BILINGUAL EDUCATION AND ENGLISH PROFICIENCY

Christopher Jepsen
Department of Economics
University of Kentucky
335BA Gatton Business and Economics Building
Lexington, KY 40506-0034
jepsen@uky.edu

Abstract

In 2001, California instituted a statewide test measuring English proficiency for English learners, students who are not proficient in English. In 2003 and 2004, nearly 500,000 English learners in grades 1–5 took this test each year. The relationship between bilingual education receipt and English proficiency is estimated using value-added regression models for each section of the test—listening and speaking, reading, and writing. In these regression models, students in bilingual education have substantially lower English proficiency of 0.3 standard deviations or more compared with other English learners in first and second grades. In contrast, the difference between bilingual education and other programs is usually less than 0.1 standard deviations for students in grades 3–5. These results hold for ordinary least squares, school fixed effects, and propensity score models.

1. INTRODUCTION

Over 10 percent of the nation's K–12 students are not proficient in English and speak a language other than English at home (Padolsky 2005). In California, these students—called English learners (ELs)—make up more than 25 percent of the K–12 student body. Although California has by far the most EL students of any state in the country, many other states are experiencing rapid growth in the population of EL students. For example, the EL population has more than doubled in Colorado, Nevada, and Oregon over the last ten years, resulting in a student population that is more than 10 percent EL in each state. Therefore, an analysis of the experience of EL programs in California can provide guidance for other states.

The academic and labor market success of EL students depends in large part on their ability to become proficient in English. Standardized tests are usually given in English, and EL students have consistently lower test scores than students whose primary language is English. English language proficiency is critical for overall academic success. Furthermore, several authors have demonstrated the positive link between English proficiency and success in the labor market (Gonzalez 2000; Trejo 2003; Bleakley and Chin 2004).

As most educators and researchers would agree, instruction targeting English proficiency is necessary for students with little English proficiency. For these students, the key research question concerns the most effective type of instructional services that they receive rather than whether they should receive services at all. Programs that use students' primary (non-English) language in academic subjects are called *bilingual education* programs. There are two important differences between bilingual programs and other EL programs. The first is the obvious difference in the language of instruction: bilingual programs teach academic subjects in the primary language, whereas other EL programs teach these subjects in English. The second difference is that bilingual programs attempt to improve academic skills and primary language skills in addition to English language skills, whereas other EL programs focus only on improving English language skills.

Previous work on bilingual education has focused on the outcome of academic achievement (Greene 1998; Gordon and Hoxby 2002; Lopez 2003).[1] As stated above, improving academic achievement is a long-run objective of bilingual and other EL programs, but it is not the only objective. The short-term objective of all EL programs—bilingual and other—is to improve English proficiency. The goal of this article is to look at the effect of bilingual education on English proficiency. Another objective of bilingual education programs is

1. Matsudaira (2005) looks at the effects of receiving any instructional services, regardless of the language of instruction, on academic achievement.

to improve primary-language proficiency, but studying that objective is beyond the scope of this article.

Using administrative data on almost 500,000 Spanish-speaking EL students in California, I compare English learners who receive bilingual education (academic instruction in the primary language) with English learners who receive EL services that provide academic instruction in English. In California, students are not randomly assigned to bilingual programs. Schools choose whether to offer bilingual education, and students need to obtain parental permission in order to participate in bilingual programs, both consequences of 1998's Proposition 227. I use multiple techniques to control for this nonrandom selection. For example, propensity score analysis matches bilingual students with students receiving other EL services based on the likelihood of participating in bilingual education.

The results of this study show that bilingual education has a sizable negative effect on English proficiency for students in first and second grades, with magnitudes often in excess of 0.3 standard deviations. This pattern is consistent across four different methods: ordinary least squares (OLS), school fixed effects, matching estimators, and instrumental variables (IV). In later elementary grades (3–5), the effect of bilingual education is usually less than 0.1 standard deviations, and it is sometimes not significantly different from zero. This negative effect is not surprising. Non-bilingual programs are narrowly focused on English proficiency, whereas bilingual education programs encourage proficiency in the student's primary language and in English. The pattern of results is also consistent with common approaches to bilingual education in California, where the percentage of instruction in English is as low as 10 percent in kindergarten compared with nearly 50 percent by fifth grade. Finally, Proposition 227 requires that all EL students spend thirty days in English-only classrooms when they first enroll in California schools, and most EL students enroll in California schools in kindergarten or first grade. Thus, students in bilingual programs have thirty fewer days of instruction in bilingual education during their first year in a California school.

2. PROPOSITION 227 AND BILINGUAL EDUCATION IN CALIFORNIA

In June 1998, California voters passed Proposition 227, which requires that students learning English be taught "overwhelmingly in English." However, the proposition does allow for students to enroll in bilingual education under special circumstances. First, a parent or guardian has to fill out a parental waiver form at the school. If there are at least twenty students in the district with waivers, the district is required to offer bilingual education. However, the district is not required to offer bilingual education at every school in the district, so students may be required to switch schools in order to receive

bilingual education. A final requirement of the proposition is that EL students who are new to California schools spend thirty days in English-only classrooms before they can transfer to other programs such as bilingual education.

Rossell (2005) documents the variation in implementation of Proposition 227 across schools and districts, particularly in the years immediately following passage of the proposition. For example, schools (and even classrooms within schools) vary greatly in their use of waivers. Some schools send waivers home with children, and others call parents about waivers even though the law requires parents and guardians to initiate the waiver process. At the other extreme, some schools do not mention waiver availability unless explicitly asked by parents. Another source of variation across the state is that the definition of "overwhelmingly in English" ranges anywhere from 60 to 100 percent of instruction in English. Overall, implementation of the proposition appears to have been more uniform a few years after the proposition passed, with the exception of continued variation in waiver policy (Rossell 2005; Parrish et al. 2006).

3. RELATION TO PREVIOUS WORK

Many researchers have written about bilingual education, but most of these studies are not able to identify the causal effect of bilingual education as opposed to other EL programs on student outcomes (Greene 1998). Some of the studies lack a comparison group (i.e., students not in bilingual programs), whereas others do not control for observable (or unobservable) differences between students in bilingual programs and other students. In this article, the discussion is limited to rigorous analyses that attempt to control for the nonrandom selection of students into bilingual education.

Among the studies with comparison groups and controls for selection, several authors—Cheng (1996), Lopez and Mora (1998), Federman (2000), Mora (2000), and Lopez (2003)—focus on the effects of primary language instruction, usually in academic subjects, on individual students using the National Educational Longitudinal Survey (NELS) or the High School and Beyond (HSB) surveys. This restrictive definition of bilingual education is quite similar to the definition of bilingual education that I employ. A limitation of these studies is that their comparison groups include all students who speak a language at home other than English.[2] The comparison groups include students who are proficient in English as well as students who are not proficient in English.

2. The comparison group is chosen because the authors cannot identify EL status, but they can identify language minority status.

The authors find mixed results of bilingual education due to differences in outcomes, methods, or data sets.[3]

The study mandated by California's Proposition 227 finds insignificant effects of bilingual education on student-level mathematics and reading test scores (Parrish et al. 2006). The study's preferred model is a hierarchical linear model for a large panel of student-level data from the Los Angeles school district. However, the authors acknowledge that this model may not fully control for the nonrandom selection of students into bilingual education programs.

Gordon and Hoxby (2002) look at the effect of bilingual education (again defined as academic instruction in the primary language) on grade-level academic achievement in California. In their preferred IV model, they find a positive effect of bilingual education on achievement in multiple subjects, although these effects are concentrated in the earlier grades. The instrument is the change in the school's percentage of students eligible for bilingual education if the school had mechanically followed Proposition 227.[4] They look separately at the effects of bilingual education on the performance of EL students and on the performance of English-proficient students.

Greene (1998) conducts a meta-analysis of eleven rigorous studies on bilingual education. He finds a positive effect of bilingual education on academic achievement (primarily reading). In other words, students in bilingual education programs outperform other EL students, students who receive EL services in English, and those who receive no EL services. August and Shanahan (2006) provide a more recent meta-analysis and reach similar conclusions. Genesee et al. (2006) summarize the education literature and find positive associations between primary language instruction and various measures of English proficiency and academic achievement.

Rather than looking at bilingual education, Matsudaira (2005) includes all English language services in a single category. He uses a regression discontinuity design based on a standardized test determining eligibility for English language services; students above the threshold do not receive English language

3. As an example, here are the differences in instruments used. Cheng (1996) uses two instruments: (1) the student's percentile distribution in the rank (by school) of time spent in the United States and (2) the year of entry in the United States. Federman (2000) uses three instruments: (1) whether the student's school is required by state law to provide bilingual education or English as a second language; (2) availability of state funding for schools choosing to offer language assistance programs for limited English proficiency (LEP) students; and (3) whether the state offered certification in bilingual education for teachers in 1987, a measure of state regulation of language assistance programs. Mora (2000) does not control for selection bias. Lopez (2003) uses two instruments: (1) first language and home language as of eighth grade (NELS) or tenth grade (HSB) and (2) the quartile rank used by Cheng (1996).
4. The change is between the school's share of students eligible for bilingual education in the 1998–99 school year, which is zero if the school mechanically followed Proposition 227, and the school's share of students eligible for bilingual education in the 1997–98 school year.

services, and students just below the threshold receive English language services. For the sample of students near the passing cutoff, he finds negligible differences in mathematics and reading achievement between students who receive English language services and students who receive no such services.

In sum, these studies find mixed effects of bilingual education on academic achievement and labor market outcomes. They use a host of techniques to control for the nonrandom placement of students into bilingual education. Their data sets either are aggregated to the grade level or contain English proficient students along with ELs. However, none of the studies considers the effect of bilingual education on English proficiency (as distinct from academic achievement in English, reading, or writing).

4. DATA

In 2001, California implemented a statewide assessment, the California English Language Development Test (CELDT), to measure proficiency in English.[5] The CELDT is administered to *all* EL students in the fall of each school year. The CELDT consists of three portions: listening and speaking, reading, and writing. The listening and speaking portion is administered individually to students in all grades; the tester asks each student a set of questions to measure both specific and general skills (Legislative Analyst's Office 2004). The reading and writing portions of the test, administered to students in grades 2–12, are standardized tests with multiple-choice and short-answer sections. An overall scale score is calculated based on the scale scores for each individual portion.

The CELDT data contain extensive student-level test score data on English proficiency, but they contain no measures of academic achievement. Scale scores are available for each portion of the English proficiency test (listening and speaking, reading, and writing), as is the overall score. These scores are available for the current school year as well as the previous school year. Comparisons of CELDT scores are valid only within each grade span (K–2, 3–5, 6–8, and 9–12). In other words, a scale score of 500 in grades 1 or 2 may or may not represent the same level of English proficiency as a scale score of 500 in grades 4 or 5. Consequently, separate regressions are estimated for each grade and year.

For each student, the CELDT data contain four categories for English language instruction (collectively referred to as EL services): (1) English language development (ELD), (2) ELD and specially designed academic instruction delivered in English (SDAIE), (3) ELD and SDAIE with primary language

5. Prior to 2001, statewide data on English proficiency are not available. Individual districts and schools may have conducted their own assessments, but data on these assessments are not available.

support, and (4) ELD and academic subjects through primary language.[6] The last category is defined as bilingual education due to the use of primary language in academic subjects.[7] For all these programs, academic instruction focuses on teaching the academic material, with the acknowledgment that the students have limited English proficiency. Bilingual programs address this issue by providing part or all of the instruction in the primary language. Other EL programs address the issue by providing additional explanation of terms, using pictures to add context, and in some cases providing small amounts of clarification in the primary language.

CELDT demographic data also contain information on primary language, gender, age, mobility, and program participation (for programs other than EL). The mobility information includes the number of years in the current school and the number of years in U.S. schools.[8] The non-EL programs include special education receipt and Title I receipt (i.e., supplemental services, typically targeting low-income students).

Student-level CELDT data are supplemented with school-level data from the California Basic Education Data System (CBEDS). These data include school-level information on teachers (because researchers cannot link EL students to their individual teachers), school demographics, and academic achievement (for non-EL students) in English and language arts.

The CELDT data are a repeated cross section rather than a panel. They contain the previous year's CELDT score (in addition to the current year's CELDT score), but all other data are from the current school year. The bilingual education variable is from the current year, which I treat as a proxy for bilingual education in the previous year. For a subset of students, I was able to merge CELDT data across years using the school identifier, birth date, grade, and gender. For these students, most had the same EL program for both years. Furthermore, the regression results were qualitatively similar when using the previous year's EL program rather than the current year's EL program. When both the current year's and previous year's EL programs are included, the coefficient for the previous year is usually statistically insignificant, whereas the coefficient for the current year is essentially unaffected by the inclusion of the previous year's EL program.[9] Therefore I use data from the current year

6. There is a fifth category for "other" EL services, treated as non-bilingual education, which contains few students.
7. The bilingual education category includes all types of bilingual programs such as early-exit programs as well as dual-immersion programs.
8. Data prior to 2004 also contain information on the years in the current district. The data on years in U.S. schools are quite noisy, so I do not use them in the analysis.
9. There are several reasons to suspect that the previous year's EL program is measured with error. Many students are not matched, and there is the possibility that other students are matched incorrectly due to the use of student characteristics rather than student identifiers (which are not

because the sample is much larger and more representative, and the results are not sensitive to this choice.

There are a few sample restrictions in the data. Students in their first year in a California school have not been exposed to any EL program and therefore are excluded from the sample. Students who skipped a grade or repeated a grade are also excluded in order to ensure that all students' previous CELDS scores are from the same grade. Because 99 percent of students in bilingual programs in California are native Spanish speakers, I only include native Spanish speakers in the analysis sample. One or two percent of EL students attend charter schools, and they are excluded from the analysis to avoid confounding the effects of charter schools with the effects of bilingual education. Similarly, students in nontraditional schools such as alternative schools, special education schools, and county offices of education also are excluded. Finally, less than 1 percent of the students have missing CELDT scores due to invalid values of either test scores or grade levels, so they are excluded from the sample. The findings are robust to alternative sample definitions such as including charter school students or students who skipped or repeated a grade.

5. DESCRIPTIVE CHARACTERISTICS

Table 1 presents the descriptive statistics for the analysis sample, a combined sample of students in the 2003 and 2004 CELDT data. The statistics are separated by grade (1–5). The average test score is between 450 and 550 for the current year, with a standard deviation between 50 and 65. The percentage of students in bilingual education programs is highest for students in first and second grades, at 14 percent. The percentage is around 9 percent for fourth-grade students and 6 percent for fifth-grade students. These small percentages are to be expected given the provision of Proposition 227 that EL students be taught "overwhelmingly" in English. Roughly 80 percent of EL students are eligible for Title I services, and between 4 percent (first grade) and 11 percent (fifth grade) receive special education services. The average EL student attends a school that has little language diversity, and the majority of students in the school are ELs.

6. PARTICIPATION IN BILINGUAL EDUCATION

This section takes a closer look at the relationship between bilingual education participation and student and school characteristics. In California, the participation decision for bilingual education is a two-part process. The first step for a student to participate in bilingual education is for the student to attend

available). A previous year EL program is often missing for students who arrived in California schools in the previous year.

Table 1. Descriptive Statistics

	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5
Current CELDT scale scores					
Listening/speaking	499.0 (58.7)	537.1 (57.9)	504.8 (56.1)	531.7 (60.5)	548.7 (63.4)
Reading			470.6 (49.4)	497.2 (49.6)	515.6 (50.5)
Writing			494.6 (53.8)	515.8 (50.5)	528.4 (49.6)
Lagged CELDT scale scores					
Listening/speaking	411.5 (91.0)	488.2 (68.2)	520.9 (70.4)	492.6 (69.1)	510.1 (75.2)
Reading			444.5 (46.9)	462.5 (49.2)	485.9 (50.9)
Writing			462.6 (59.5)	478.7 (59.6)	498.4 (59.4)
Bilingual education	0.138 (0.345)	0.136 (0.343)	0.122 (0.328)	0.091 (0.288)	0.061 (0.239)
Student demographics					
Female	0.489 (0.500)	0.492 (0.500)	0.492 (0.500)	0.489 (0.500)	0.483 (0.500)
Title I	0.797 (0.402)	0.811 (0.392)	0.823 (0.381)	0.837 (0.369)	0.848 (0.359)
Special education	0.044 (0.205)	0.060 (0.238)	0.077 (0.266)	0.093 (0.291)	0.108 (0.310)
Years in current school	1.89 (0.310)	2.66 (0.624)	3.33 (1.003)	3.85 (1.415)	4.27 (1.812)
School demographics					
Language homogeneity index	0.861 (0.177)	0.865 (0.175)	0.866 (0.175)	0.866 (0.174)	0.866 (0.175)
Class size	19.4 (1.51)	19.3 (1.56)	19.4 (1.62)	19.0 (3.02)	18.5 (4.13)
Percent EL	56.4 (22.2)	56.9 (22.0)	57.0 (21.9)	56.8 (21.9)	56.5 (21.8)
Average teacher experience	11.2 (2.90)	11.2 (2.87)	11.2 (2.87)	11.2 (2.85)	11.2 (2.88)
Percent BA or less	25.5 (18.9)	25.8 (18.8)	26.1 (18.8)	26.5 (18.8)	26.4 (18.8)
Percent BA + 30	47.4 (21.0)	46.9 (21.0)	46.6 (20.8)	46.2 (20.5)	46.2 (20.5)
Percent fully credentialed	89.4 (11.2)	89.4 (10.9)	89.3 (10.9)	89.0 (11.0)	89.1 (10.9)
EL teacher access	0.090 (0.100)	0.088 (0.103)	0.087 (0.104)	0.086 (0.094)	0.085 (0.087)
Percent new students	17.3 (8.84)	17.5 (8.92)	17.7 (9.41)	17.6 (9.28)	17.8 (9.59)
Average CST score	331.0 (16.1)	330.7 (16.1)	330.8 (16.0)	331.1 (15.9)	331.0 (15.9)
Observations	224,608	241,005	235,584	222,607	201,386

Note: Standard deviations are in parentheses.

a school that offers bilingual education, as schools are not required to offer it. If the student attends a school that offers bilingual education, he or she may participate in bilingual education if at least one parent visits the school and signs a permission slip. Thus the placement of students into bilingual programs is a subjective process by the schools (whether to offer bilingual education) and by the parents (whether to enroll their children in bilingual education). CELDT data identify only the outcome of this process: whether or not the student participates in bilingual education.

Table 2 contains the marginal effects from a probit on bilingual participation, presented in percentage terms. The marginal effects are calculated at the mean for continuous variables; for dichotomous variables, the marginal effect is the change in probability of bilingual participation from changing the dichotomous variable from zero to one.[10] Recall that bilingual participation is largest in the lower grades—around 14 percent in first grade—and smallest in the highest grades—around 6 percent in fifth grade. Participation in bilingual education is negatively associated with previous CELDT scores in reading/listening and writing, but the association is positive for reading scores. For first-grade students, a one standard deviation increase in the previous listening and speaking score (about ninety points; see table 1) is associated with a decreased likelihood of participation in bilingual education of about 4 percentage points—approximately one-third of the mean participation rate. The effect is approximately 5 percentage points for second-grade students. When measured as a percentage of the bilingual participation rate, the effect of previous test scores is sizable, ranging from 27 to 49 percent. In other words, more proficient students are not as likely to participate in bilingual education.

Other student characteristics have significant differences in bilingual education participation. Females in third and fourth grades have slightly higher participation. Students receiving special education services correspond to noticeably lower participation in bilingual education. On the other hand, students who have spent more time in their current school correspond with somewhat higher participation.

Selected school characteristics have significant relationships with bilingual education. The percentage of EL students in the school has a large, positive relationship with bilingual education participation: the one standard deviation effect is similar to that of lagged CELDT scores. Similarly, the language homogeneity index—higher for schools with little diversity in the home language of EL students—also has a large and positive relationship. Perhaps the positive coefficients on these two variables reflect economies of scale—more

10. Although not reported, results from a linear probability model are quite similar.

Table 2. Probit Marginal Effects for Bilingual Education Participation, as Percentages

	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5
Lagged listen/ speak	−0.040 (0.003)	−0.073 (0.004)	−0.012 (0.002)	−0.020 (0.002)	−0.010 (0.002)
Lagged reading			−0.007 (0.004)	0.029 (0.004)	0.016 (0.004)
Lagged writing			−0.080 (0.005)	−0.062 (0.005)	−0.033 (0.004)
Female	0.07 (0.19)	−0.07 (0.18)	0.72 (0.17)	0.41 (0.15)	0.15 (0.13)
Title I	0.63 (1.26)	0.57 (1.19)	−0.46 (1.27)	−0.05 (1.01)	0.23 (0.82)
Special education	−2.35 (0.55)	−3.09 (0.48)	−5.25 (0.39)	−3.41 (0.32)	−2.31 (0.29)
Years in current school	1.66 (1.00)	2.54 (0.51)	2.30 (0.31)	1.50 (0.20)	0.48 (0.11)
Language index	15.76 (3.88)	14.78 (3.51)	14.48 (3.32)	13.51 (2.78)	7.65 (2.13)
Class size	−1.66 (0.45)	−1.79 (0.40)	−1.59 (0.36)	−0.32 (0.16)	−0.37 (0.09)
Percent EL	0.23 (0.03)	0.22 (0.03)	0.20 (0.03)	0.13 (0.02)	0.06 (0.02)
Average teacher experience	−0.66 (0.93)	−0.40 (0.89)	0.12 (0.89)	0.12 (0.81)	−0.38 (0.70)
Percent BA or less	0.02 (0.05)	−0.04 (0.05)	−0.06 (0.05)	−0.03 (0.04)	−0.07 (0.03)
Percent BA + 30	0.08 (0.03)	0.02 (0.03)	0.01 (0.03)	0.01 (0.03)	−0.02 (0.02)
Percent fully credentialed	0.12 (0.06)	0.12 (0.05)	0.11 (0.05)	0.15 (0.04)	0.11 (0.03)
EL teacher access	−10.85 (9.55)	−12.18 (9.85)	−5.94 (8.34)	2.19 (3.76)	2.74 (3.17)
Percent new students	−0.03 (0.06)	−0.13 (0.06)	−0.06 (0.06)	0.07 (0.05)	0.05 (0.04)
Average CST score	−0.13 (0.05)	−0.16 (0.04)	−0.11 (0.04)	−0.12 (0.03)	−0.09 (0.03)
Observations	114,259	121,637	115,188	112,624	100,010

Notes: Standard errors clustered by school are in parentheses. Each column contains results from a separate model. Each model also contains controls for age, grade, squared teacher experience, and dummy variables for missing variables.

bilingual education in schools with higher concentrations of students speaking the same non-English language. For all grades, schools with higher California Standards Test (CST) scores, a test of academic achievement in English and language arts, correspond with substantially lower participation.

7. OLS MODEL AND RESULTS

The pattern of results from the participation probit suggests that students in bilingual programs differ from students in other instructional programs along many dimensions. In order to control for these differences, the starting point for the analysis is an OLS regression, as shown in equation 1:

$$ENGPROF_i = BILINGUAL_i\alpha + X_i\beta + \varepsilon_i. \quad (1)$$

In this equation, $ENGPROF_i$ is student i's scale score from the current school year of the CELDT (2003 or 2004). Recall that separate models are estimated for each grade. Because students in kindergarten and first grade take only the listening/speaking portion of the test, the only outcome variable for grades 1 and 2 is the listening/speaking scale score. For grades 3–5, separate models are estimated for each of the three portions: listening/speaking, reading, and writing.

$BILINGUAL_i$ is a dummy variable for whether the student receives academic instruction in his or her native language (Spanish, as the sample is limited to Spanish speakers). In other words, the comparison group contains EL students who do not receive academic instruction in their native language.

X_i can be partitioned into three components (along with a dichotomous variable for year and a constant term), $LAGPROF_i$ (lagged English proficiency), $STUDENT_i$, and $SCHOOL_i$. $LAGPROF_i$ contains the scale scores from the previous year's CELDT. Again, for grades 1 and 2, $LAGPROF_i$ has only one test score, the previous year's listening/speaking portion. For grades 3–5, $LAGPROF_i$ contains three test scores, one for each portion of the test. This structure allows for a more comprehensive effect of lagged English proficiency.

$STUDENT_i$ is a vector of student-level characteristics such as sex, special education receipt, and participation in the Title I program (a proxy for family income), as listed in tables 1 and 2. $SCHOOL_i$ is a vector of school-level characteristics from the previous school year.[11]

The OLS results are in table 3. The top panel contains the results for the listening and speaking test, the middle panel has the results for the reading test, and the bottom panel has the results for the writing test. Each table contains a separate column for each grade. Appendix tables A.1a and A.1b contain the results for the covariates other than bilingual education.

Bilingual education has a negative coefficient for the listening/speaking test, although the coefficient is much larger for grades 1 and 2. Specifically, the

11. Because the CELDT is administered toward the beginning of the school year, most of the value added since the last administration of the CELDT occurred in the previous school year. For example, the 2004 CELDT contains school characteristics from the 2003–4 school year. There is an implicit assumption that students attended their current school in the previous school year. The results are similar if I use current school year characteristics instead.

Table 3. OLS Results for English Proficiency, by Test and Grade

	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5
Dependent variable—listening/speaking proficiency					
Bilingual education	−27.00	−15.45	−2.29	−2.08	−2.54
	(1.14)	(0.68)	(0.70)	(0.71)	(1.10)
	[−0.46]	[−0.27]	[−0.04]	[−0.03]	[−0.04]
Dependent variable—reading proficiency					
Bilingual education			3.60	0.45	−1.78
			(0.43)	(0.45)	(0.58)
			[0.07]	[0.01]	[−0.04]
Dependent variables—writing proficiency					
Bilingual education			−5.43	−3.64	−2.99
			(0.49)	(0.51)	(0.63)
			[−0.10]	[−0.07]	[−0.06]
Observations	224,608	241,005	235,584	222,607	201,386

Notes: Standard errors clustered by school are in parentheses. The brackets measure the bilingual effect in standard deviations of the dependent variable. Unlike previous tables, each cell (coefficient, standard error, and standard deviation effect) is from a separate regression. See the variables and notes to appendix tables A.1a and A.1b for a more complete list of additional control variables.

effect is −0.46 standard deviations for grade 1 and −0.27 standard deviations for grade 2. In contrast, the coefficient is less than 0.05 standard deviations in grades 3–5. In other words, the negative association between bilingual education and listening/speaking proficiency is concentrated in first and second grades; by third grade, the effect is extremely small.

The coefficients for bilingual education in the reading proficiency regressions vary by grade. In third grade, bilingual education is associated with a small, positive effect of 0.07 standard deviations. In fourth grade, the coefficient is not statistically different from zero at conventional levels of 10 percent or less. In fifth grade, bilingual education is associated with a small, negative effect of −0.04 standard deviations. This pattern of results suggests that bilingual education does not have a strong association—either positive or negative—with reading test scores.

The bottom panel of table 3 illustrates that bilingual education has a modest, negative association with writing proficiency. The coefficient is negative and significant in all years and grades. In third grade, bilingual education is associated with a 0.10 standard deviation decrease in writing proficiency. For fourth grade, the effect size is a 0.07 standard deviation decrease, and in fifth grade the effect size is a 0.06 standard deviation decrease.

Bilingual education is not the only determinant of English proficiency, as illustrated by the results in appendix tables A.1a and A.1b. As expected, previous

Table 4. School Fixed Effects Results for English Proficiency, by Test and Grade

	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5
Dependent variable—listening/speaking proficiency					
Bilingual education	−24.55	−16.22	−4.41	−4.69	−6.20
	(0.40)	(0.40)	(0.39)	(0.48)	(0.62)
	[−0.42]	[−0.28]	[−0.08]	[−0.08]	[−0.10]
Dependent variable—reading proficiency					
Bilingual education			2.02	−0.60	−3.29
			(0.33)	(0.35)	(0.43)
			[0.04]	[−0.01]	[−0.07]
Dependent variables—writing proficiency					
Bilingual education			−5.51	−3.71	−4.29
			(0.33)	(0.35)	(0.43)
			[−0.10]	[−0.07]	[−0.09]
Observations	224,608	241,005	235,584	222,607	201,386

Notes: Standard errors are in parentheses. The brackets measure the bilingual effect in standard deviations of the dependent variable. Each cell (coefficient, standard error, and standard deviation effect) is from a separate regression. See the variables and notes to appendix tables A.1a and A.1b for a more complete list of additional control variables.

test scores have strong positive effects: students do not lose English proficiency from one year to the next. Females and students who have spent more time in their current school are associated with higher test scores, whereas disadvantaged students—Title I students and students receiving special education services—are associated with lower test scores.

At the school level, average test scores (of non-EL students) in English/language arts are also positively associated with English proficiency. The school's percentage of EL students is associated with lower listening and speaking proficiency. The percent of teachers who are fully credentialed has a positive association with reading proficiency and writing proficiency. Other school characteristics, including the access to teachers specifically authorized to teach EL students, do not have consistent effects on English proficiency.

Table 4 contains the results from a school fixed effects estimator. This model controls for between-school differences in fixed (observable and unobservable) factors. Rossell (2005) illustrates how California schools had great variation in defining bilingual education and other EL instructional programs. In other words, two schools may offer the same program, but one school calls it bilingual education and one school calls it something else. Rather than identifying the bilingual education effects of such situations, the school fixed effects model is identified only off of within-school differences in bilingual education. The results are similar to the OLS results in table 3, although the magnitudes of the effect may vary slightly. These findings suggest that between-school variation in bilingual education programs is not driving the OLS results.

8. PROPENSITY SCORE MODELS AND RESULTS

In this section, propensity score analysis provides a robustness check on the OLS results in table 3. For bilingual education, the propensity score equals the likelihood of participating in bilingual education, calculated as the predicted probability from the probit equation of bilingual education receipt reported in table 2. The propensity scores obtained from a logit model are nearly identical. The independent variables are defined as in equation 1.[12]

Figure 1 illustrates the predicted propensity scores for fall 2004 separately by grade. As expected, propensity scores are higher for students who participate in bilingual education than for those who do not. The area of common support is 100 percent for both cohorts (and for 2003, although not shown in the figure). Both of these factors are consistent with a valid propensity score estimator, but the validity of the estimator ultimately hinges on the assumption that observables capture all relevant determinants of bilingual education participation.

Propensity scores dominate the matching literature. However, there are many types of propensity score estimators, and no single estimator dominates the literature (see, for example, Frolich 2004). Because the focus of the article is on the effectiveness of bilingual education, not propensity score estimators, I employ three common methods: kernel density, nearest neighbor, and caliper.[13] Essentially, the kernel density estimator creates a weighted average of the comparison group (in this case, EL students not in bilingual education) for each student in the treatment group (in this case, EL students in bilingual education).[14]

In the nearest neighbor approach, each student in bilingual education is matched with the four students in other EL programs who have the most similar propensity scores. I choose four because simulation results in Abadie and Imbens (2002) worked best, in terms of mean-squared error, for four neighbors, although my results for four are quite similar to results with other numbers of matches (from one to five).[15]

12. Researchers often use interaction terms or higher-ordered terms (for continuous independent variables) in order to improve the goodness of fit of the probit (or logit) generating the propensity score. However, there is no theoretical basis for including these extra terms. Their goal is to eliminate any statistically significant differences in variables between the treatment and control groups. However, given the size of my sample, it is extremely difficult to eliminate *all* statistically significant differences between students in bilingual education and other EL students.
13. I use a matching program in Stata called psmatch2 (see Leuven and Sianesi 2003) to calculate the propensity score estimators. For a recent review of matching estimators, with an application to job training programs, see Mueser, Troske, and Gorislavsky (2007) and the references cited therein.
14. I use the default options in psmatch2: an Epanechnikov kernel (as recommended by Silverman 1986) with a bandwidth of 0.08.
15. I match with replacement, and I include ties (i.e., students with identical propensity scores), even if including them raises the number of closest neighbors above four. The results are not sensitive to either of these assumptions.

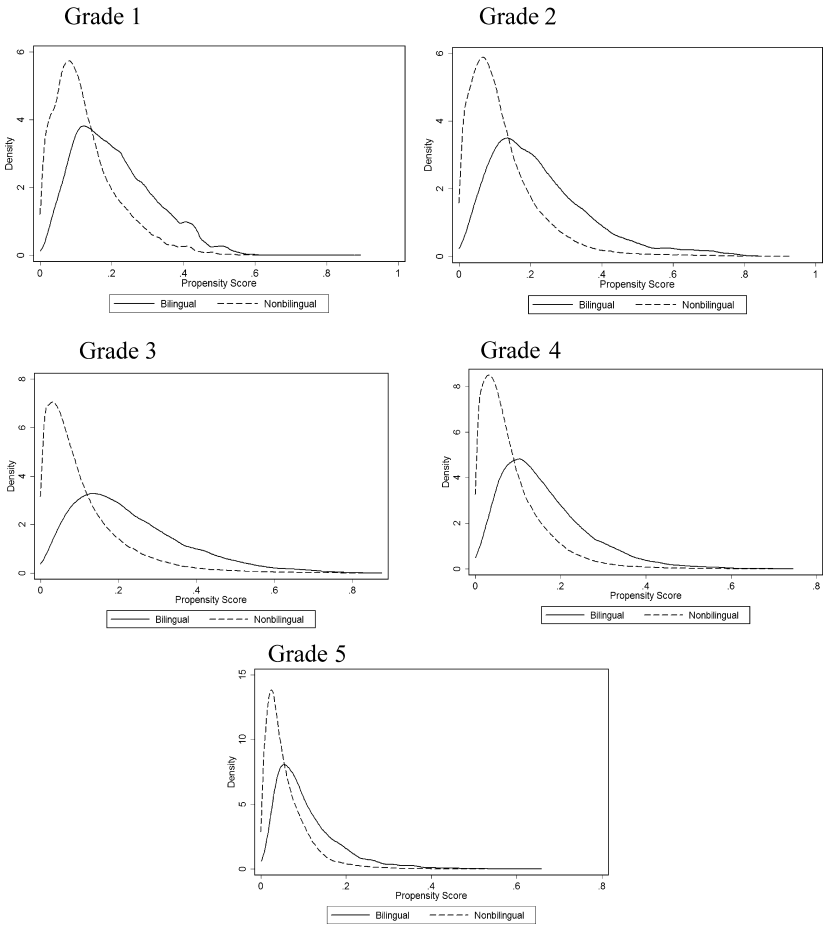

Figure 1. Propensity Scores by Grade and Bilingual Education Participation

The third method is a caliper or radius method. I match each student in bilingual programs with all students in other EL programs who have a propensity score with a given amount, called the radius. I choose a radius of 0.000004 (which produces a similar average number of matches to the four nearest neighbors), although results with radii of 0.000001 to 0.000005 produce nearly identical results.

Table 5 contains the results from all three types of propensity score estimators. The reported results are simple differences in means between the two groups, the full set of students in bilingual education and the matched set of EL students in other programs. Regression-adjusted estimates are similar to the simple averages and therefore are not reported. The results do not account for the variation resulting from the matching process. Given the similarity of the results to the OLS results for the full sample, it is extremely doubtful that

Table 5. Propensity Score Estimators for English Proficiency, by Test and Grade

	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5
Dependent variable—listening/speaking proficiency					
Kernel density	−30.05	−19.24	−3.50	−2.05	−5.05
	(0.62)	(0.53)	(0.58)	(0.68)	(0.90)
	[−0.51]	[−0.33]	[−0.06]	[−0.03]	[−0.08]
Four nearest neighbors	−28.65	−16.59	−2.10	−1.74	−2.70
	(0.49)	(0.42)	(0.44)	(0.53)	(0.72)
	[−0.49]	[−0.29]	[−0.04]	[−0.03]	[−0.04]
Radius ≤ 0.00004	−25.55	−14.53	−3.12	−3.64	−3.28
	(0.50)	(0.43)	(0.47)	(0.54)	(0.70)
	[−0.44]	[−0.25]	[−0.06]	[−0.06]	[−0.05]
Dependent variable—reading proficiency					
Kernel density			1.92	−1.09	−4.79
			(0.47)	(0.52)	(0.67)
			[0.04]	[−0.02]	[−0.09]
Four nearest neighbors			2.82	0.24	−2.31
			(0.37)	(0.42)	(0.54)
			[0.06]	[0.005]	[−0.05]
Radius ≤ 0.00004			3.64	−0.56	−2.10
			(0.41)	(0.44)	(0.55)
			[0.07]	[−0.01]	[−0.04]
Dependent variables—writing proficiency					
Kernel density			−6.34	−4.27	−4.90
			(0.53)	(0.54)	(0.68)
			[−0.12]	[−0.08]	[−0.10]
Four nearest neighbors			−5.65	−3.75	−3.34
			(0.41)	(0.44)	(0.57)
			[−0.10]	[−0.07]	[−0.07]
Radius ≤ 0.00004			−6.27	−4.89	−3.50
			(0.44)	(0.45)	(0.55)
			[−0.12]	[−0.10]	[−0.07]

Notes: Simple standard errors (i.e., not bootstrapped) are in parentheses. These standard errors are used for calculating statistical significance. The brackets measure the bilingual effect in standard deviations of the dependent variable. Each cell (coefficient, standard error, and standard deviation effect) is from a separate propensity score estimator.

controlling for that variation will substantially alter the findings. Furthermore, Abadie and Imbens (2008) show that bootstrapping the standard errors is not necessarily appropriate.

The estimators in table 5 are quite similar to the OLS estimators (measured in standard deviation effects) for the full sample (table 3). With one exception, the full sample OLS standard deviation effect lies between the three matching estimators.[16] For example, the full sample OLS coefficient is −0.46 standard

16. In that exception, grade 4 reading proficiency, the estimates from the propensity score are from −0.02 to 0.005 standard deviations, whereas the OLS estimate is 0.01 standard deviations.

deviations for first-grade students (table 3, top panel, column 1). This coefficient falls between the kernel density estimate of −0.51 and the radius/caliper estimate of −0.44 (both from table 5, column 1). There are several potential explanations for the similarity between matching estimators and OLS coefficients. The area of common support is 100 percent for the kernel and nearest neighbor matching but falls to roughly 75–90 percent for radius matching. This high percentage suggests that propensity score samples may be similar, but not identical, to the overall sample (controlling for observables). A related explanation is that selection on observables is largely linear. In this case, OLS models control for selection, and therefore the propensity score, a nonlinear function of observables, provides little additional information.

A third explanation is that the matching procedure based on the propensity score is not appropriate. Propensity scores are convenient because they reduce the matching problem to one variable, the propensity score. Matching across multiple variables—that is, directly on the observables (or a subset of them) rather than on a nonlinear function of them—is difficult to implement successfully (Smith and Todd 2005). For example, Mueser, Troske, and Gorislavsky (2007) and Frolich (2004) find that Mahalanobis distance methods perform worse than propensity score methods.[17] Therefore I do not pursue multidimensional matching estimators.

9. INSTRUMENTAL VARIABLES MODELS AND RESULTS

One concern with the OLS and matching estimators is that the bilingual education coefficient may be picking up the effect of other variables that cannot be measured. The results in this section use an IV approach to address this concern using two instruments based on school-level changes in the percentage of EL students participating in bilingual education as a result of Proposition 227.[18] For more information on the IV technique, see Jepsen (2009). In order for these instruments to be valid, they must be correlated with bilingual education participation at the student level, but they must not be directly correlated with student-level English proficiency.

The IV results for these two instruments are in table 6. The table reports the full model, where each cell is from a separate IV model. The layout is the same

17. Similarly, Frolich (2004) finds that using the odds ratio of the propensity score, as suggested by Imbens (2000), performed more poorly than directly using propensity score methods.

18. The first instrument is the mechanical change in the school-level percentage of EL students participating in bilingual education as a result of Proposition 227. As in Gordon and Hoxby (2002), the mechanical percentage is set to zero for the post-proposition period. The pre-proposition measure is the school-level percentage of students enrolled in bilingual education in spring 1998 (i.e., right before Proposition 227). Because the post-proposition percentage is set to zero, the mechanical change (i.e., post-proposition minus pre-proposition) is equal to the opposite of the pre-proposition percentage. The second instrument is the actual change in the percentage bilingual between spring 1998 and spring 1999 (at the school level).

Table 6. Instrumental Variables Models for English Proficiency, by Test and Grade

	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5
Dependent variable—listening/speaking proficiency					
Bilingual education	−38.46	−18.94	4.02	3.86	6.50
	(2.53)	(1.79)	(1.91)	(2.49)	(4.22)
	[−0.65]	[−0.33]	[0.07]	[0.06]	[0.10]
Dependent variable—reading proficiency					
Bilingual education			10.15	4.15	1.61
			(1.16)	(1.24)	(1.90)
			[0.21]	[0.08]	[0.03]
Dependent variables—writing proficiency					
Bilingual education			−5.10	−4.56	3.04
			(1.18)	(1.31)	(2.04)
			[−0.09]	[−0.09]	[−0.06]
Observations	218,746	234,528	229,311	216,339	191,706

Notes: Instruments are measures of Proposition 227–induced changes in school-level bilingual education availability. Standard errors clustered by school are in parentheses. The brackets measure the bilingual effect in standard deviations of the dependent variable. Each cell (coefficient, standard error, and standard deviation effect) is from a separate regression. See the variables and notes to appendix tables A.1a and A.1b for a more complete list of additional control variables.

as table 4 (the results for the school fixed effects model). Bilingual education has a large, negative association with listening and speaking proficiency in first and second grades. The IV coefficients show no evidence of a negative effect for listening and speaking proficiency in grades 3–5, as the coefficients are always positive and in some cases statistically significant. For reading proficiency, bilingual education has a positive association for students in grades 3 and 4, and the coefficient for grade 5 is not statistically different from zero. For writing proficiency, bilingual education has a modest, negative association for grades 3 and 4, and the coefficient for grade 5 is not statistically different from zero.

In sum, the results are generally consistent across all four methods (OLS, school fixed effects, propensity score, and IV). Compared with other EL programs, bilingual education has a sizable, negative association with English listening/speaking proficiency for EL students in first and second grades. For EL students in grades 3–5, bilingual education has a small and often insignificant association with English proficiency.

10. BILINGUAL EDUCATION BY PREVIOUS PROFICIENCY SCORE

As seen above, EL students in bilingual education have dramatically lower English listening and speaking proficiency than EL students in other EL programs in first and second grades. Bilingual education has small and often insignificant associations with English proficiency for students in grades 3–5. These findings are robust across years and across estimation methods. This

Table 7. OLS Results for Listening/Speaking Proficiency, by Grade: Bilingual Education Interacted with Previous Listening/Speaking Quartile

	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5
Dependent variable—listening/speaking proficiency					
Bilingual * 1st quartile	−42.08	−27.22	−13.11	−10.13	−10.35
	(1.76)	(0.85)	(0.87)	(0.87)	(1.23)
Bilingual * 2nd quartile	−31.78	−9.85	4.42	1.13	2.04
	(1.18)	(0.67)	(0.80)	(0.87)	(1.38)
Bilingual * 3rd quartile	−9.01	−3.42	8.98	6.39	4.74
	(0.91)	(0.80)	(0.83)	(1.01)	(1.68)
Bilingual * 4th quartile	3.35	−0.04	8.51	7.03	2.13
	(1.20)	(1.24)	(1.09)	(1.37)	(2.22)
Observations	224,608	241,005	235,584	222,607	201,386

Notes: Standard errors clustered by school are in parentheses. Each column is from a separate regression. See the variables and notes to appendix tables A.1a and A.1b for a more complete list of additional control variables.

section contains an exploration of whether bilingual education has differential effects based on students' prior English proficiency levels.

Specifically, I estimate an OLS model similar to equation 1, except for the addition of variables interacting the bilingual education variable with dummy variables for each quartile of the previous year's listening/speaking score. The results for the listening/speaking test are in table 7, and results for reading and writing are in appendix table A.2.[19]

The results for all three English proficiency tests show the same pattern: the bilingual education coefficient improves dramatically with the quartile of lagged listening and speaking proficiency. In other words, the students who appear to benefit the most from bilingual education relative to other EL programs are the students who had strong English listening/speaking proficiency in the previous year. For students in the lowest quartile, bilingual education is associated with lower listening and speaking test scores of more than forty points (approximately 0.7 standard deviations) among first-grade students. For students in the third-lowest quartile, the bilingual coefficient is only −9 points (0.06 standard deviations). Other grades also have a positive relationship between lagged listening and speaking proficiency and the bilingual coefficient, although the range of effects is much smaller.

11. ATTRITION

One concern with the results in all the tables is the relationship between student attrition and bilingual education. By using lagged CELDT scores, the

19. The results are similar when the quartile interaction terms are replaced with a single interaction term between bilingual education and the previous year's listening/speaking test score.

sample is restricted to students who took the CELDT in the current year and in the previous year. If attrition from one year to the next is systematically related to bilingual education status, the estimates in the table are not valid. There are two sources of attrition in the CELDT data. First, attrition occurs if a student leaves the California public school system (CELDT test score data are available for students who switch public schools within California). Second, attrition occurs if a student is no longer considered an EL. Once a student is considered proficient in English, that student is no longer considered an EL and stops taking the CELDT.

Because the data are a repeated cross section, I cannot directly measure attrition. Instead, I calculate attrition indirectly by attempting to merge the fall 2003 cross section with the fall 2004 cross section. Student identifiers are not available, so students are matched based on school, grade, gender, and birth date. Specifically, the attrition analysis is limited to students in the fall 2003 cross section who can be uniquely identified based on the matching variables (school, grade, gender, and birth date). The focus is on students in grades K–4 in 2003 because these students would be in grades 1–5 in the fall 2004 cohort. Table 8 contains the percentage of students who were successfully matched in the fall 2004 cohort. The top panel combines all students, whereas the bottom two panels provide separate information by bilingual status.

The table shows that substantial attrition occurs: 65 percent of students in the fall 2003 cohort are in the fall 2004 cohort. Students in bilingual programs have lower attrition than students in other EL programs, although the difference declines as the grade increases. The results are similar for each quartile of the 2003 listening/speaking test.[20] This pattern of results is not consistent with the hypothesis that the large, negative bilingual education coefficients in grades 1 and 2 and small, often insignificant, bilingual coefficients in grades 3–5 occur because the high-achieving, non-bilingual students in grades 3–5 were more likely to become proficient in English. On the other hand, the lower attrition for students in bilingual programs is not consistent with the hypothesis that the bilingual education coefficients are biased down because bilingual education students are economically disadvantaged and therefore more likely to leave California schools. Thus the table suggests that the results in this article are not being driven by systematic attrition of EL students.

12. DISCUSSION

This article studies the causal effect of bilingual education versus other EL programs on English proficiency. The comparison group is the set of students

20. Note that the quartile is calculated for all students in the appropriate grade. Separate quartiles are not calculated for students in bilingual programs and for students in other EL programs.

Table 8. Percentage of Students in Fall 2003 Cohort Who Are in Fall 2004 Cohort, by EL Program, Grade, and Fall 2003 Listening/Speaking Quartile

	Fall 2003 Grade					
	K–4	**K**	**1**	**2**	**3**	**4**
All students						
All quartiles	65.1%	52.0%	69.9%	70.5%	66.9%	64.6%
1st quartile	55.2%	41.6%	64.2%	67.3%	64.5%	65.6%
2nd quartile	68.5%	53.2%	72.5%	73.7%	72.3%	71.7%
3rd quartile	71.7%	55.8%	72.8%	73.3%	71.8%	66.7%
4th quartile	65.0%	56.5%	70.3%	69.1%	61.8%	57.1%
Not bilingual						
All quartiles	64.5%	51.3%	69.2%	70.0%	66.5%	64.7%
1st quartile	53.4%	40.0%	61.5%	66.0%	63.8%	66.0%
2nd quartile	67.7%	52.3%	71.9%	73.2%	72.1%	72.0%
3rd quartile	71.5%	55.2%	72.3%	72.8%	71.7%	66.8%
4th quartile	64.8%	56.1%	69.9%	68.7%	61.4%	57.2%
Bilingual						
All quartiles	70.1%	60.6%	74.9%	74.4%	70.3%	63.5%
1st quartile	65.0%	54.4%	71.9%	71.6%	68.1%	63.1%
2nd quartile	73.9%	61.9%	77.4%	76.3%	74.5%	69.3%
3rd quartile	74.5%	65.5%	78.9%	78.1%	73.3%	65.8%
4th quartile	69.1%	65.3%	76.5%	76.1%	67.1%	55.6%

receiving other EL programs because most educators and policy makers agree that these EL students need some sort of EL program. Not offering them any EL program is no more plausible an option than not offering them a mathematics program. The study uses a unique administrative data set of California EL students in elementary grades. Students are not randomly assigned to bilingual education: in California, Proposition 227 requires parental permission slips before students are allowed in bilingual education classrooms. In addition to OLS, I use propensity score analysis to control for selection on observables and school fixed effects to control for between-school attributes that do not vary across students.

The results from the propensity score technique and the school fixed effects model are similar to the OLS results. In grades 1 and 2, students in bilingual education programs have lower listening and speaking proficiency scores than other EL students. The size of this effect is large—over 0.3 standard deviations in many cases. Furthermore, the effect of bilingual education is inversely related to students' previous listening and speaking proficiency. Students with low English listening and speaking proficiency in the previous year have particularly large, negative effects of bilingual education, whereas students with higher levels of English proficiency suffer little if any negative consequences of bilingual education. Policy makers should consider this adverse

effect on English listening and speaking when evaluating bilingual education programs because they represent a cost to the youngest and weakest students.

For students in grades 3–5, the effects of bilingual education versus other EL programs are much smaller. In several cases, the effects on English proficiency are insignificant or even positive. In terms of standard deviations, nearly all the effects are under 0.1 standard deviations, and many are under 0.05 standard deviations. Although these effects are sometimes statistically significant, their small size suggests fewer concerns about the negative consequences of bilingual education on English proficiency for older students.

Why would bilingual education have a large, negative effect of English listening and speaking proficiency for students in early elementary grades but smaller (and sometimes insignificant) effects in later grades? Proposition 227 requires that EL students enroll in English-only classrooms for thirty days when they first enroll in California schools. English-only classrooms are much more similar to English-based EL programs than bilingual programs, so students in bilingual programs lose thirty days of access to bilingual programs in their first year. Most students in bilingual programs enter California schools in kindergarten or first grade.

Bilingual programs in kindergarten and first grade typically use primary language much more extensively than English; the percentage of instruction in English gradually increases so that by fifth grade at least half the instruction is in English (Ovando and Collier 1998). For example, schools in San Diego County and in San Francisco use such an approach in their bilingual programs (Gold 2006; SFUSD 2008). If this approach is common in California, lower English proficiency gains might be expected for bilingual education students in first and second grades compared with other EL programs where a higher proportion of overall instruction takes place in English.

Because student-level academic achievement data are not available, this article looks at only one outcome: English proficiency as measured by the CELDT data. Bilingual education has many goals, including overall academic proficiency. Still, English proficiency in the younger grades is essential for overall academic achievement. Although bilingual education is not as successful with regard to English proficiency as other EL programs for early elementary school students, it may have other benefits that are not captured in the administrative data used here. Previous work finds mixed effects for academic achievement (for example, Gordon and Hoxby 2002; Lopez 2003). Further research on all relevant outcomes (not just English proficiency and academic achievement) is needed.

I thank the editors, two anonymous referees, Laura Brien, Arturo Gonzalez, Scott Hankins, Brian Jacob, Lisa Jepsen, Robert McMillan, Abigail Payne, Heather Rose, Jon Sonstelie, Christopher Taber, Kenneth Troske, and seminar participants at the Bureau

of Labor Statistics, the Public Policy Institute of California, the University of California, Santa Barbara, and the University of Kentucky for useful comments. Thanks also to Mark Fetler and Michelle Center for providing access to—and assistance in using—restricted-use data from the California Department of Education. Shelley de Alth and Ria Sengupta provided outstanding research assistance. All opinions and mistakes are mine alone.

REFERENCES

Abadie, Alberto, and Guido W. Imbens. 2002. Large sample properties of matching estimators for average treatment effects. National Bureau of Economic Research Technical Working Paper No. 283.

Abadie, Alberto, and Guido W. Imbens. 2008. On the failures of the bootstrap for matching estimators. *Econometrica* 76(6): 1537–57.

August, Diane, and Timothy Shanahan. 2006. *Developing literacy in second language learners: Report of the National Literacy Panel on Language Minority Children and Youth*. New York: Lawrence Erlbaum Associates.

Bleakley, Hoyt, and Aimee Chin. 2004. Language skills and earnings: Evidence from childhood immigrants. *Review of Economics and Statistics* 86 (2): 481–96.

Cheng, Chang-Kuo. 1996. Impact of bilingual education on educational attainment. PhD dissertation, University of Maryland, College Park, MD.

Federman, Maya. 2000. Bilingual education, college major choice, and college remediation. PhD dissertation, Harvard University, Cambridge, MA.

Frolich, Markus. 2004. Finite-sample properties of propensity-score matching and weighting estimators. *Review of Economics and Statistics* 86 (1): 77–90.

Genesee, Fred, Kathryn Lindholm-Leary, William M. Saunders, and Donna Christian. 2006. *Educating English language learners: A synthesis of research evidence*. New York: Cambridge University Press.

Gold, Norm. 2006. Successful bilingual schools: Six effective programs in California. San Diego: San Diego County Office of Education.

Gonzalez, Arturo. 2000. The acquisition and labor market value of four English skills: New evidence from the NALS. *Contemporary Economic Policy* 18 (3): 259–69.

Gordon, Nora, and Caroline Hoxby. 2002. Achievement effects of bilingual education versus English immersion: Evidence from California's Proposition 227. Unpublished paper, University of California, San Diego.

Greene, Jay. 1998. A meta-analysis of the effectiveness of bilingual education. Los Angeles: Tomas Rivera Policy Institute.

Imbens, Guido. 2000. The role of the propensity score in estimating dose-response functions. *Biometrica* 87: 706–10.

Jepsen, Christopher. 2009. Bilingual education and English proficiency. University of Kentucky Center for Poverty Research Discussion Paper No. DP2009-01. Available www.ukcpr.org/Publications/DP2009-01.pdf. Accessed 20 August 2009.

Legislative Analyst's Office (LAO). 2004. *A look at the progress of English learner students*. Sacramento, CA: LAO.

Leuven, Edwin, and Barbara Sianesi. 2003. Psmatch2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. Available http://ideas.repec.org/c/boc/bocode/s432001.html. Accessed 12 August 2009.

Lopez, Mark Hugo. 2003. Do English language assistance programs affect educational attainment and labor market outcomes? Evidence from the National Education Longitudinal Study of 1988 and High School and Beyond. Unpublished paper, University of Maryland.

Lopez, Mark Hugo, and Marie T. Mora. 1998. The labor-market effects of bilingual education among Hispanic workers. *READ Perspectives* 5: 33–51.

Matsudaira, Jordan. 2005. Sinking or swimming? Evaluating the impact of English immersion versus bilingual education on student achievement. Unpublished paper, Cornell University.

Mora, Marie T. 2000. English-language assistance programs, English skill acquisition, and the academic progress of high school language minority students. *Policy Studies Journal* 28 (4): 721–38.

Mueser, Peter R., Kenneth R. Troske, and Alexey Gorislavsky. 2007. Using state administrative data to measure program performance. *Review of Economics and Statistics* 89 (4): 761–83.

Ovando, Carlos J., and Virginia P. Collier. 1998. *Bilingual and ESL classrooms: Teaching in multicultural contexts*. San Francisco: McGraw Hill.

Padolsky, Daniel. 2005. National Clearinghouse for English Language Acquisition and Language Instruction Educational Programs. FAQ [No. 1]. Available www.ncela.gwu.edu/faqs/. Accessed 24 September 2009.

Parrish, Thomas B., Amy Merickel, María Pérez, Robert Linquanti, Miguel Socias, Angeline Spain, Cecilia Speroni, Phil Esra, Leslie Brock, and Danielle Delancey. 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; San Francisco: WestEd.

Rossell, Christine H. 2005. The educational effectiveness of bilingual education. *Education Next* 3 (4): 44–52.

San Francisco Unified School District (SFUSD). 2008. Multilingual programs in SFUSD. San Francisco: SFUSD.

Silverman, Bernard W. 1986. *Density estimation for statistics and data analysis*. New York: Chapman and Hall.

Smith, Jeffrey A., and Petra E. Todd. 2005. Does matching overcome LaLonde's critique of nonexperimental estimators? *Journal of Econometrics* 125 (1–2): 305–53.

Trejo, Stephen J. 2003. Intergenerational progress of Mexican-origin workers in the U.S. labor market. *Journal of Human Resources* 38 (3): 467–89.

APPENDIX

Table A.1a. OLS Results for Listening/Speaking Proficiency, by Grade

	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5
Bilingual education	−27.00	−15.45	−2.29	−2.08	−2.54
	(1.14)	(0.68)	(0.70)	(0.71)	(1.10)
Lagged listening/speaking	0.305	0.408	0.240	0.281	0.239
	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)
Lagged reading			0.227	0.224	0.228
			(0.004)	(0.004)	(0.005)
Lagged writing			0.152	0.141	0.160
			(0.004)	(0.004)	(0.005)
Female	1.485	2.054	1.231	1.679	1.170
	(0.216)	(0.206)	(0.195)	(0.214)	(0.240)
Title I	−1.403	−1.597	−1.739	−1.241	−1.775
	(0.659)	(0.607)	(0.598)	(0.643)	(0.669)
Special education	−25.43	−19.75	−17.63	−12.99	−9.86
	(0.70)	(0.54)	(0.43)	(0.44)	(0.47)
Years in current school	4.781	1.651	0.760	0.640	0.583
	(0.586)	(0.242)	(0.144)	(0.119)	(0.096)
Language index	3.394	1.154	2.102	2.641	0.734
	(1.562)	(1.432)	(1.395)	(1.532)	(1.732)
Class size	0.268	0.015	−0.337	−0.082	−0.140
	(0.203)	(0.218)	(0.138)	(0.128)	(0.087)
Percent EL	−0.202	−0.132	−0.129	−0.109	−0.099
	(0.018)	(0.015)	(0.014)	(0.016)	(0.017)
Average teacher experience	−0.011	−0.250	0.014	0.624	−0.181
	(0.522)	(0.411)	(0.446)	(0.483)	(0.515)
Percent BA or less	−0.064	0.008	−0.018	0.027	−0.009
	(0.028)	(0.025)	(0.024)	(0.025)	(0.028)
Percent BA + 30	−0.073	−0.033	0.004	0.031	−0.011
	(0.020)	(0.017)	(0.017)	(0.018)	(0.020)
Percent fully credentialed	−0.008	0.016	−0.026	−0.030	0.009
	(0.041)	(0.031)	(0.031)	(0.040)	(0.040)
EL teacher access	2.252	2.542	2.045	7.072	6.154
	(2.077)	(1.727)	(1.725)	(2.178)	(2.597)
Percent new students	−0.108	−0.008	−0.044	−0.044	0.0001
	(0.043)	(0.025)	(0.024)	(0.028)	(0.032)
Average CST score	0.103	0.127	0.104	0.094	0.110
	(0.024)	(0.018)	(0.018)	(0.020)	(0.022)
2003–4 school year	1.845	4.193	6.112	0.069	−0.653
	(0.499)	(0.462)	(0.410)	(0.494)	(0.521)
Observations	224,608	241,005	235,584	222,607	201,386

Notes: Standard errors clustered by school are in parentheses. R^2s range from 0.29 to 0.35. All models also contain controls for squared teacher experience, age, grade, and dummy variables for missing variables (except bilingual education).

Table A.1b. OLS Results for Reading and Writing Proficiency, by Grade

	Reading			Writing		
	Grade 3	**Grade 4**	**Grade 5**	**Grade 3**	**Grade 4**	**Grade 5**
Bilingual education	3.60	0.45	−1.78	−5.43	−3.64	−2.99
	(0.43)	(0.45)	(0.58)	(0.49)	(0.51)	(0.63)
Lagged listening/speaking	0.022	0.025	0.026	0.044	0.029	0.026
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Lagged reading	0.416	0.444	0.467	0.318	0.261	0.243
	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)	(0.004)
Lagged writing	0.217	0.204	0.167	0.346	0.358	0.347
	(0.004)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)
Female	3.132	4.356	4.309	4.139	4.766	4.574
	(0.159)	(0.156)	(0.166)	(0.163)	(0.155)	(0.162)
Title I	−1.770	−1.688	−0.994	−1.094	−0.389	−0.492
	(0.348)	(0.335)	(0.349)	(0.359)	(0.327)	(0.347)
Special education	−16.67	−17.80	−18.93	−24.94	−19.16	−17.01
	(0.41)	(0.35)	(0.36)	(0.49)	(0.39)	(0.38)
Years in current school	−0.237	−0.339	−0.236	0.292	0.214	0.154
	(0.109)	(0.067)	(0.060)	(0.116)	(0.071)	(0.063)
Language index	−1.675	−2.063	−2.459	1.771	0.777	0.120
	(0.880)	(0.759)	(0.767)	(0.944)	(0.806)	(0.833)
Class size	−0.159	0.023	0.048	−0.230	−0.006	0.002
	(0.089)	(0.049)	(0.037)	(0.083)	(0.046)	(0.042)
Percent EL	−0.005	0.012	0.012	−0.007	0.007	0.015
	(0.008)	(0.007)	(0.007)	(0.009)	(0.008)	(0.008)
Average teacher experience	0.141	0.683	−0.187	0.390	0.664	0.019
	(0.256)	(0.237)	(0.238)	(0.276)	(0.251)	(0.252)
Percent BA or less	−0.020	0.014	−0.012	−0.023	0.007	−0.023
	(0.014)	(0.013)	(0.014)	(0.015)	(0.013)	(0.015)
Percent BA + 30	−0.017	0.002	−0.016	0.0002	0.022	0.007
	(0.010)	(0.009)	(0.009)	(0.010)	(0.009)	(0.010)
Percent fully credentialed	0.069	0.071	0.096	0.127	0.109	0.115
	(0.020)	(0.016)	(0.017)	(0.019)	(0.017)	(0.018)
EL teacher access	−2.087	−0.576	−0.389	−1.401	−0.791	0.849
	(0.940)	(1.188)	(1.341)	(1.080)	(1.290)	(1.453)
Percent new students	−0.004	0.006	0.021	0.029	0.006	0.034
	(0.015)	(0.014)	(0.013)	(0.016)	(0.017)	(0.014)
Average CST score	0.108	0.103	0.119	0.107	0.104	0.104
	(0.011)	(0.010)	(0.009)	(0.012)	(0.010)	(0.010)
2003–4 school year	0.981	2.718	0.739	−1.503	1.038	−0.198
	(0.265)	(0.240)	(0.246)	(0.266)	(0.236)	(0.248)
Observations	235,584	222,607	201,386	235,584	222,607	201,386

Notes: Standard errors clustered by school are in parentheses. R^2s range from 0.44 to 0.52. All models also contain controls for squared teacher experience, age, grade, and dummy variables for missing variables (except bilingual education).

Table A.2. OLS Results for Reading and Writing Proficiency, by Grade: Bilingual Education Interacted with Previous Listening/Speaking Quartile

	Grade 3	Grade 4	Grade 5
Dependent variable—reading proficiency			
Bilingual * 1st quartile	−1.06	−2.19	−4.91
	(0.58)	(0.61)	(0.86)
Bilingual * 2nd quartile	4.57	−0.49	−2.60
	(0.56)	(0.58)	(0.68)
Bilingual * 3rd quartile	8.94	3.31	1.13
	(0.60)	(0.61)	(0.82)
Bilingual * 4th quartile	11.19	7.45	4.14
	(0.63)	(0.73)	(1.02)
Dependent variable—writing proficiency			
Bilingual * 1st quartile	−10.05	−7.31	−6.49
	(0.64)	(0.72)	(1.04)
Bilingual * 2nd quartile	−3.07	−2.18	−2.55
	(0.60)	(0.64)	(0.73)
Bilingual * 3rd quartile	−0.17	−0.28	0.46
	(0.60)	(0.63)	(0.82)
Bilingual * 4th quartile	0.21	1.20	1.35
	(0.72)	(0.72)	(0.98)
Observations	235,584	222,607	201,386

Notes: Standard errors clustered by school are in parentheses. The results from each column and subject are from a separate regression (there are a total of twelve regressions). See the variables and notes to table 3 for a more complete list of additional control variables.