.094
FS Coefficient	0.202***	0.201***	0.204***	0.202***
FS Standard Error	0.053	0.053	0.053	0.052
RF Coefficient	-0.002	0.010	0.047**	0.037**
RF Standard Error	0.021	0.019	0.023	0.016
Outcome Mean	-0.096	-0.059	-0.061	-0.026
Observations	381,838	381,902	370,282	379,696
Outcome Sample	Elementary Math All	Elementary Reading All	Middle Math All	Middle Reading All

*Note: *=10% Significance Level, **=5% Significance Level, ***=1% Significance Level*

Table 12: Effect of Bilingual Education on Test Score Outcomes (Spanish, ELL)

	(1)	(2)	(3)	(4)
IV Coefficient	-0.034	-0.013	0.242*	0.201*
IV Standard Error	0.121	0.110	0.137	0.112
FS Coefficient	0.267***	0.264***	0.264***	0.264***
FS Standard Error	0.051	0.051	0.050	0.050
RF Coefficient	-0.009	-0.003	0.064*	0.053*
RF Standard Error	0.032	0.029	0.034	0.027
Outcome Mean	-0.293	-0.340	-0.148	-0.191
Observations	54,365	54,284	53,733	54,694
Outcome Sample	Elementary Math Spanish, ELL	Elementary Reading Spanish, ELL	Middle Math Spanish, ELL	Middle Reading Spanish, ELL

*Note: *=10% Significance Level, **=5% Significance Level, ***=1% Significance Level*

Table 13: Effect of Bilingual Education on Test Score Outcomes (Non-Spanish, Non-ELL)

	(1)	(2)	(3)	(4)
IV Coefficient	-0.006	0.053	0.219	0.177*
IV Standard Error	0.111	0.099	0.145	0.102
FS Coefficient	0.189***	0.189***	0.191***	0.189***
FS Standard Error	0.054	0.054	0.054	0.054
RF Coefficient	-0.001	0.010	0.042*	0.033*
RF Standard Error	0.021	0.019	0.023	0.017
Outcome Mean	-0.063	-0.011	-0.046	0.003
Observations	324,245	324,397	313,524	321,874
Outcome Sample	Elementary Math Non-Spanish, Non-ELL	Elementary Reading Non-Spanish, Non-ELL	Middle Math Non-Spanish, Non-ELL	Middle Reading Non-Spanish, Non-ELL

*Note: *=10% Significance Level, **=5% Significance Level, ***=1% Significance Level*

Table 14: Elementary Test Score Effects, 2003–2010 Cohort:
Comparison to Chin, Daysal, and Imberman (2013)

Sample	Our Estimates (2003–2010 cohort)		Chin et al. (2013), Table 3	
	Elem. Math	Elem. Reading	Elem. Math	Elem. Reading
All	0.185 (0.118)	0.164 (0.102)	0.206* (0.112)	0.143 (0.095)
Spanish, ELL	0.020 (0.143)	0.011 (0.140)	0.107 (0.155)	0.040 (0.156)
Non-Spanish, Non-ELL	0.217* (0.125)	0.181* (0.108)	0.213* (0.114)	0.169* (0.092)
Obs. (Ours)	All: 347,988; Sp., ELL: 52,717; Non-Sp., Non-ELL: 292,289		See Chin et al. (2013)	

*Note: Ours are 2SLS estimates from the 2003–04 through 2009–10 first grade cohorts (individual-level data). Chin et al. (2013) estimates are from their Table 3, Panel B, using school-grade-year cell data for 3rd–5th graders in the 2002–03 through 2009–10 school years. *=10% Significance Level, **=5% Significance Level, ***=1% Significance Level.*

Table 15: Middle School Test Score Effects, 2003–2010 First Grade Cohorts

Sample	Middle Math	Middle Reading
All Students	0.203 (0.137)	0.265** (0.120)
Spanish, ELL	0.220 (0.145)	0.305** (0.144)
Non-Spanish, Non-ELL	0.194 (0.144)	0.252** (0.124)
Obs. (Math)	All: 327,823; Sp., ELL: 50,717; Non-Sp., Non-ELL: 274,429	
Obs. (Reading)	All: 344,058; Sp., ELL: 52,590; Non-Sp., Non-ELL: 288,591	

*Note: 2SLS estimates from the 2003–04 through 2009–10 first grade cohorts (individual-level data). Middle school pools students in grades 6–8. *=10% Significance Level, **=5% Significance Level, ***=1% Significance Level.*

Figure 1: First Stage by First Grade Cohort

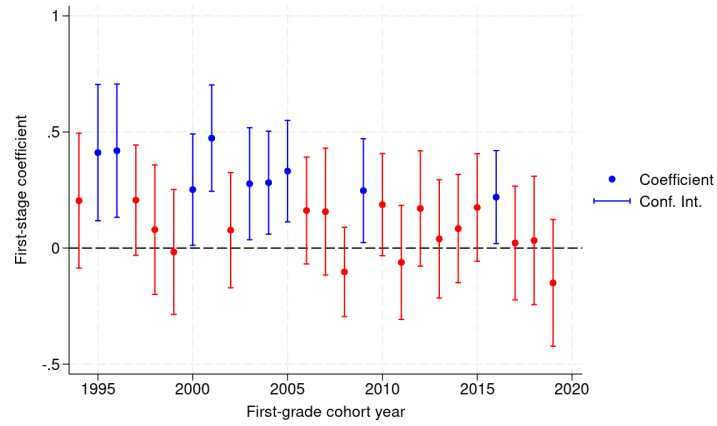


Figure 2: Discontinuous Change in Probability of Providing Bilingual Education at the Cut-off

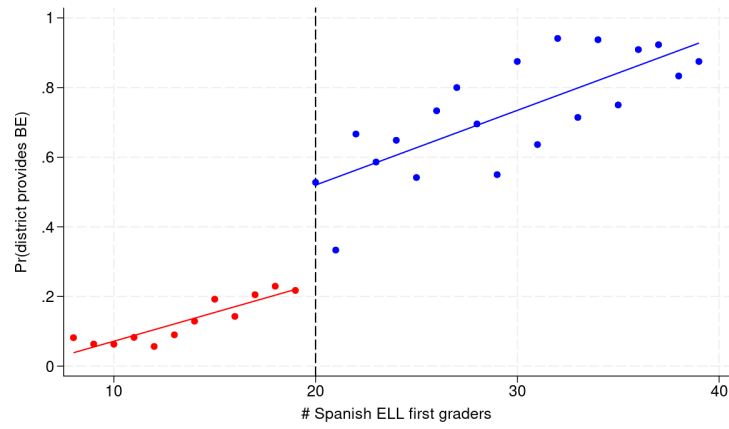


Figure 3: McCrary Density Test

