preliminary
do not quote without authors' permission

ACHIEVEMENT EFFECTS OF BILINGUAL EDUCATION *VS.* ENGLISH IMMERSION: EVIDENCE FROM CALIFORNIA'S PROPOSITION 227

Nora Gordon and Caroline Hoxby*

Abstract

We examine the effect of bilingual education, as opposed to English immersion, on the achievement of limited English proficient (LEP) and English proficient (non-LEP) students. On the one hand, students enrolled in bilingual education may learn better when taught at least partly in their mother tongue. On the other hand, students in bilingual classes may lose valuable spillovers associated with English-proficient classmates and may learn English more poorly because they do not need English to meet their immediate needs. Many studies of bilingual education have been crippled by selection problems because students assigned to bilingual education tend to be different than other students in many unobservable as well as observable ways. We avoid such problems using the sharp, dramatic change in California's bilingual education policy following the passage of Proposition 227, which banned bilingual education in favor of English immersion. Our identification strategy relies on the sharpness of the change (regression discontinuity), and we are careful to account for potentially confounding factors such as class size reduction and test habituation. We also distinguish several channels by which bilingual education affects LEP and non-LEP students. We generally find that the shift from bilingual education to English immersion had a modest negative effect on LEP students and a modest positive effect on non-LEP students. The effects of the ban on bilingual education are more positive for higher elementary grades, among both LEP and non-LEP students.

* Department of Economics, University of California - San Diego and Department of Economics, Harvard University. The authors welcome comments on this preliminary draft. Please send them to negordon@weber.ucsd.edu or choxby@harvard.edu. The authors especially gratefully acknowledge seminal discussions with Michael Kremer. In addition, the authors thank Lauri Burnham-Massey and David Dolson of the California Department of Education for details of California's policies related to limited English proficiency, the Arizona Department of Education for help with data, and Mireya Almazan for excellent research assistance. Finally, the authors gratefully thank David Cutler, Lawrence Katz, Claudia Goldin, and seminar participants at the American Educational Research Association, Carnegie Mellon University, and Cornell University for comments. All errors are the authors' own.

I. Introduction

In this paper, we investigate how putting limited English proficient (LEP) students into English language classrooms with English proficient (non-LEP) students affects the educational achievement of both groups. When LEP students are taught academic subjects in the English language with English proficient classmates, their experience is called "English immersion." This is opposed to bilingual education, in which LEP students are enrolled in classes that are conducted at least partly in their mother tongue and that include no students who are fully proficient in English.[1] The question of whether LEP students are better off in bilingual education or English immersion is a longstanding and controversial one. On the one hand, LEP students may learn material better when it is taught at least partly in their mother tongue. On the other hand, when LEP students are isolated in bilingual classrooms, they may lose valuable resources and spillovers associated with English-proficient peers. They may also learn English poorly because they are less exposed to English speakers and have little incentive to learn English to meet immediate needs. Because many career opportunities and advanced courses require English proficiency, students who do not learn English well may eventually be denied access to advanced material and/or have little incentive to learn it. There are other trade-offs associated with bilingual education and immersion, but these are the main ones. We do not think that it is possible to resolve the controversy over bilingual education using theory alone, so in this paper we present empirical evidence.

Specifically, we examine the effects of Proposition 227, which mandated a switch from bilingual education to English immersion for Californian LEP students, beginning in the 1998-99 school year. Proposition 227 provides us with a uniquely good opportunity to evaluate bilingual education because it was a major policy change that happened very quickly. The proposition was enacted, by referendum, in June 1998. Although it did not exist in the 1997-98 school year, the proposition was fully implemented in the fall of the 1998-99 school year. California's policy shock (to which we hereafter refer as the "ban" on bilingual education) is very useful because it allows us to avoid identification problems that have crippled previous studies of the effects of bilingual education.

As a rule, previous studies have suffered from the following problem. Students are enrolled in bilingual

[1]Throughout the paper, we follow California in defining bilingual enrollments as the number of students receiving instruction in at least two academic subjects (as distinct from English as a second language) in their native language. We refer to other instructional settings as English immersion generally; this includes sheltered English immersion and other variants.

classes because they have a mother tongue other than English and do poorly on English language tests (this is the definition of LEP), because their teachers (and sometimes also parents) judge them to be suitable candidates for bilingual education, and because their school is required to support a bilingual class for their grade (this is a function of the number of LEP students in their grade). Thus, students who are enrolled in bilingual classes are different in many ways, observable and unobservable, from those who are not. Bilingual classes contain students whose parents are, on average, more likely to be immigrants, poor, less-educated, less connected to American job networks, and less sophisticated about educational opportunities in America. An LEP student is more likely to be in a bilingual class if he lives in a neighborhood densely populated by non-English speakers. A student whose first language is not English is more likely to end up in a bilingual class if he is perceived by teachers to be a slow learner; in contrast, a very smart student whose first language is not English is more likely to be considered suitable for English immersion. In short, there are many factors correlated with assignment to a bilingual class and these factors might affect achievement on their own. Many of the correlated factors are unobservable to the researcher or at least very poorly measured. Research tends not to be convincing when it simply compares students in and out of bilingual education, controlling for the factors that we can measure.

For identification, we depend on the *sharpness* of California's policy change. Intuitively, the idea is that little else changed dramatically and discretely between the 1997-98 and 1998-99 school years. Also, to the extent that other changes did occur, the size of the change experienced by a school was not a function of the share of its students eligible for bilingual education. Formally, we are using regression discontinuity for identification–a sharp change in the midst of gradual changes, an event where the scale of the change was determined by pre-existing factors. We discuss this identification strategy in more depth later, but it is useful to add that we do *not* rely on the idea that the ban on bilingual education was unmotivated or on the idea that every school reacted similarly to the ban (some schools did not fully implement the ban). Relying on the sharpness of the change frees us from having to assume that the change was exogenous. We instrument for actual changes in bilingual enrollment with the changes that would have resulted if schools mechanically complied with California's policies for LEP students. This frees us from having to assume that all schools reacted similarly.

We recognize that there are several channels by which the ban on bilingual education may have affected students. Thus, after we estimate a simple, overall effect of the policy change, we attempt to disaggregate the change into several channels by which students were affected. In addition, we are careful to control for other factors, such as class size reduction and test habituation, that may have altered measured student achievement in California in recent years. None of these other factors changed greatly, and sharply, between 1997-98 and 1998-99–as a result, we find that controlling for them does not much affect our results.

II. What is Known about the Effects of Bilingual Education

There are literally ten of thousands of articles published by academic journals on the subject of bilingual education. The vast majority are either opinion or advice. The remaining studies, which are still in the thousands, do examine the effects of bilingual education. Almost none of these are authored by economists, and the vast majority are case studies, as opposed to studies that attempt to estimate a treatment effect of bilingual education.[2] What distinguishes the case studies is the not the small number of students being analyzed. Many of the "treatment" studies examine small numbers of students, too. What distinguishes the case studies is the intimacy and involvement of the researcher with the students being studied. As a result, researchers tend not to encourage treatments that are consistent and stable; instead, they have a fluid sense of what is best for each child and want to see interventions altered as seems best. Perhaps because malleable treatments do not lend themselves to quantitative estimates, case studies rarely produce estimates that are derived in sufficiently uniform way to be interpreted as treatment effects. Moreover, none of the case studies we have read attempts to control for selection into bilingual education treatment; indeed, the researchers are typically interested in observing how treatments are assigned, evolve, and reassigned. They have no interest in identifying similar students who are arbitrarily assigned to different treatments. It is impossible to pick out a representative sample of case studies from the many available, but recent studies include Carlisle and Beeman (2000), Calderon (1999), Johnson (2000), Alanis (2000), Riojas-Cortez (2001), and Kucer (1999).

Many of the studies that do attempt to estimate a treatment effect of bilingual education also show no or little awareness of the problems caused by selection into bilingual education. It is still very common for researchers to compare students assigned to different treatments for limited English proficiency without attempting to control for

[2] The existing research by economists appears to be limited to four doctoral dissertations that I discuss later: Federman (2000), Cheng (1996), Lopez (1996), and Mora (1996).

family background or initial differences in achievement. This may be because it is nearly impossible to assess initial achievement on an English-language test (which will be the later measure of achievement) among kindergarteners who know very little English. Also, within a school, there may be few if any children with very similar family background (for instance, two immigrant parents) who are *not* assigned to the same treatment. Treatment and background are often indistinguishable, therefore.

Among the studies that do show some awareness of the selection problem, two approaches are employed: longitudinal analysis and controls for selection. Two of the most influential longitudinal studies were commissioned by the United States Department of Education: Burkheimer *et al.* (1989) and Ramirez *et al.* (1991). These two are among the many studies ably reviewed by Meyer and Fienberg (1992) in their comprehensive assessment of evaluation studies of bilingual education. As a rule, longitudinal studies follow children for four years or longer and examine students' outcomes as they move from full bilingual education, through an "exit" program, and into regular education. Although longitudinal studies examine *changes* in a student's achievement, thereby controlling for aptitude differences that affect every year similarly, they have no way to combat ongoing selection problems. Within a school, longitudinal studies often compare students who leave bilingual education earlier and later, but students' exit times are determined by teachers' assessments of the student's ability to succeed in an English language classroom. Thus, early and late exiters from bilingual education are very unlikely to have similar unobservable determinants of achievement. For this reason, within-school longitudinal studies often find negative effects of bilingual education. *Between schools*, differences in treatment are often generated by differences in the school's resources or attitude toward bilingual education. Thus, students tend to be in more intensive bilingual treatments and stay in them longer when they attend a school that has a generous budget for bilingual education and has hired specialist teachers who are enthusiastic about the program. For this reason, between-school longitudinal studies often find positive effects of bilingual education.

Meyer and Fienberg (1992) also review numerous studies that are essentially cross-sectional and that attempt to control for selection into bilingual education by introducing (a) linear control variables, (b) a selection correction that is identified by functional form alone, or (c) a selection correction that is identified by implausible exclusion restrictions–such as the idea that the share of students assigned to bilingual education in a school is a valid instrument for the student's own assignment to bilingual education. The vast majority of the cross-sectional studies they survey use controls, not an explicit selection correction. None of these methods is acceptable, especially since highly desirable control variables (such as parents' immigrant status and education) are often excluded because they are unavailable or highly collinear with assignment to bilingual education.

Of the four recent doctoral dissertations on bilingual education by economists, three use a combination of longitudinal analysis and the selection corrections already discussed: Cheng (1996), Lopez (1996), and Mora (1996). Federman (2000), using the National Education Longitudinal Study, employs by far the most credible method to remedy selection. She exploits differences in state laws that affect the funding that schools receive for bilingual education and that affect the threshold at which schools must provide bilingual education. For instance, some states require schools to provide bilingual education when a language minority has at least 20 students in a grade in a school; other states look for 20 students in a grade in a district; and so on. The resulting instruments are unfortunately rather weak, and it is of course impossible to control for state fixed effects in order to remove unobserved differences between states that affect achievement. Nevertheless, Federman finds that bilingual education and language assistance programs (such as English as a Second Language (ESL)) improve students' later mathematics performance. She finds statistically insignificant or very weakly significant effects for reading, history, science, drop-out behavior, and grade retention.

Although economists have devoted little attention to bilingual education *per se*, the struggle between bilingual education and English immersion is closely related to a large economics literature on whether immigrants benefit or lose when they live in ethnic enclaves. In fact, we think of bilingual education as a version of the enclave problem that contains the same trade-offs as the larger problem, but is fortunately more tidy, well-defined, well-documented, and amenable to evidence gathering. Much has been written on immigrant enclaves, but we note two recent contributions by Lazear (1999) and Borjas (2000), both of whom model the trade-off between the insider's edge that an immigrant has in his ethnic enclave and the loss of economic opportunity facing an immigrant who restricts himself to an enclave. Borjas models the reduced incentives for immigrants to acquire the dominant language and culture when ethnic enclaves are available. Lazear shows that policies that promote acquisition of English language skills are welfare enhancing if the resulting access to a broader labor market outweighs the cost of acquiring the skills (including the opportunity cost generated by not working in the ethnic enclave while learning English). Making causal inferences about the impact of enclaves is extremely difficult, for the same reasons that it is difficult to gather evidence on the causal effect of bilingual education. If immigrants with lower unobserved human capital (and corresponding higher costs of learning English) are more likely to concentrate in ethnic enclaves, then the observed relationships between ethnic concentration, English language proficiency, and labor market outcomes could be entirely due to selection. Clearly, if a state were somehow to ban ethnic enclaves, we would have good opportunities for learning about their effects. Edin *et al.* (2002) exploit a Swedish policy that provides geographic placements for new immigrants, reducing the ability of immigrants to sort themselves into enclaves. They find that less skilled immigrants derive labor market benefits from enclaves, and that high-income ethnic enclaves are more helpful than low-income enclaves. The ban on bilingual education may provide us with some of the best evidence, in this spirit, on what happens to non-English speakers when they do not have opportunities for self-segregation.

III. Background on Bilingual Education and Proposition 227

In the United States, 9.3 percent of elementary school students are classified as LEP, and both the absolute number and share of LEP students are growing. LEP students are more likely to be poor than their native English-speaking peers and to perform below their peers on achievement tests.[3] LEP students are concentrated by geography and by ethnicity, with 25 percent of all public school students in California classified as LEP in the 1997-98 school year and about one-third of all Hispanic children nationwide speaking English with difficulty in 1995.[4]

While improving outcomes for LEP students is an important concern for policy makers, the federal government and most states do not legislate particular curricular methods for teaching LEP students. Federal law mandates that some effort be made to address the specific needs of LEP students, but it allows state and local governments to exercise discretion in determining the form of that effort. Prior to 1998, eleven states including California required schools to offer LEP students a bilingual program, which they could accept or not.[5] Three states, Arkansas, Delaware, and Nebraska, specifically forbade bilingual education, and the remaining states were relatively neutral on the subject of bilingual education. Beginning in the fall of 1998, California banned bilingual education in favor of structured English immersion. Arizona voted to do the same in November of 2000, and opponents of

---

[3] Standardized tests will not fully capture proficiency in spoken English, another important determinant of later outcomes.

[4] This comes from a survey question and is distinct from the school-based LEP classification.

[5] These states were: Alaska, California, Connecticut, Illinois, Indiana, Massachusetts, New Jersey, New York, Texas, Washington, and Wisconsin.

bilingual education are actively campaigning for similar changes in Colorado and Massachusetts. The campaigns typically claim that bilingual education fails to teach students English. Unfortunately, there is no research-based consensus as to whether bilingual education or English immersion generates higher achievement among LEP students: there exists a study that supports almost any claim an advocate of a particular program may make.

It is relatively easy to hold a referendum on a policy issue in California because the ballot will include any proposition whose sponsors gather a modest number of signatures. Proposition 227, which was on the June 1998 ballot, was passed and declares that, "all children in California public schools shall be taught English by being taught in English. In particular, this shall require that all children be placed in English language classrooms. Children who are English learners shall be educated through sheltered English immersion during a temporary transition period not normally intended to exceed one year." Proposition 227 went into effect in the fall of 1998. Table 1 shows that 33 percent of LEP students were enrolled in bilingual education in the 1997-98 school year; this share dropped sharply to 11 percent in the fall of the 1998-99 school year. It has since held steady at about 11 percent.

Table 1 also shows that the ratio of LEP students to all students remained constant throughout the implementation period. This fact is *very* important because it indicates that the classification of students into the LEP and non-LEP categories did not change because of the ban on bilingual education. It would cause us considerable difficulty in measuring the effect of bilingual education on LEP (or non-LEP) students if selection into the two categories changed as a result of the ban. There appears to have been no change in selection, however. The trends in bilingual and LEP enrollments presented in Table 1are not weighted; when weighted by student enrollment, the levels and trends are extremely similar (the weighted and unweighted shares LEP and bilingual are within three percentage points of each other for all years).

In this time period in California, a student was classified as LEP if he had *both* a primary language other than English and was judged to have insufficient English language skills for the school's regular instructional program, based on state-approved oral language and (for grades 3-12) literacy assessments.[6]

[6] In 2001 (not affecting our sample) California introduced the California English Language Development Test (CELDT), a statewide standardized test for English learner (the term switched from LEP to English learner) classification. Even under the new CELDT regime, the school will have some discretion in classifications: students with particularly high or low scores will be classified automatically, but if a student's overall score is in the "upper end" of the intermediate category, the student fall into the "possible fluent English proficent" category and "other test scores, report card grades, (and) input from parents/teachers are taken into consideration."

California's bilingual education rolls did not fall to zero because of a provision of the law allowing parents to petition schools for waivers to keep their LEP children in bilingual programs. Schools were given discretion to grant or deny these waiver requests. In some cases, parents, teachers, or administrators organized school-based waiver campaigns, resulting in some schools having very high levels of waivers requested and granted. In the majority of cases, however, all or most LEP students at a given school did transition from bilingual education to structured English immersion following the implementation of Proposition 227. Table 2 shows that the majority of schools that had less than half of their LEP students in bilingual education before Proposition 227 had no students in bilingual education by 2000-01. Schools that had very high shares (80 to 100 percent) of their LEP students in bilingual education before Proposition 227 were more likely to apply for waivers and keep large shares of their LEP students in bilingual education.

We ensure that our empirical strategy takes account of two concurrent changes in California schools. California implemented a class size reduction in the 1996-97 school year, providing incentives for schools to limit kindergarten through third grade classes to twenty students. We control for class size reductions, but find little effect, partly because the policy change preceded the ban on bilingual education and partly because class size did not change in most of the grades we examine. In addition, California began a new program of statewide testing in 1997-98, and it is normal for test scores to rise across the board during the first few years of a test, purely because of habituation. We are careful to account for this as well in our empirical strategy.

IV. The Costs and Benefits of Eliminating Bilingual Education

In order to make our empirical predictions precise, we propose a simple model of the trade-offs associated with eliminating bilingual education in favor of immersion. It is important to recognize that the use of bilingual education for LEP students is likely to affect not only their achievement, but also the achievement of non-LEP students. Therefore, we propose a model of achievement that works for both types of students, and we demonstrate how the same ban on bilingual education might affect each group differently.

We particularly want to model four channels by which bilingual education might affect students. The first is, of course, acquisition of the English language. LEP students who are enrolled in bilingual education may learn English more slowly than those in immersion programs because they are less exposed to English use and have weaker incentives to learn it in order to conduct themselves, communicate with their teacher, and grasp the material presented to them. The opposite may also be true: if the teaching of English is well structured in bilingual education classrooms and excessively casual in immersion classrooms, LEP students might learn English more quickly when they are enrolled in bilingual education.

Second, both LEP and non-LEP students may learn faster when they are in classrooms where teaching is oriented toward their needs. However, a classroom that is oriented toward LEP students' needs could be inappropriate for non-LEP students' needs, and *vice versa*. Therefore, when LEP students are enrolled in English language classrooms, it is likely that some group or both groups will find the teaching environment to be less congenial, although it is hard to know which group (LEP or non-LEP) will be most affected. Non-LEP students may be strongly affected if their school adjusts its English language classrooms greatly to the needs of LEP students (which can be very pressing). On the other hand, LEP students may be strongly affected by the loss of materials and teaching oriented towards their home language and culture.

The third channel by which bilingual education could affect students has nothing to do with English language acquisition *per se*. Bilingual education can effectively segregate students into classrooms of children whose incoming achievement was poor and classrooms of children whose incoming achievement was good. Low incoming achievement may be highly correlated with deficiencies in English, even if English language proficiency has no effect on achievement. For instance, students may have lower achievement if they come from families with low income or low parental education, and such traits may be highly correlated with LEP status. Having peers who are generally low achieving can affect a student's own achievement, regardless of the peers' knowledge of the English language. Although we have little evidence on the forms that peer effects take, it is normal to hypothesize that having high achieving peers is probably good or only slightly bad. The spillovers from high achieving students are generally thought to be good, but the existence of high achieving students may cause teachers to use pedagogical methods that are inappropriate for low achieving students. Thus, the net effect of high achieving peers is theoretically ambiguous but is unlikely to be very negative. In addition, it is normal to hypothesize that having low achieving peers is probably bad for a student's achievement. Not only is the student exposed to peers whose level of knowledge is low, but also the student may be taught using methods geared too low.

The fourth channel by which bilingual education could affect a student's achievement is also indirect: the existence of a bilingual education program can change a school's level of resources or a school's allocation of the resources it has. A school may get extra revenues from the state or federal government if it runs a bilingual education program, as opposed to an immersion program. The effect of bilingual education on a school's total revenues is nearly always non-negative. Running a bilingual education program, however, involves considerable administrative effort. The program may be such a distraction for schools that the extra revenues they obtain are largely or even more than consumed by extra paperwork and organization. Because a school's efficiency may fall when it operates bilingual education, bilingual education has an ambiguous effect on the resources that students *experience*, even though its effect on revenues is definitely non-negative. Moreover, the effect of bilingual education on the resources experienced by a *particular* student is highly ambiguous. On the one hand, LEP students may experience greater resources under bilingual education than under immersion if bilingual classrooms get extra materials, contain fewer students, or contain better teachers. (It is quite possible that teachers with bilingual certification are better. Among the teachers surveyed in the United States Department of Education's Schools and Staffing Survey, bilingual certified teachers were 8 percent more likely to have their baccalaureate degree from a competitive college, paid up to 10 percent more for their certification, and were 5 years younger than the average non-bilingual-certified teacher.) On the other hand, if non-LEP parents are better at inducing schools to direct resources towards their children than LEP parents are, then the segregation of LEP students under bilingual education may allow non-LEP parents to concentrate resources on their children.

A. A Simple Model of English Language Acquisition

Now let us make each of these four channels precise. Consider English language acquisition first, and let E_{ijkt} index the English language ability of student *i* in classroom *j* in school *k* in period *t*. We construct the index E_{ijkt} so that it is equal to one for native English speakers, zero for students who know no English, and between zero and one for students who have some but not complete deficiency in English. Let $\bar{E}_{jkt}$ be the average English language ability of students in classroom *jkt*; and let $I_{jkt}^{English}$ be an indicator variable equal to one if classroom *jkt* is conducted in the English language and zero otherwise. (That is, $I_{jkt}^{English}=0$ for bilingual education classrooms.) Then, we can embody language acquisition with the following equation:

(1) $$E_{ijk,t+1} = \max[E_{ijkt}, \kappa(\bar{E}_{jkt}-E_{ijkt})\cdot(1-I_{jkt}^{English})\cdot E_{ijkt}, \lambda(\bar{E}_{jkt}-E_{ijkt})\cdot I_{jkt}^{English}\cdot E_{ijkt}],$$

where κ and λ are functions such that $\kappa' > 0$ and $\lambda' > 0$.

Equation (1) specifies that an LEP student learns more English if his classmates are more proficient in English, regardless of whether he is in a bilingual education or immersion classroom. If $k(\bar{E}_{jkt} - E_{ijkt}) < \lambda(\bar{E}_{jkt} - E_{ijkt})$ for all $(\bar{E}_{jkt} - E_{ijkt})$, then LEP students always learn more English in immersion than in bilingual education classrooms, and *vice versa*. It could be, however, that the English language advantages of an immersion classroom depend on how well the student's peers know English. That is, it could be that $k(\bar{E}_{jkt} - E_{ijkt}) < \lambda(\bar{E}_{jkt} - E_{ijkt})$ for some levels of $(\bar{E}_{jkt} - E_{ijkt})$ and that $k(\bar{E}_{jkt} - E_{ijkt}) > \lambda(\bar{E}_{jkt} - E_{ijkt})$ for other levels of $(\bar{E}_{jkt} - E_{ijkt})$.

Equation (1) specifies that English language proficiency does not deteriorate once acquired. For instance, natives never lose their proficiency. In order to allow the speed of language acquisition to be determined by data, we purposely have not specified the length of the periods in equation (1). Language acquisition is faster if each period is shorter and there are consequently more periods in a school year. Language acquisition is slower if there are fewer periods in a school year.

B. A Simple Model of Achievement

Now consider achievement, by which we mean students' skills in mathematics, reading, writing, and so on–indeed, all skills that are not English proficiency *per se*. Let us abstract from a particular skill and let μ_{ijkt} be the achievement of student *i* in classroom *j* in school *k* at the beginning of period *t*. We are interested in the evolution of achievement from μ_{ijkt} to $\mu_{ijk,t+1}$ over period t and how it is affected by on learning. Thus, we now add components to a basic model $\mu_{ijk,t+1} = \mu_{ijkt} + \ldots + \epsilon_{ijkt}$, in which ϵ_{ijkt} represents random factors that affect achievement and the ellipses will be filled in terms that represent the channels by which bilingual education potentially affects achievement.

B.1 A Student's Achievement is Higher When Instruction is Oriented to His Language Proficiency

When a student's understanding of a language is very different from that of his classmates, the instruction he receives is unlikely to be appropriate. This principle applies to both LEP and non-LEP students in a straightforward way. An LEP student is unlikely to find instruction geared toward him in a class where most children are more much proficient in English than he is. Equally, a native English speaker is likely to find that instruction inevitably becomes oriented toward English learners if he is immersed in a class of LEP students.[7] We can embody these effects by adding the following term to our basic achievement equation:

$$\mu_{ijk,t+1} = \mu_{ijkt} + \ldots + \beta_6(|E_{ijkt} - \bar{E}_{jkt}| \cdot I_{jkt}^{English}) + \ldots + \epsilon_{ijkt} \tag{2}$$

In equation (2), β_6 is a function. We expect β_6' to be negative if a student is worse off, all else equal, when his English language knowledge is very different from the mean level of English language knowledge in the class. This is true both for a student who knows English well and a student who know English poorly–thus the absolute value bars around the difference between the student's English knowledge and the class's mean English knowledge. For simplicity, the β_6 function imposes the restriction that the student's deviation from his class's average English knowledge does not affect his achievement unless he is in a classroom conducted in English.

<u>B.2. A Student's Achievement May Be Affected by the Achievement of His Peers</u>

Above, we noted that one effect of removing bilingual education is that LEP students are generally immersed in classrooms with higher achieving students (regardless of language). A student who is in a class with higher achieving peers is likely to get some positive knowledge spillovers but may suffer because his teacher gears her pedagogy towards higher achievers. The net effect is ambiguous. A student who is in a class with lower achievers may get few positive or even negative knowledge spillovers and may also suffer because his teacher's methods are geared towards lower achievers. The net effect is probably but not necessarily negative.[8] In any case, the effect of having lower achieving peers is likely to be different than the effect of having higher achieving peers.

We embody peer effects by adding two terms to the achievement equation, so that it becomes:

$$\begin{aligned}\mu_{ijk,t+1} = \mu_{ijkt} + \ldots + \beta_4((\bar{\mu}_{jkt} - \mu_{ijkt}) \cdot I^{\mu_{ijkt} < \bar{\mu}_{jkt}}) + \beta_5((\mu_{ijkt} - \bar{\mu}_{jkt}) \cdot I^{\mu_{ijkt} > \bar{\mu}_{jkt}}) + \\ + \beta_6(|E_{ijkt} - \bar{E}_{jkt}| \cdot I_{jkt}^{English}) + \ldots + \epsilon_{ijkt}\end{aligned} \tag{3}$$

β_4 is a function that embodies the effect of being in a class with higher achieving peers, and we have assumed the effect is a function of the difference between peers' mean achievement and the student's own achievement.

[7]Note that these effects are due entirely to the composition of the class and not the method of instruction.

[8] Some education experts claim that knowledge spillovers from lower achieving students are positive because the higher achievers are forced to *teach* lower achievers, thus honing their own understanding.

Similarly, β_5 is a function that embodies the effect of being in a class with lower achieving peers. Observe that, if the effect of being with higher and lower achieving peers were symmetric, we would find that $\beta_4' = -\beta_5'$.

<u>B.3. Achievement is Affected by the Resources a Student Experiences and the Efficiency with which the Resources are Used</u>

We propose that a student's achievement may be raised by experiencing more resources or by experiencing resources that are more efficiently used. Bilingual education can affect a school's overall level of resources, the distribution of resources among LEP and non-LEP students, and the efficiency with which a school's resources are used overall. We add three terms to the achievement equation in order to embody these possibilities:

$$\begin{aligned} &\mu_{ijk,t+1} = \mu_{ijkt} + \beta_1(X_{kt}) + \beta_2(X_{kt} \cdot (\bar{I}_{kt}^{English} - 1)) + \beta_3(X_{kt} \cdot (I_{jkt}^{English} - \bar{I}_{kt}^{English})) \\ &\beta_4((\bar{\mu}_{jkt} - \mu_{ijkt})\, I^{\mu_{ijkt} < \bar{\mu}_{jkt}}) + \beta_5((\mu_{ijkt} - \bar{\mu}_{jkt})\, I^{\mu_{ijkt} > \bar{\mu}_{jkt}}) + \beta_6(|E_{ijkt} - \bar{E}_{jkt}| \cdot I_{jkt}^{English}) + \ldots + \epsilon_{ijkt} \end{aligned} \tag{4}$$

β_1 is a function that embodies the effect of school *k*'s resources in period *t* (X_{kt}). β_2 is a function such that $\beta_2(0) = 0$; it embodies the effect that bilingual education classes have on the efficiency with which a school uses its resources. Notice that $\bar{I}_{kt}^{English}$ is school *k*'s mean of the indicator variable for English language classrooms in period *t*. Thus, $\bar{I}_{kt}^{English}$ is equal to one if all classes are conducted in English and is less than one otherwise. If all classes are in English, then the β_2 term is equal to zero and the resource effect is given by β_1. Finally, β_3 is a function such that $\beta_3(0) = 0$; it allows the effect of a school's resources to depend on whether a student is in a classroom that conducted in English. (To differentiate this effect from the efficiency effect, notice that β_3 function depends on the *difference* between a classroom's own English language designation ($I_{jkt}^{English}$) and the school's mean classroom designation ($\bar{I}_{kt}^{English}$). Thus, our construction of the β_3 function is such that the average classroom in a school *cannot* experience resource discrimination ($I_{jkt}^{English} = \bar{I}_{kt}^{English}$ so $\beta_3(0) = 0$). Also observe that the β_3 function allows resource discrimination to be in favor of *either* English language or bilingual education classrooms. There is no a priori reason to believe that discrimination works favor of one or the other type of classroom. On the one hand, schools are typically given extra resources for LEP students and are told to use them in favor of LEP students; on the other hand, non-LEP parents may be good at getting resources directed away from bilingual education classrooms.

<u>B.4. Measured Achievement is Higher when Students are Habituated to a Test</u>

We will be using standardized tests to measure achievement, and we need to add one other effect to our equation to account for a circumstance peculiar to California's ban on bilingual education. Coincidentally, it happens that California began using its new STAR testing program, which is based on the Stanford 9 exam, just one year before it banned bilingual education. It is normal for students to do slightly better on an exam for each of the first few years it is given, simply because they become habituated to its idiosyncracies. We wish to allow such habituation to affect measured achievement, so we add two additional terms and obtain our complete model of achievement:

(5)
$$\mu_{ijk,t+1} = \mu_{ijkt} + \beta_1(X_{kt}) + \beta_2(X_{kt} \cdot (\bar{I}_{kt}^{English} - 1)) + \beta_3(X_{kt} \cdot (I_{jkt}^{English} - \bar{I}_{kt}^{English}))$$
$$\beta_4((\bar{\mu}_{jkt} - \mu_{ijkt})\, I^{\mu_{ijkt} < \bar{\mu}_{jkt}}) + \beta_5((\mu_{ijkt} - \bar{\mu}_{jkt})\, I^{\mu_{ijkt} > \bar{\mu}_{jkt}}) + \beta_6(\,|E_{ijkt} - \bar{E}_{jkt}| \cdot I_{jkt}^{English}) +$$
$$\beta_7(t - t_0) + \beta_8((t - t_0) \cdot (1 - E_{ijkt})) + \epsilon_{ijkt}\ ,$$

where t_0 is the period in which the exam is initially administered. The β_7 function embodies the habituation effect for a native English speaking student. The β_8 function is included so that habituation effect can be different for LEP students. For instance, if β_8' is positive, then the habituation effect is larger for LEP students than non-LEP students; and *vice versa*. The β_7 and β_8 functions are in terms of $(t - t_0)$ for simplicity. This would be a somewhat restrictive functional form if we observed numerous years of test-taking, but we actually observe such a small number of years after the ban on bilingual education that we are fully flexible about the form of the habituation effect in our empirical work.

B.5. Summing Up: The Effect of Bilingual Education on Achievement

Between them, equations (5) and (1) contain all of the channels by which bilingual education may affect achievement. Equation (5) shows achievement *within* a period *t* so that students' level of English proficiency is held constant. Bilingual education may change a student's exposure to English proficient peers (that is, may affect the β_6 term). Bilingual education may change a student's exposure to higher or lower achieving peers (that is, may affect the β_4 and β_5 terms). Bilingual education may affect a school's resources (affect the β_1 term), the efficiency with which resources are used (affect the β_2 term), or the distribution of resources across LEP and non-LEP students (affect the β_3 term). *Moreover*, through equation (1), bilingual education may affect the rate at which LEP students learn English, thereby affecting their ability to derive knowledge from English language instruction and changing them as peers. If LEP students learn English faster in period *t*, then all students' achievement is affected by their new proficiency through the β_6 term in period *t+1*.

V. An Empirical Strategy for Estimating the Effects of Bilingual Education

We set up the model of the last section so that it would be clear that bilingual education can affect achievement through several channels. Some of the key lessons are that (1) the ban on bilingual education may affect non-LEP students as well as LEP students, (2) the *magnitude* of the effect of the ban will depend on the share of students who were initially enrolled in bilingual education, (3) the magnitude of the effect of the ban may also depend on the English proficiency of the student body overall (for any given share enrolled in bilingual education), (4) the ban may have indirect effects through the allocation of resources such as teachers, (5) the ban may differentially affect different grades if younger children learn English faster than older children, and (6) the ban may differentially affect measured achievement in subjects depending on whether English language knowledge is crucial for doing well on the tests (reading) or merely helpful (mathematics). Thus, even though we analyze a policy change in California that is simple, we need an empirical strategy that does not force the policy effect to be simple. In particular, it is *not* appropriate to use a simple differences-in-differences approach in which non-LEP students are *assumed* to have been unaffected by ban on bilingual education.

In this section, we attempt to do two things. First, we explain how we remedy a few identification problems that arise in our study of California's ban on bilingual education. Second, we explain how we estimate the simple effect of the schools' reducing enrollment in bilingual education and then how we break down the effect into the channels suggested by our model.

A. Endogenous Changes in Bilingual Education Enrollment

Our identification of the effects of bilingual education relies mainly on the *discreteness* of the change in California's policy on bilingual education. That is, we are not claiming that the existence of the policy change is really exogenous or even that the timing of the policy change is arbitrary. Also, we are not claiming that the schools greatly affected by the ban are similar to the schools hardly affected by the ban. What we rely upon is that the ban was *abrupt*: it encouraged schools to make large, discrete changes in the share of their students enrolled in bilingual education at a time when the same schools were not experiencing other large, *discrete* changes. Put another way, we use regression discontinuity for identification, relying on the fact that other determinants of student achievement did not change *dramatically* and *abruptly* between the 1997-98 and 1998-99 school years.

For this identification strategy to be convincing, we need to confront the problem that schools' actual changes in bilingual enrollment were potentially correlated with their attitudes toward bilingual education. This is for two reasons. First, although a student is determined to be *eligible* for bilingual education based on well-defined rules, a school could influence what share of eligible children actually enrolled in bilingual education before the ban.[9] A school that was enthusiastic about bilingual education might enroll a larger share of its eligible students in bilingual classes. Second, although schools that mechanically implemented the ban eliminated bilingual education altogether, schools were permitted to try to obtain and grant requests for waivers from parents of eligible students in order to keep children in bilingual education. A school's effort in seeking and granting waiver requests and parents' willingness to sign these requests are likely correlated with local attitudes toward bilingual education. We do not want local attitudes toward bilingual education to alter our measure of the "treatment" that the school receives: we want to compute the effect of bilingual education, not the effect of *attitudes toward* bilingual education.

An example may clarify our logic. Suppose that schools A and B have identical student populations so if the two schools act mechanically, they experience exactly the same change in bilingual education as a result of the ban. Suppose, however, school A is enthusiastic about bilingual education and school B is not. Prior to the ban, school A enrolls every student who is eligible for bilingual education in a bilingual class. After the ban, school A grants as many waivers as possible, has a decrease in the share enrolled in bilingual education (despite keeping "waived" students in bilingual education), and implements English immersion reluctantly and poorly. Prior to the ban, school B enrolls only some of its bilingual eligible students enrolled in bilingual classes; and, after the ban, it decreases the share enrolled to zero and implements English immersion enthusiastically and well. Simple least

[9] A student had to be classified as LEP and have at least nine other LEP students of the same native language in his grade at his school to be eligible for bilingual education under the Chacon-Moscone Bilingual-Bicultural Education Act of 1976. In 1987, Governor Deukmejian did not reauthorize the act, but its provisions were still technically binding. See Baker and Hakuta (1997) for further discussion.

squares estimates would suffer from bias of unknown sign. Because school B implements English immersion better than school A, the effects of its English immersion will be better after the ban. However, enthusiasm for English immersion may be either positively or negatively correlated with a school's change in bilingual enrollment. On the one hand, school A may experience a larger decrease in bilingual enrollment because it initially enrolled a larger share of eligible students in bilingual classes. On the other hand, school A may experience a smaller decrease in bilingual enrollment because it sought waivers.

We want to compute an unbiased estimate of the effect of bilingual education. Thus, although we start by showing simple least squares results in which the explanatory variable is a school's *actual* change in its share of students in bilingual education, we move immediately to instrumental variables results. The instrument is the decrease in a school's share of students enrolled in bilingual education that would occur if the school followed state guidelines in a mechanical way. This instrument is formed by computing the school's share of students who were *eligible* for bilingual education in 1997-98 and subtracting that number from zero. (Zero is the share of each school's students who would have been enrolled in bilingual education in 1998-99 if each school had mechanically implemented the ban.) Note that we use the share of students *eligible* for bilingual education *prior* to the ban. We do this so that the effect of the ban varies only with the school's population of students, not with the school's attitude toward bilingual education or the school's beliefs about how the ban should be implemented. A school's population of students is not under its control and is likely to be very similar before and after the ban. By instrumenting with zero minus the share of students who were *eligible* for bilingual education *before* the ban, we use only the change in bilingual education that is exogenous. Indeed, the two schools in our example would have exactly the same value for the instrumental variable, so their attitudes could not be correlated with their instrumented treatment, by definition.

B. California's Class Size Reduction

Our identification strategy depends on the abrupt, large change in California's policies toward bilingual education. If other policies also changed abruptly and significantly at exactly the same time, they would be a problem for us. Of course, the vast majority of school-related policies were constant or changed only slightly from 1997-98 to 1998-99. The one policy that gives us slight pause is California's class size reduction, which was implemented between the 1995-96 and 1996-97 school years. We are not overly concerned because most districts that participated in class size reduction were already doing so before the ban on bilingual education: 98 percent of districts participated in class size reduction in 1997-98 and 99 percent participated in 1998-99. Thus, class size reduction does not qualify as an abrupt change that is simultaneous with the ban on bilingual education. Also, class size reduction only affected grades of kindergarten through three, and we will be examining the effects of the bilingual ban on other grades. Nevertheless, we wish to take every precaution against our estimated bilingual effects becoming contaminated by the effects of class size reduction. Therefore, after presenting our simple estimates, we focus on estimates that control for actual changes in schools' class size. For reference, Table 3 shows how class size changed in California from 1997-98 to 2000-01.[10] Only grades up through sixth grade are shown because class size is not available (and, indeed, is not well defined) for higher grades in which the single classroom model is not followed.

C. Habituation to the Test

California began its current system of testing, known as the STAR system, in the 1997-98 school year. The core of the STAR system is the Stanford 9 Achievement Test, a well-known standardized test that has been administered to grades 2 through 11 from the 1997-98 school year onwards. We use mean percentile rank scores from the Stanford 9 as our measure of achievement. We should emphasize that the percentile rank scores are based on the performance of a very large, national "norm" group of test-takers whose test-taking precedes California's administration. Thus, the percentage rank scores are absolute scores: by performing well, California students cannot raise the bar for themselves.

It is well known that students score better on tests that are familiar to them in terms of test set-up, answer sheet set-up, and so on. For this reason, students' scores generally rise slightly on a test during the first few years it is administered, even if their knowledge is exactly the same. After the first few years, the benefits of becoming habituated to the test are exhausted, and scores rise only with real improvements in knowledge.

Because California began administering the Stanford 9 in 1997-98 (the year before the bilingual ban took effect), we would expect to see habituation-driven score increases in the 1998-99 through 2000-01 school years. We do not want to attribute these habituation-driven improvements to the ban on bilingual education. Therefore, we need to calculate the counterfactual increase in California scores *purely from habituation* over the 3 years in

[10]Class size changed for third grade between 1997-98 and 1998-99, despite the policy having been implemented in the previous year. The policy affected first and second grades immediately with a lagged effect on third grade and kindergarten.

question.

The introduction of the Stanford 9 test and the ban on bilingual education were both state-wide events, so we must look outside California to get data for our counterfactual calculation. Fortunately, Arizona began administering the Stanford 9 test to its students in 1996-97 (one year before California), and Arizona did not have a ban on bilingual education during the first three years of its test administration. We therefore compute an habituation effect for each California school by running a regression of Arizona schools' scores on their demographics; indicator variables for the second, third, and fourth year of test administration; and interactions between each demographic variable and each year indicator variable:

(9)
$$\begin{aligned}\mu_{subject,gkt} &= \delta_0^{subject,g} + \boldsymbol{X}_{kt}\boldsymbol{\delta}_1^{subject,g} \\ &+ \delta_2^{subject,g} I_t^{Test\ Year\ 2} + \delta_3^{subject,g} I_t^{Test\ Year\ 3} + \delta_4^{subject,g} I_t^{Test\ Year_4} + \\ & I_t^{Test\ Year\ 2}\boldsymbol{X}_{kt}\delta_5^{subject,g} + I_t^{Test\ Year\ 3}\boldsymbol{X}_{kt}\delta_6^{subject,g} + I_t^{Test\ Year\ 4}\boldsymbol{X}_{kt}\delta_7^{subject,g} + \omega_{subject,gkt}\end{aligned}$$

where $\boldsymbol{X}_{kt}$ is a vector of demographic variables, like percent black and percent Hispanic, and $I_t^{Test\ Year\ n}$ is an indicator variable for the n^{th} year of test administration. We apply the estimated coefficients on the year indicator variables and interaction variables to California data, and compute a predicted effect of habituation for each California school. Observe that if white students habituate faster, say, this difference in habituation will be accounted for in our estimates. All systematic differences in habituation that are related to demographics will be embodied in our estimates. For instance, the habituation effect for the third year of test administration (1999-00) in reading in grade four in the California school numbered 100 would be:

$$habit_{reading,grade\ 4,100,1999-00} = \hat{\delta}_3^{reading,grade\ 4} + \boldsymbol{X}_{\mathbf{100,1999-00}} \cdot \hat{\delta}_6^{reading,grade\ 4}.$$

Although we do not yet have Stanford 9 data from Texas, some districts there have also introduced the Stanford 9 in recent years. We plan to use Texas data to get yet more estimates of the habituation effect.

<u>D. From Simple Estimates to Estimates that Distinguish Channels of Effect</u>

We propose to start with the simplest possible method of estimating the effect of the ban on bilingual education. First, we linearize equation (5)–that is, we assume that all of the functions β_1 through β_8 in equation (5) are linear in their arguments. Then we estimate just the effect of the change in a school's percentage of students enrolled in bilingual education. This one effect picks up several of the channels mentioned in the model. After presenting these simple estimates, we add variables to pick up other channels of effect and to distinguish channels of effect. In other words, we start with a very simple estimating equation and gradually flesh it out so that looks more and more like the model.

We start by simply estimating the effect of the change in a school's percentage of students in bilingual education, computing separate estimates for LEP and non-LEP students, for each grade, and for each subject. Thus, our most basic estimating equation is:

(6) $$\Delta\mu_{subject,language,gkt} = \alpha_0^{g,subj,lang} + \alpha_1^{g,subj,lang}\Delta PctBE_{gkt} + \Delta\epsilon_{gkt}$$

where

(7) $$subject \in \{reading, math, language\}\ ,\ language \in \{LEP, non\text{-}LEP\}$$

In equation (6), *g* indexes the grade being tested, Δ indicates that we take the first difference in the variable (after the ban minus before the ban), and *PctBE* is the percentage of students in bilingual education. "Subject" designates the subject being tested, and "language" designates whether the students are non-LEP or LEP. Thus, we estimate equations like:

$$\Delta\mu_{reading,LEP,3,kt} = \alpha_0^{reading,LEP,3} + \alpha_1^{reading,LEP,3}\Delta PctBE_{3,kt} + \Delta\epsilon_{3,kt}$$

$$\Delta\mu_{math,non-LEP,4,kt} = \alpha_0^{math,non-LEP,4} + \alpha_1^{math,non-LEP,4}\Delta PctBE_{4,kt} + \Delta\epsilon_{4,kt}$$

In order to allow bilingual education time to affect achievement, we mainly show three-year changes: from the 1997-98 school year ("before") to the 2000-01 school year ("after"). However, we do show one- and two-year results in appendix tables.

After estimating equation (6) by least squares, we then estimate it by instrumental variables, where the instrument for $\Delta PctBE_{gkt}$ is $(0 - PctEligibleBE_{gk,t-3})$. We then proceed to add variables to the estimating equation, always keeping the implied first stage for the instrumental variables procedure exactly parallel to the implied second stage. Table A-1 in the appendix shows the results of the first stage equation. Not surprisingly, the instrumental variable is always a very powerful predictor for the endogenous variable for which it is instrumenting (t-statistics on the instrumental variable are uniformly above 15). The great power of the instrument is due to the fact that many schools *do* mechanically implement California's bilingual policies. Even the schools that deviate from mechanical implementation do so only partially.

Our first modification to the basic equation is the removal of the habituation effect. We simply subtract the predicted habituation effect from the change in test scores that is our dependent variable:

(8) $$\Delta\mu_{subject,language,gkt} - habit_{subject,gkt} = \alpha_0^{g,subject,language} + \alpha_1^{g,subject,language}\Delta PctBE_{gkt} + \Delta\epsilon_{gkt}.$$

In other words, our dependent variable is now the change in achievement *above and beyond* predictable habituation.

Our next modification to the basic equation is the addition of the actual change in class size as an explanatory variable:

(9) $$\Delta\mu_{subject,language,gkt} - habit_{subject,gkt} = \alpha_0^{g,subject,language} + \alpha_1^{g,subject,language}\Delta PctBE_{gkt} + \alpha_5\Delta ClassSize_{gkt} + \Delta\epsilon_{gkt}\ .$$

Changes in class size are trivial for grades four and up, in any case.

Our model predicts that the ban on bilingual education may affect achievement through the channel of exposure to English proficient students. Thus, we add a measure of the change in the percentage of non-LEP students to whom a student is exposed in class. For exactly the same reasons we need to instrument for the change in a school's share of student enrolled in bilingual classes, we need to instrument for the change in a student's *actual* exposure to non-LEP students with the change that the student would experience if his school were to have mechanically implemented bilingual education and mechanically implemented the ban. That is, for non-LEP students, the instrument is equal to the share of non-LEP students in the school overall (exposure to non-LEP after a mechanical implementation of the ban) minus the non-LEP share that would have existed in English language classrooms before the ban if the school had mechanically enrolled all bilingual eligible students in bilingual classes:

(10) $$\frac{NonLEP_{gk,t-3}}{NonLEP_{gk,t-3} + LEP_{gk,t-3}} - \frac{NonLEP_{gk,t-3}}{NonLEP_{gk,t-3} + (LEP_{gk,t-3} - BEeligible_{gk,t-3})}.$$

Notice that the students who are eligible for bilingual education ($BEeligible_{gk,t-3}$) are a subset of the LEP students. Also observe that the instrument is based entirely on data from before the ban, so that it is the mechanical change in exposure and cannot include any reaction to the ban. For LEP students, there is a parallel instrument that looks more complicated but is actually completely parallel.

$$(11)\quad \frac{LEP_{gk,t-3}-BEeligible_{gk,t-3}}{LEP_{gk,t-3}}\cdot\left[\frac{NonLEP_{gk,t-3}}{NonLEP_{gk,t-3}+LEP_{gk,t-3}}-\frac{NonLEP_{gk,t-3}}{NonLEP_{gk,t-3}+(LEP_{gk,t-3}-BEeligible_{gk,t-3})}\right]+\frac{BEeligible_{gk,t-3}}{LEP_{gk,t-3}}\cdot\left[\frac{NonLEP_{gk,t-3}}{NonLEP_{gk,t-3}+LEP_{gk,t-3}}-0\right]$$

The instrument shown in equation 11 is just the weighted average of the mechanical change in exposure for LEP students who were not eligible for bilingual education and the mechanical change in exposure for LEP students who were eligible for bilingual education. The first-stage equations for the instrumental variables procedure just described are shown in Table A-2. The excluded instrumental variables are uniformly powerful (with t-statistics above 10) because most schools do follow California's bilingual policies closely.

Summing up, we add the change in exposure to non-LEP students to the estimating equation, and we always instrument for this variable with the change in exposure that would have occurred if schools mechanically implemented California's bilingual policy. Our estimating equation thus becomes:

$$(12)\quad \Delta\mu_{subject,language,gkt}-habit_{subject,gkt}=\alpha_0^{g,subject,language}+\alpha_1^{g,subject,language}\Delta PctBE_{gkt}+\alpha_2\Delta PctNonLEPinClass_{language,gkt}+\alpha_5\Delta ClassSize_{gkt}+\Delta\epsilon_{gkt}\ .$$

Finally, we want to add variables that can pick up the resource channel through which the ban on bilingual education might operate. We can add a vector of changes in measured resources for each school, $\Delta \boldsymbol{Z}_{kt}$. The vector includes changes in real per-pupil spending and teacher experience, but we do not expect adding this vector to affect the other coefficients much. This is because there were no major changes in California's resources for schools over the period in question, other than class size. Also, California school finance is highly centralized and equal across schools, so many changes that were experienced by all schools in California were already absorbed by the constant, α_0.

However, there is one peculiar change in resources that we think *may* matter. When bilingual classes were eliminated, their would-be students were merged into other classrooms. Bilingual classes were generally smaller than English language classes and their teachers were paid significantly more than other teachers (because there is a salary step associated with bilingual certification in California teachers' contracts). Thus, class size generally rose and teacher quality probably fell for LEP students; and class size generally fell and teacher quality probably rose for non-LEP students as schools reallocated resources in response to the ban. Notice that the reallocation occurs for a *given* average class size. Thus, our current control for class size does not pick up the reallocation of resources that we want to capture. We do not actually know the changes in class size or teacher quality experienced by each student, so we cannot control for them directly. We do know how much the range of class size shrunk in each grade in each school following the ban in bilingual education, and this is an indicator of reallocation for which we can control. For instance, if a school had two fourth grade classes before the ban, a bilingual class of 15 students and an English language class of 25 students, then its "before" range in fourth grade class size was 10. If, after the ban, the school had two fourth grade classrooms of 20 students each, then its "after" range in fourth grade class size would be zero. The change in a school's range would be good indicator of its reallocation of resources, and we would expect a falling range to hurt LEP students (because they were previously in smaller classes with better-paid teachers) and help non-LEP students (because they were previously in larger classes with worse-paid teachers).

Adding the vector $\Delta \boldsymbol{Z}_{kt}$ and the change in the class size *range*, we obtain our final estimating equation:

(13)
$$\Delta\mu_{subject,language,gkt} - habit_{subject,gkt} = \alpha_0^{g,subject,language} + \alpha_1^{g,subject,language}\,\Delta PctBE_{gkt}$$
$$+\alpha_2\,\Delta PctNonLEPinClass_{language,gkt} + \alpha_3\,\Delta RangeClassSize_{gkt} + \Delta\boldsymbol{Z}_{gkt}\,\boldsymbol{\alpha_4} + \alpha_5\,\Delta ClassSize_{gkt} + \Delta\epsilon_{gkt}\ .$$

Equation (13) comprises all of the channels described in our model, albeit with a simplified, linear functional form. α_1 is the effect of bilingual education that does *not* work through exposure to English-proficient students or resource changes; α_2 is the effect of bilingual education that operates through exposure to English-proficient students; and α_3 is the effect of resource reallocation that is probably a result of bilingual education. The other coefficients in the equation may be interesting in and of themselves, but they are essentially nuisance coefficients so far as the effects of bilingual education are concerned. We have not attempted to parameterize the speed at which students learn English, but we allow the effects of bilingual education to vary by grade and by subject. We can examine the estimated effects to see if there are patterns that suggest how English proficiency itself is changing. For instance, if the effects of English immersion are better for students in younger grades, this may be because younger people learn new languages more readily. Similarly, if the effects of English immersion are worse for mathematics than for reading, this may be because students were getting more out of mathematics material when it was partly taught in their mother tongue but students were not learning how to read well in English when their reading lessons were partly taught in their mother tongue.

V. Data

Achievement, demographic, and resource data at the school level are available annually for the 1997-98 through 2000-2001 school years from the California Department of Education (CDE) and the Common Core of Data Public School Universe. Arizona test score data come from the Arizona Department of Education, and demographic data come from the Common Core of Data Public School Universe.

California mandated statewide use of the Stanford Achievement Test, Ninth Edition, Form T (Stanford 9) beginning in the 1997-98 school year, one year before Proposition 227 took effect. This provides us with one year of pre-reform data. The latest post-reform data available are for three years later, the 2000-01 school year. We use test scores for reading, math, and language, which are reported for each school at the grade level. Micro-level data are not available. We use school-level data on the national percentile rank of the mean student and on the share of students exceeding the 25^{th} and 75^{th} national percentiles for each subject and grade. For confidentiality, scores are reported for LEP students as a separate group only if at least there are more than ten LEP students in that grade at the school, and we limit our sample to schools with sufficient LEP students to report separate data. LEP students who first enrolled in a California public school less than twelve months prior to the testing date were tested beginning in 1998-99 in Spanish, and thus are excluded from the Stanford 9 sample.

Demographic data from the CDE comes from the Language Census, the California Basic Educational Data System (CBEDS), and CalWORKS. The Language Census collects information on the number of limited English proficient students, by grade and primary language, each year. We use these data to determine the number of children eligible for bilingual education in the 1997-98 school year, prior to the implementation of Proposition 227. Schools were required to offer bilingual instruction to LEP students if it was considered reasonable. The criteria for reasonableness were ill-defined in the 1997-98 school year but many schools used a more specific past rule as a focal point: if at least ten LEP students in one grade shared the same primary language, the school had to provide bilingual instruction in that language in that grade. CBEDS includes data on the number of LEP students in different instructional settings each year. The CalWORKS data include information on the number of free and reduced lunch eligible students in each school.

The CBEDS system includes a Professional Assignment Information Form, with data on the number of students and grade/subject matter of each class assigned to each teacher. Teacher identification numbers change each year so it is not possible to track changes in the assignments of individual teachers. We use these data to examine changes in the mean class size within a grade at a school.

We also use Stanford 9 test score data and demographic data from the Public School Universe of the Common Core of Data from Arizona schools, to estimate a test habituation effect. Arizona began using the Stanford 9 in the 1996-97 school year. We use school report card data on mean Stanford 9 scores by grade and subject at the school level from 1996-97 through 1999-2000, [11] and corresponding data from the Common Core on the racial and ethnic composition of schools.[12]

Tables 4 and 5 contain descriptive statistics for our dependent variables and key explanatory variables. We use these statistics when interpreting the results of our regressions.

VI. Results

Our main predictions are that Proposition 227 will change the share of students enrolled in bilingual education and that this may differentially affect students with varying levels of English proficiency. We allow the effects to vary by the student's grade and by the subject in which the student is tested. As described above, we begin with a very simple specification and then make our specification more sophisticated in order to account for possible

[11] Arizona voters passed Proposition 203, similar to California’s Proposition 227, in November of 2000; our use of Arizona data is restricted to school years before the policy change.
[12] Arizona did not report free and reduced lunch eligibility data in the necessary years of the Common Core.

sources of endogeneity, confounding factors, and the multiple channels through which bilingual education could affect achievement.

A. The Effects of Bilingual Education on the Achievement of LEP Students in the Fourth Grade

In Tables 6 and 7, we present results for our full array of specifications, from our most basic estimating equation in the left-hand column to our most sophisticated specification in the right-hand column. We show results for fourth graders only because the table would be too complicated if we showed all grades. We concentrate on the results for LEP students in Table 6 and the results for non-LEP students in Table 7.

The far left-hand column of Table 6 contains least squares results for equation (6), our most basic estimating equation in which changes in achievement are predicted solely by changes in the share of all students in a school who are enrolled in bilingual education. We find that the least squares effect of changes in the share of students in bilingual education is statistically insignificantly different from zero for reading and mathematics. The change in the share bilingual is statistically significantly different from zero for language, and the sign suggests that reductions in percent bilingual *raise* LEP students' achievement on the language test. However, all of the least squares coefficients are small in magnitude. For instance, the mean decline in bilingual enrollment over this period was about 12 percent; so the mean decline in bilingual education would generate a 0.035 point increase in the mean percentile rank score of LEP fourth graders on the language test. This is an increase of only 0.0036 standard deviations on the language test because it has a standard deviation of 9.7 for fourth graders. In column (2), we remove the habituation effect but still use least squares, and we find that reductions in percent bilingual apparently reduce LEP students’ achievement in reading and mathematics but have a statistically insignificant effect in language.

However, we do not wish to give much interpretation to the least squares because we are confident that they are biased, and we do not even know the direction of the bias. On the one hand, schools with a positive attitude toward bilingual education would have had a larger share of their eligible students enrolled in bilingual classes before Proposition 227; on the other hand, schools with a positive attitude toward bilingual education were more likely to be obtain waivers in order to keep bilingual enrollment unchanged. Thus, even if we knew the correlation between a school's unobserved attitude toward bilingual education and its unobserved factors that improve achievement, we would not know the sign of the bias.

In column (3) of Table 6, we eliminate the effects of schools' endogenous responses to Proposition 227 by instrumenting for a school's change in the percent bilingual with the change that would have occurred if the school had mechanically followed California's policies. Thus, column (3) shows the first estimates that we consider to be reliable. Our instrumental variables estimates indicate that reductions in percent bilingual *reduce* LEP fourth graders' achievement by statistically significant amounts in reading and mathematics, but have no statistically significant effects on language. For instance, the mean change in percent bilingual (a reduction of 0.12) would lower mean reading scores by 0.758 percentile rank points (0.097 standard deviations) and lower mean mathematics scores by 1.805 percentile rank points (0.17 standard deviations).

Next, we wish to remove the effect of habituation to the test. Recall that we use data from Arizona to predict the habituation effect, after three years, for fourth graders who are demographically like the fourth graders in each California school. We then subtract the predicted habituation effect from the observed change in achievement for each school to get the change in achievement purged of the effect of habituation. Column (4) of Table 6 presents instrumental variables results for achievement purged of habituation. Observe that using changes in achievement purged of habituation increases the standard errors of the coefficients only slightly. This suggests that the predicted habituation effect is not measured with a great deal of error because, if it were, it would greatly increase noise in the regression. We believe that the habituation effect is accurately predicted because we use an entire state's worth of data.

When we use achievement purged of habituation, we find that reductions in a school's percent bilingual *reduces* LEP fourth graders' reading, mathematics, and language scores by statistically significant amounts. A school that reduced percent bilingual by 0.12 (the average school's reduction) would see its mean LEP student's reading score decline by 1.103 percentile points (0.14 standard deviations), his mathematics score decline by 3.091 percentile points (0.30 standard deviations), and his language score decline by 1.094 percentile points (0.12 standard deviations). The effects of bilingual education on LEP students are the same in sign, whether or not they are purged of habituation effects. However, the effects are larger in absolute value when purged of habituation. Notice that mathematics achievement is most hurt by a reduction in bilingual education–that was to be expected because mathematics is the one subject in which instruction in a mother tongue might reasonably be expected to be helpful on an English language examination. After all, much of what is taught in mathematics is the symbolic language of mathematics, and the symbolic language is universal.

In column (5) of Table 6, we add controls for the change in mean class size for fourth graders at the school and the change in the fraction of students at the school who are eligible for free or reduced-price lunch. Controlling for these changes produces a negligible change on the estimated effect of bilingual education on achievement. Also, class size itself does not have a statistically significant effect on achievement, except for mathematics scores. (The actual mean reduction of 0.27 students per class would raise mathematics scores by only 0.05 percentile rank points, or 0.005 of a standard deviation.) We find that schools with decreasing shares of students eligible for free or reduced price lunch have significantly greater achievement gains.

We would like to know how much of the effect of bilingual education operates through the channel of exposure to non-LEP students. Thus, in column (6) we add the change in percent non-LEP students experienced by the typical LEP student. Recall that this variable is always instrumented with the change that an LEP student would experience if his school were to implement California policies mechanically. We find that this variable has a statistically insignificant effect. Also, its inclusion only negligibly changes the estimated effect of percent bilingual. These two facts suggest that exposure to non-LEP students *per se* is not an important channel by which bilingual education affects the achievement of LEP students.

Finally, in column (7) we consider the possibility that eliminating bilingual education changes the *allocation* of resources within a school, most often by spreading small class sizes and higher-paid teachers evenly across LEP and non-LEP students (as opposed to having them allocated disproportionately to LEP students). We control for the change in the *range* of class size, an indicator of resource reallocation. We find that this explanatory variable does not affect LEP fourth graders' achievement in any subject; nor does it change the estimated effect of percent bilingual on their achievement. Thus, we conclude that reallocation of resources is probably not a key channel through the effect of bilingual education operates. Admittedly, we measure reallocation crudely, and a finer measure might produce statistically significant results.

In summary, we find that California's reduction in bilingual education decreased the achievement of LEP students in the fourth grade. Put another way, bilingual education was good for LEP fourth graders. The effects of bilingual education are modest (about 1/8th of a standard deviation in test scores for the mean change in percent bilingual) except in mathematics where the effects are non-modest (about 1/3rd of a standard deviation in test scores for the mean change in percent bilingual) in specifications where the habituation effect is purged. At this point, it may be useful to remember that the ban on bilingual education did not change how students were classified into the LEP and non-LEP categories. Not only did the classification rules remain unchanged; there is no indication in the data that the number or share of children being classified as LEP changed. Thus, our finding that the ban on bilingual reduced LEP students' achievement is *not* due to changing selection into the LEP group.

B. The Effects of Bilingual Education on the Achievement of Non-LEP Students in the Fourth Grade

Table 7 is just like Table 6, except that we show outcomes for fourth graders who are non-LEP (English proficient). As before, column (1) displays the least squares estimates of our simplest estimating equation for exposition, although it is likely that they are biased. The least squares results, both with and without the effect of habituation, suggest that eliminating bilingual education is good for the achievement of non-LEP students. We are confident that the least squares results are biased, however, because schools implementation of the ban was non-random. We focus on the results of the other columns in Table 7, in which we instrument for a school's actual change in percent bilingual with the change a school would have experienced if it had mechanically followed California policy. The column (3) results indicate reducing bilingual education had a strong, positive effect on the achievement of non-LEP fourth graders. A typical reduction of 0.12 in a school's percent bilingual raised the mean non-LEP fourth grader's reading score by 3.137 percentile points (0.22 standard deviations), his mathematics score by 2.255 percentile points (0.15 standard deviations), and his language score by 4.023 percentile points (0.30 standard deviations). Because the non-LEP results go in the opposite direction to the LEP results, a simple difference-in-differences estimator (in which non-LEP scores were *assumed* to be unaffected by bilingual education) would have greatly exaggerated the negative effect of bilingual education on LEP students' achievement. It is important, therefore, to allow bilingual education to affect both LEP and non-LEP students, as we do.

In column (4) of Table 7, we show what happens when we purge non-LEP students' achievement of the effects of test habituation. The estimated effects of bilingual education do shrink in absolute value, but the coefficients remain statistically significant. Moreover, we see very only minor changes in the estimated effects of bilingual education when we control for class size, control for changes in the percent of non-LEP students experienced by the average non-LEP student (correctly instrumented), or control for changes in resource reallocation (indicated by the change in the *range* of class size). For instance, our most sophisticated specification, in column (7), suggests that the mean non-LEP student in a school with the mean (0.12) reduction in bilingual education had his reading score rise by 2.745 percentile points (0.20 standard deviations), his mathematics score rise by 1.158 percentile points (0.09 standard deviations), and his language score rise by 3.405 percentile points (0.23 standard deviations). Quite simply, columns (5) through (7) give us little or no support for the ideas that class size is getting confounded with bilingual education, that bilingual education works largely through exposure to non-LEP students, or that bilingual education works largely through resource reallocation.

In summary, Table 7 suggests that–at least for fourth graders–the elimination of bilingual education improved non-LEP students' achievement. For a typical change in percent bilingual, the effects range from modest (about 1/10th of a standard deviation in mathematics) to non-modest (about 1/3rd of a standard deviation in language). It is interesting to observe that mathematics is the subject in which LEP students lost the most through the ban on bilingual education and that mathematics is the subject in non-LEP gained the least through the ban. We speculate that the addition of LEP children who knew some mathematics but had considerable trouble following English language explanations may have slowed down mathematics instruction for all students. Perhaps non-LEP students were less affected by the presence of LEP students in reading and language because schools recognized that their English language deficiencies *were* reading and language deficiencies and did not allow them to change the pace of learning for the whole class.

How do we explain the fact that the elimination of bilingual education appears to have improved achievement for non-LEP students? We doubt that exposure to LEP students is the channel for the improvement, not only because it seems unlikely behaviorally but because the exposure variable has negligible effects in Table 7. It is possible that reallocation of resources in the form of class size is the channel for the improvement, but our (admittedly crude) measure of reallocation gives us no support for such a conclusion. We find that controlling for schools' level of resources does not change the effect of bilingual education at all (not shown). Having eliminated all of the other channels through which bilingual education might affect non-LEP students, we conclude that at least one of following explanations are likely: (1) introducing LEP students into the non-LEP classroom may prompt changes in instructional style or curricular content experienced by non-LEP students (more on this in Section D); (2) eliminating bilingual education may have allowed schools to focus on the task of raising achievement in core subjects; (3) giving non-LEP students greater access to bilingual certified teachers (who are better educated, more highly paid, and younger) has beneficial effects for these students.

We would like to check that pre-existing school-level achievement trends, which could differ systematically with the treatment effect of the bilingual ban, do not bias our results. In particular, we want to demonstrate that prior to the implementation of Proposition 227, non-LEP test scores were not rising disproportionately in schools with higher shares of bilingual-eligible students. Because California implemented its STAR testing system only one year before the ban on bilingual education, we only have one year of test scores under bilingual education and cannot test for pre-existing non-LEP achievement trends in the state overall. We do have administrative data for the Los Angeles Unified School District's (LAUSD) administration of the California Test of Basic Skills (CTBS) in the 1993-94 to 1995-96 school years. In those years, LAUSD was testing Spanish-speaking students with a separate test administered in Spanish (the APRENDA test), so the CTBS scores are largely determined by non-LEP students. We find no pre-existing trends related to the share of a school's enrollment that is LEP or bilingual-eligible in non-LEP scores on the reading, math, or language CTBS tests for the fourth grade.

C. Results for Grades Three Through Eight, For Both LEP and Non-LEP Students

We now focus on one specification that we believe to be highly reliable, and we compare the results of this specification for all available grades and subjects, for both LEP and non-LEP students. We show these comparisons in Tables 8 and 9. We focus on the specification shown in column (6) of Tables 6 and 7. It is an instrumental variables specification in which we purge the achievement variables of habituation, control for changes in mean class size, and control for changes in exposure to non-LEP students. We also present results from variants of this main specification for grades three through five in the appendix tables.[13] Our main findings are robust to the various specifications.

Among LEP students, the negative effects of California's reduction in bilingual education seem to have been confined to the earlier grades. The achievement of third and fourth grade LEP students was reduced in all three subjects: reading, math, and language. Among fifth and sixth grade LEP students, only mathematics achievement

[13]Table A-3 shows OLS estimates without purging habituation effects and Table A-4 shows IV estimates without purging habituation effects. Table A-5 gives IV estimates weighted by enrollment. Table A-6 presents IV estimates for changes in the first year following the implementation of Proposition 22,7 and Table A-7 presents IV estimates for changes in the first two years following implementation.

was reduced by the ban; there were no statistically significant effects on reading or language. Among seventh and eighth grade LEP students, achievement was not statistically significantly affected in any subject. The differences between the lower and upper grade LEP students are probably not caused by different rates of English language acquisition: most commentators believe that younger children learn languages slightly faster than older children. If we keep in mind the fact that *all* LEP students in the "after-the-ban" group have spent at least 3 years in English immersion, we can speculate usefully on the differences between the results for the lower and upper grades. An LEP fifth grader is likely to know more English than an LEP first grader, even if the fifth grader was in bilingual education up through fourth grade. Thus, students who had already had some schooling when the ban occurred may have been better equipped for English immersion. Alternatively, it may be that the effects of English immersion are simply more positive in higher grades where material is more challenging. In the third grade, recognizing simple words, numbers, and letters guarantees a reasonable score. By grades five through eight, a student cannot do well unless he is comfortable with English sentence structure, has a reasonably wide vocabulary, and can process mathematical word problems. It is not at all surprising that the negative effects of the ban persist most (to the highest grade) in mathematics, where material learned in another language is likely to be useful on an English language test. Conversely, the negative effects of the ban die out most quickly (at the lowest grade) in reading and language where an LEP student's greater knowledge of English offsets any confusion he experiences when being taught in the English language.[14]

Among non-LEP students, California's reduction in bilingual education improved achievement consistently up through the seventh grade and then had no statistically significant effect for the eighth grade. The positive effect on reading is statistically significant for third through seventh grades; the positive effect on math is statistically significant for third through fifth grades and also for seventh grade; the positive effect on language is statistically significant for third through sixth grades. Non-LEP students benefitted the most in their reading and language achievement, but also benefitted to a statistically significant degree in mathematics achievement. The differences by

[14] We also considered whether the better results for fifth graders could be due to their being better prepared for the shock of English immersion when it arrives. Thus, even if fifth graders learned English more slowly than third graders, they would know more English than third graders when they were suddenly thrust into English language classes. However, we dismissed this line of reasoning when we realized that all of the LEP students represented as "after the ban" in Table 8 had experienced at least 2.75 school years of English immersion before being tested in the 2000-01 school year. No "after the ban" student in Table 8 is still in the immediate shock stage.

subject in non-LEP students' achievement may be caused by differences in how teachers integrate LEP students in different subjects. In mathematics, the typical LEP student knows more than he can express in English, so teachers may struggle to include them in mathematics lessons that are appropriate for their level of knowledge but that confuse them because of language difficulties. In reading and language, the typical LEP student is obviously behind grade level so perhaps he can be smoothly integrated into lessons for native English speakers who need remedial attention.

D. Results throughout the Achievement Distribution

Thus far, we have examined changes in mean test performance. It is unlikely, however, that the various changes induced by the new English immersion regime would affect all LEP students uniformly, or all non-LEP students uniformly. For example, the effect of introducing an English immersion curriculum on an LEP student's academic performance may well depend on *how* limited the student's English proficiency is, as well as the student's achievement level. The peer effects of introducing LEP students into the classroom may differ among non-LEP students by their levels of academic proficiency. In Tables 10 and 11, we present results for how the switch from bilingual education to English immersion affects mean test scores in reading, math, and language for students in grades three through five, as well as how it affects the share of students scoring above the 25th and 75th national percentiles. Table 10 shows results for LEP students and Table 11 for non-LEP students.

In these tables, we have not accounted for students' becoming habituated to the test over the three-year period. We were able to do so in the analysis of changes in *mean* test scores presented in Tables 6 through 9 because the Arizona data include variables for mean scores by grade and subject, but cannot do so for other measures of change in the distribution because Arizona did not report them. Because students scoring in different parts of the test score distribution likely habituate at distinct paces, rather than making any assumptions about the degree of habituation, we analyze the data as they are. The effects on mean scores presented in Tables 10 and 11 thus differ from the mean score results presented earlier, but are defined consistently with the other distributional measures in these tables.

The first three columns of Table 10 present three-year IV results for effects on reading scores, the next three for math, and the final three for language, all for LEP students only. Within each subject, the first column shows estimated effects on mean test scores, the second column shows the effect on the percent of students scoring above the 25th percentile, and the third column on the percent of students scoring above the 75th percentile. The results are presented by grade for grades three through five. Within each grade, the first row provides the mean of each measure for the 1997-98 school year, before the implementation of Proposition 227. The second row gives the effect of a 100 percent decline in bilingual enrollments, the third row estimates the effect (in percentage points) given the mean decline in bilingual enrollments for the relevant grade (between 11 and 13 percent for grades three through five), and the fourth row gives this mean estimated effect in standard deviations. Table 11 is set up identically for non-LEP students only.

Table 10 shows that the modest but significant declines observed in mean LEP performance when we accounted for habituation become smaller and less frequently significant when we no longer account for habituation. These small effects at the mean often mask absolutely larger effects at different points in the distribution, however, with the switch to English immersion typically modestly increasing or having no effect on the share of students who exceed the 25th (national) percentile and decreasing the share of students exceeding the 75th percentile. In contrast, Table 11 shows that significant mean gains in achievement for non-LEP students persist even when we do not account for habituation, and that these mean gains generally are driven by large increases in the share of students scoring above the 25th percentile, with little effect on the share of students scoring above the 75th percentile.

The patterns in which parts of the LEP and non-LEP distributions benefit by subject area suggest possible explanations of *how* the shift to English immersion affects students. For example, the strongest pattern among LEP students is a significant decline in the share of students exceeding the 75th percentile. For fourth and fifth grade LEP students, there are no significant effects of the switch on the share of students exceeding the 25th percentile in math, and for third graders, there is a small but significant *increase* in that share. Meanwhile, lower achieving non-LEP students appear to be benefitting, with significant increases in the share scoring above the 25th percentile (about one-quarter of a standard deviation for third through fifth graders). This is consistent with the typical non-LEP student experiencing a change in the math classroom as more LEP students are introduced. The teacher may teach more slowly or repetitively because of LEP students' English skills, and may also gear the curriculum more towards the lower end of the achievement distribution. This change benefits lower achieving non-LEP students, but hurts LEP students who were performing well in math when they were taught in their native language and were possibly receiving a different quality of instruction (at a faster pace or higher level).

Results for reading and language are relatively similar; here we discuss the language results, which are somewhat stronger for LEP students, only. (Note that both reading and language results are less consistent for LEP students across grades than are the math results.) For LEP students, the most consistent finding for language scores is a modest boost to low achievers (those exceeding the 25th percentile), with mixed findings for high achievers (those exceeding the 75th percentile). This suggests that for those with very limited English proficiency, English immersion is somewhat helpful. Low-achieving non-LEP students benefit from this change even more than low-achieving LEP students. The share of non-LEP students exceeding the 25th percentile increases nearly uniformly across grades by about 0.30 standard deviations; LEP students experience increases in this measure ranging from 0.08 to 0.21 standard deviations. As in math, this suggests that the teacher's teaching style may change to reflect the introduction of English learners into the typical classroom experienced by non-LEP students.

E. Summing Up

It is a mistake to expect the LEP and non-LEP students to be parallel or opposite or have any other simple pattern. Non-LEP students experience few offsetting effects: they enjoy the efficiency benefits of the ban and any reduction in spillovers from their peers is likely to be small simply because LEP students are usually a minority. LEP students experience numerous offsetting effects: English skills and peers may be better but teaching is likely to be less well targeted to their immediate needs. We are not surprised, for instance, to find no statistically significant effects in reading and language for LEP students in grades five and up, while we do find positive effects for non-LEP students in the corresponding subjects and grades. For LEP students, effects cancel one another out at the mean; for non-LEP students, the key negative effects simply do not exist at the mean.

Are these changes in achievement relatively small or large? In Table 12, we redisplay our results by the showing the effect of the typical decrease in percent bilingual as a result of Proposition 227, which is different for different grades. The typical decrease was 0.12 or 12 percent in grades three through five, 5.6 percent in grade six, and 3.7 percent in grades seven and eight. We present the results in terms of both percentile rank points and fractions of a standard deviation in scores (for the relevant test). We show the percentile rank points results simply so that readers can see that the numbers came from the previous table. We focus on the effects measured in fractions of a standard deviation because these allow us to judge magnitudes. For LEP students, the most substantial effects of the ban were in mathematics, where up to 0.27 of a standard deviation was lost in grades three through five. All of the reading and language effects are less substantial, and no effect is statistically significant in grades six through eight. For non-LEP students, the most substantial effects of the ban were in reading and language, where up to 0.3 of a standard deviation was gained in grades three through seven. The mathematics effects are less substantial for non-LEP students. Our results are, thus, neither negligible nor dramatic. Certainly, the worst "gloom and doom" predictions for the ban are not born out. Similarly, the most glowing encomiums of bilingual education are not born out.

Because Table 12 indicates that the effects of the ban were systematically more positive for LEP students in higher grades, it is possible that English immersion is initially hard on students who have limited knowledge of English but that, as a student's knowledge of English grows, positive effects of English immersion increasingly counterbalance negative effects.

It is worth highlighting the fact that, if we do not account for test habituation, the effects of English immersion on LEP students are generally *not* negative, even for students in the lower grades. Without accounting for habituation, the effect of LEP students are generally statistically insignificant and small in magnitude. As a result, observers of California schools will not have seen a decline in LEP students' performance in the first few years of English immersion. Instead, the LEP students have not improved as much as we would have expected them to improve, given that their scores would have predictably have risen due to habituation alone.

VII. Conclusions

We investigate the effects of Proposition 227, which eliminated bilingual education in many schools in California. We effectively use Proposition 277 as a natural experiment to learn whether bilingual education is good or bad for the achievement of LEP and non-LEP students. We find that LEP students' achievement is hurt in grades three and four when bilingual education is reduced, especially in mathematics where the negative effect extends to the fifth grade. For grades six through eight, LEP students suffer no averse consequences of the ban on bilingual education. We find that non-LEP students' achievement is generally improved when bilingual education is reduced. This improvement holds grades three through seven (not eight) and is especially large in reading and language. The overall effect of the ban on bilingual education is, thus, more positive for higher elementary grades. This pattern can bear several interpretations, but one that is particularly interesting is the idea that English immersion has different dynamic effects depending on a student's initial knowledge of English. For instance, the English immersion might be a good long-term investment even for students who initially lose ground because they know so little English when "immersed."

We set up a model that illustrates all of the various channels through which bilingual education can work. This proves helpful when we come to explain the results for non-LEP students, who cannot have experienced *direct* effects of the elimination of bilingual education (that is, their language of instruction did not change). We do not find evidence, however, that the effects of bilingual education work largely through exposure to non-LEP students or through the reallocation of resources (such as small class sizes and higher-paid teachers) away from LEP and toward non-LEP students. Also, our results supply little support for the hypothesis that younger LEP students will do better in the switch to English immersion because they learn new languages faster. We are very careful to eliminate bias due to schools' implementing the ban on bilingual education differently, according to their attitudes towards bilingual education. We are also careful to eliminate the potentially confounding effects of habituation to the test and California's 1996 class size reduction.

References

Arizona Department of Education. Stanford 9 Achievement Test Results. 1998, 1999, 2000. http://www.ade.state.az.us/standards/stanford9/

Alanis, Iliana. "A Texas Two-Way Bilingual Program: Its Effects on Linguistic and Academic Achievement." *Bilingual Research Journal*, Vol 24, Number 3 (Summer 2000), pp. 225-48.

Baker, Susan and Kenji Hakuta, "Bilingual Education and Latino Civil Rights," paper presented at Latino Civil Rights Conference, the Civil Rights Project, Harvard University, December 1997.

Borjas, George J., "Ethnic Enclaves and Assimilation," *Swedish Economic Policy Review*, Vol. 7, No. 2 (2000), 89-122.

Borjas, George J., "Ethnicity, Neighborhoods, and Human Capital Externalities," *American Economic Review*, Vol. 85, No. 3 (June 1995), 365-390.

Burkheimer, Jr., G.J., A.J. Conger, G.H. Dunteman, B.G. Elliott, and K.A. Mowbray, "Effectiveness of Services for Language-Minority Limited-English-Proficiency Students." Volumes I and II. Technical report. Research Triangle Institute, Research Triangle Park, 1989.

Calderon, Margarita Espino. "Promoting Language Proficiency and Academic Achievement through Cooperation," *ERIC Digest*, December 1999.

California Department of Education. STAR Research Files. 1998, 1999, 2000, 2001. http://star.cde.ca.gov/

California Department of Education, Educational Demographics Office. CBEDS Professional Assignment Information Form. 1998, 1999, 2000, 2001. http://www.cde.ca.gov/demographics/files/paif.htm

California Department of Education, Educational Demographics Office. Free and Reduced Price Meals/CalWORKS School Level Files. 1998, 1999, 2000, 2001. http://www.cde.ca.gov/demographics/files/afdc.htm#fs

California Department of Education, Educational Demographics Office. Language Census Data Files. 1998, 1999, 2000, 2001. http://www.cde.ca.gov/demographics/files/census.htm

Cheng, Chang-kuo. *Impact of Bilingual Education on Educational Attainment*. Ph.D. Dissertation. University of Maryland, 1996.

Chiswick, Barry R., "Speaking, Reading, and Earnings among Low-Skilled Immigrants," *Journal of Labor Economics*, Vol. 9, No. 2 (April 1991), 149-170.

Chiswick, Barry R. and Paul W. Miller, "Language in the Immigrant Labor Market," in *Immigration, Language, and Ethnicity: Canada and the United States* (Chiswick, Barry R., ed.) American Enterprise Institute, Washington (1992), 229-296.

Edin, Pers-Anders, Peter Fredriksson, and Olof Aslund, "Ethnic Enclaves and the Economic Success of Immigrants – Evidence from a Natural Experiment," unpublished manuscript, January, 2002.

Federman, Maya. *Bilingual Education, College Major Choice, and College Remediation*.. Ph.D. Dissertation. Harvard University, 2000.

Garcia, Anita and Cynthia Morgan, "A 50 State Survey of Requirements for the Education of Language Minority Children," mimeo, The READ Institute, November 1997.

Garcia, Eugene, “Implementation of Proposition 227: 1998-2000,” *Bilingual Research Journal*, Vol. 24, Nos. 1 and 2 (Winter and Spring 2000).

Johnson, Robert J. Case Studies of Expectation Climate at Two Bilingual Education Schools. *Bilingual Research Journal*, Vol. 24, Number 3 (Summer 2000) , pp. 261-76.

Kucer, Stephen B., and Cecilia Silva. "The English Literacy Development of Bilingual Students within a Transitional Whole-Language Curriculum," *Bilingual Research Journal*, Volume 23, Number 4 (Fall 1999), pp. 347-71.

Lazear, Edward P., “Culture and Language,” *Journal of Political Economy*, Vol. 107, No. 6, Pt. 2 (December 1999), S95-S126.

Lopez, Mark H. *The Educational and Labor Market Impacts of Bilingual Education in the Short and Long Run: Evidence from the National Education Longitudinal Study of 1988 and High School and Beyond*. Ph.D. Dissertation. Princeton University, 1996.

Los Angeles Unified School District, Information Technology Division. CTBS and APRENDA Test Scores: Norm-Referenced Test Results. 1993-1994, 1994-1995, 1995-1996.

Manzo, Kathleen K., “California Text Adoption Puts Emphasis on Phonics,” *Education Week*, January 15, 1997.

Meyer, M. M. and S. E. Fienberg, eds. *Assessing Evaluation Studies: The Case of Bilingual Education Strategies*, National Research Council, 1992.

Mora, Marie. *English Proficiency, Bilingual Education, and the Earnings of Hispanic Americans*. Ph.D. Dissertation. Texas A&M University, 1996.

Ramírez, J., S. Yuen, D. Ramey, and D. Pasta. *Final Report: Longitudinal study of structured English immersion strategy, early-exit and late exit bilingual education programs for language-minority children*. Vol. I and II. Prepared for the US Department of Education. San Mateo, CA: Aguirre International, 1991.

Riojas-Cortez, Mari. "It's All about Talking: Oral Language Development in a Bilingual Classroom.," *Dimensions of Early Childhood*, Volume 29, Number 1 (Winter 2001), pp. 11-15.

U.S. Department of Education, National Center for Education Statistics. Public Elementary/Secondary School Universe Survey Data. http://nces.ed.gov/ccd/pubschuniv.html

Vaishnav, Anand, “Bilingual ed plan offered by legislators,” *Boston Globe*, January 30, 2002.

Yettick, Holly, “Survey shows support for bilingual limits,” *Denver Rocky Mountain News,* February 25, 2002.

Table 1
Bilingual, Limited English Proficient (LEP), and Total Enrollments for the Mean* California School

	PRE-227		POST-227	
	1997-98	1998-99	1999-2000	2000-01
Bilingual enrollment	152.130	48.185	51.369	49.041
	(181.776)	(103.619)	(110.879)	(105.618)
Bilingual as % of LEP	0.351	0.114	0.112	0.113
	(0.287)	(0.207)	(0.204)	(0.209)
Bilingual as % of total	0.179	0.059	0.061	0.059
	(0.175)	(0.120)	(0.124)	(0.123)
LEP enrollment	386.396	375.702	381.384	379.141
	(276.069)	(270.194)	(271.418)	(266.700)
LEP as % of total	0.454	0.461	0.467	0.467
	(0.205)	(0.205)	(0.203)	(0.198)
Total enrollment	773.038	782.900	782.889	787.848
	(279.583)	(286.049)	(288.048)	(291.191)
Schools in sample	1345	1345	1345	1345

*Table shows means and standard deviations for all California public schools that have at least ten LEP fourth graders. Standard deviations are in parentheses. Data are not weighted.

Source: California Department of Education Language Census.

Table 2
LEP Students in Bilingual Education After Proposition 227 in California

	Percent of LEP Enrollment in Bilingual Education, 2000-01					
	0%	0-20%	20-50%	50-80%	80-100%	sum
Percent of LEP Enrollment in Bilingual Education, 1997-98						
0%	94.43	3.43	1.71	0.43	0.00	1.00
0-20%	74.59	13.26	13.26	8.84	2.76	1.00
20-50%	60.89	18.39	16.91	3.81	0.00	1.00
50-80%	44.67	15.78	25.20	12.50	1.84	1.00
80-100%	17.50	5.00	15.00	37.50	25.00	1.00

Note: Table includes all California schools that have at least ten LEP fourth graders.
Source: California Department of Education Language Census.

Table 3
Class Size, By Grade, Before and After Proposition 227

	PRE-227	POST-227		
	1997-98	1998-99	1999-2000	2000-01
Grade 3				
Mean class size	23.794	20.700	19.921	19.635
Standard deviation, class size	5.844	4.001	3.003	3.266
Grade 4				
Mean class size	29.495	29.471	29.333	29.152
Standard deviation, class size	4.366	2.915	3.418	3.473
Grade 5				
Mean class size	29.577	29.845	29.577	29.520
Standard deviation, class size	3.312	3.302	3.228	3.405
Grade 6	30.000	30.394	30.330	30.020
Mean class size	4.165	5.251	6.249	4.117
Standard deviation, class size				

Note: Table includes all California schools that have at least ten LEP students in relevant grade.
Source: California Department of Education Professional Assignment Information Form.

Table 4
Descriptive statistics, changes from 1997-98 to 2000-01

	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
change % classes bilingual	-0.116	-0.121	-0.124	-0.056	-0.037	-0.037
	(0.154)	(0.155)	(0.157)	(0.116)	(0.090)	(0.089)
change % non-LEP students in classes experienced by LEP students	0.078	0.085	0.083	0.048	0.028	0.030
	(0.138)	(0.143)	(0.143)	(0.121)	(0.103)	(0.100)
change % non-LEP students in classes experienced by non-LEP students	-0.144	-0.152	-0.149	-0.091	-0.045	-0.042
	(0.182)	(0.180)	(0.175)	(0.154)	(0.104)	(0.103)
change mean class size	-4.400	-0.269	-0.075	-0.048		
	(5.265)	(3.646)	(3.514)	(4.889)		
change reading score, LEP[a]	8.342	6.680	4.298	4.454	3.505	2.865
	(7.915)	(7.846)	(6.847)	(7.268)	(6.820)	(6.310)
change reading score, non-LEP[a]	14.527	12.401	9.221	7.512	7.410	5.825
	(9.131)	(9.165)	(8.766)	(8.019)	(6.501)	(7.141)
change math score, LEP[a]	17.184	12.526	11.184	10.545	5.857	4.625
	(12.342)	(10.809)	(11.277)	(10.478)	(8.574)	(7.548)
change math score, non-LEP[a]	21.598	17.709	16.070	13.010	9.334	7.198
	(12.496)	(11.441)	(11.363)	(10.801)	(7.332)	(8.929)
change language score, LEP[a]	12.149	8.364	6.677	6.666	5.163	3.557
	(10.039)	(9.064)	(9.327)	(9.251)	(8.910)	(7.797)
change language score, non-LEP[a]	17.476	11.943	10.827	8.998	8.951	6.813
	(10.656)	(9.933)	(10.224)	(9.460)	(7.638)	(8.385)
number schools in sample	1348	1286	1195	522	558	546

Notes: Table contains means and standard deviations for all California schools that have at least ten LEP students in the relevant grade. Standard deviations are in parentheses.
a. This table reports changes in percentile rank scores that have not yet been purged of the habituation effect.
Sources: California Department of Education Language Census, Professional Assignment Information Form, and STAR Research Files.

Table 5
Descriptive statistics, base year (1997-98) test scores by LEP status

	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
base year reading score, LEP	16.239	16.207	14.963	18.379	14.814	17.629
	8.568	8.252	6.727	(7.688)	(7.621)	(7.518)
base year reading score, non-LEP	31.257	33.309	33.804	40.622	40.716	43.996
	14.294	14.680	13.430	(14.249)	(15.729)	(14.439)
base year math score, LEP	27.338	22.350	21.639	27.569	26.005	26.975
	13.185	10.898	10.549	(12.662)	(13.584)	(14.038)
base year math score, non-LEP	36.496	33.979	35.709	45.468	43.156	43.842
	16.406	15.269	14.939	(16.277)	(16.499)	(15.890)
base year language score, LEP	21.384	24.599	22.096	24.648	22.259	21.914
	10.858	9.831	9.170	(9.650)	(10.194)	(9.314)
base year language score, non-LEP	34.034	39.301	39.013	45.004	47.859	46.663
	14.850	14.119	14.416	(14.716)	(16.334)	(15.168)
number schools in sample	1408	1345	1250	904	582	571

Notes: Table contains means and standard deviations for all California schools that have at least ten LEP students in the relevant grade. Standard deviations are in parentheses.
Source: California Department of Education STAR Research Files.

Table 6

Effect of Bilingual Education on Fourth Graders who are LEP: Three-Year Changes

Regression type	Mean of Var	(1) OLS	(2) OLS	(3) IV	(4) IV	(5) IV	(6) IV	(7) IV
Habit. purged		no	yes	no	yes	yes	yes	yes
READING (dependent variable is mean percentile rank in reading of LEP fourth graders in school)								
change, share bilingual	-0.121	-0.286	2.420*	6.478**	9.425***	9.376***	8.804***	8.820***
	(0.155)	(1.245)	(1.373)	(2.953)	(3.237)	(3.220)	(2.855)	(3.240)
change, mean class size	-0.269					-0.021	-0.022	-0.036
in grade	(3.646)					(0.060)	(0.061)	(0.066)
change, share free/red	0.004					-5.397*	-6.568**	-6.577**
lunch elig	(0.089)					(2.964)	(2.729)	(3.062)
change, share non-LEP	0.085						-9.175	-9.122
students experienced	(0.143)							(5.594)
change, range of class	0.362							-0.029
size in grade	(4.677)							(0.059)
constant		6.684***	2.497***	7.507***	3.341***	3.309***	4.025***	4.029***
		(0.280)	(0.302)	(0.478)	(0.509)	(0.507)	(0.477)	(0.731)
effect for mean change in share bilingual (-0.12)		0.033	-0.283	-0.758	-1.103	-1.097	-1.030	-1.032
MATH (dependent variable is mean percentile rank in math of LEP fourth graders in school)								
change, share bilingual	-0.121	2.321	6.778***	15.430***	26.417***	25.609***	25.120***	25.145***
	(0.155)	(1.854)	(1.978)	(4.095)	(4.414)	(4.336)	(4.414)	(4.415)
change, mean class size	-0.269					-0.190**	-0.191**	-0.214**
in grade	(3.646)					(0.082)	(0.084)	(0.088)
change, share free/red	0.004					-8.827**	-9.828**	-9.843**
lunch elig	(0.089)					(3.891)	(4.020)	(4.032)
change, share non-LEP	0.085						-7.845	-7.764
students experienced	(0.143)						(6.584)	(6.558)
change, range of class	0.362							-0.045
size in grade	(4.677)							(0.077)
constant		12.654***	3.509***	14.234***	5.876***	5.701***	6.313***	6.319***
		(0.382)	(0.400)	(0.622)	(0.651)	(0.643)	(0.824)	(0.826)
effect for mean change in share bilingual (-0.12)		-0.272	-0.793	-1.805	-3.091	-2.996	-2.939	-2.942
LANGUAGE (dependent variable is mean percentile rank in language of LEP fourth graders in school)								
change, share bilingual	-0.121	-2.904**	1.297	1.615	9.351**	8.642**	8.180**	8.192**
	(0.155)	(1.468)	(1.577)	(3.485)	(3.791)	(3.715)	(3.698)	(3.703)
change, mean class size	-0.269					-0.037	-0.038	-0.048
in grade	(3.646)					(0.068)	(0.069)	(0.073)
change, share free/red	0.004					-8.513**	-9.459***	-9.466***
lunch elig	(0.089)					(3.489)	(3.571)	(3.578)
change, share non-LEP	0.085						-7.413	-7.375
students experienced	(0.143)						(6.031)	(6.009)
change, range of class	0.362							-0.021
size in grade	(4.677)							(0.066)
constant		8.025***	2.663***	8.574***	3.636***	3.535***	4.114***	4.117***
		(0.323)	(0.340)	(0.549)	(0.581)	(0.575)	(0.799)	(0.801)
effect for mean change in share bilingual (-0.12)		0.340	-0.152	-0.189	-1.094	-1.011	-0.957	-0.958
Observations		1286	1286	1284	1284	1248	1248	1248

Robust standard errors in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 7

Effect of Bilingual Education on Fourth Graders who are not LEP

	Mean of Var	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Regression type		OLS	OLS	IV	IV	IV	IV	IV
Habit. purged		no	yes	no	yes	yes	yes	yes
READING (dependent variable is mean percentile rank in reading of LEP fourth graders in school)								
change, share bilingual	-0.121	-10.906***	-8.213***	-26.813***	-23.231***	-22.252***	-23.406***	-23.463***
	(0.155)	(1.712)	(1.744)	(3.552)	(3.658)	(3.665)	(4.230)	(4.259)
change, mean class size in grade	-0.269					-0.023	-0.017	-0.005
	(3.646)					(0.072)	(0.074)	(0.082)
change, share free/red lunch elig	0.004					-9.775***	-10.428***	-10.440***
	(0.089)					(2.878)	(3.274)	(3.284)
change, share non-LEP students experienced	-0.152						3.23	3.346
	(0.180)						(7.269)	(7.341)
change, range of class size in grade	0.362							0.024
	(4.677)							(0.068)
constant		11.083***	6.866***	9.144***	5.027***	5.153***	5.500***	5.506***
		(0.328)	(0.340)	(0.458)	(0.478)	(0.477)	(0.942)	(0.945)
effect for mean change in share bilingual (-0.12)		1.276	0.961	3.137	2.718	2.603	2.739	2.745
MATH (dependent variable is mean percentile rank in math of LEP fourth graders in school)								
change, share bilingual	-0.121	-8.965***	-4.400**	-19.270***	-7.881*	-7.670*	-9.739*	-9.901**
	(0.155)	(2.178)	(2.193)	(4.295)	(4.210)	(4.224)	(4.992)	(5.024)
change, mean class size in grade	-0.269					-0.149*	-0.139	-0.103
	(3.646)					(0.083)	(0.084)	(0.093)
change, share free/red lunch elig	0.004					-11.576***	-12.727***	-12.759***
	(0.089)					(3.791)	(4.083)	(4.083)
change, share non-LEP students experienced	-0.152						5.824	6.154
	(0.180)						(8.530)	(8.645)
change, range of class size in grade	0.362							0.069
	(4.677)							(0.078)
constant		16.391***	7.244***	15.124***	6.801***	6.789***	7.418***	7.434***
		(0.410)	(0.414)	(0.586)	(0.588)	(0.586)	(1.124)	(1.130)
effect for mean change in share bilingual (-0.12)		1.049	0.515	2.255	0.922	0.897	1.139	1.158
LANGUAGE (dependent variable is mean percentile rank in language of LEP fourth graders in school)								
change, share bilingual	-0.121	-14.522***	-10.315***	-34.387***	-26.166***	-25.024***	-26.271***	-26.313***
	(0.155)	(1.951)	(1.935)	(4.002)	(3.932)	(3.951)	(4.581)	(4.613)
change, mean class size in grade	-0.269					-0.049	-0.042	-0.033
	(3.646)					(0.079)	(0.081)	(0.087)
change, share free/red lunch elig	0.004					-8.747***	-9.441***	-9.449***
	(0.089)					(3.276)	(3.621)	(3.628)
change, share non-LEP students experienced	-0.152						3.511	3.597
	(0.180)						(7.424)	(7.495)
change, range of class size in grade	0.362							0.018
	(4.677)							(0.075)
constant		10.188***	4.803***	7.766***	2.864***	3.021***	3.401***	3.405***
		(0.347)	(0.351)	(0.493)	(0.502)	(0.505)	(0.979)	(0.982)
effect for mean change in share bilingual (-0.12)		1.699	1.207	4.023	3.061	2.928	3.074	3.079
Observations		1273	1273	1271	1271	1242	1242	1242

Robust standard errors in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 8
Effect of Bilingual Education on Third, Fourth, and Fifth Graders

	Third Graders		**Fourth Graders**		**Fifth Graders**	
	LEP	Non-LEP	LEP	Non-LEP	LEP	Non-LEP
Regression type	IV	IV	IV	IV	IV	IV
Habit. effect purged	yes	yes	yes	yes	yes	yes
READING (dependent variable is mean percentile rank in reading)						
change, share bilingual	16.144***	-15.407***	9.376***	-22.252***	2.646	-32.511***
	(3.230)	(3.328)	(3.220)	(3.665)	(3.036)	(4.709)
change, mean class size in grade	-0.04	0.016	-0.021	-0.023	-0.022	0.033
	(0.044)	(0.048)	(0.060)	(0.072)	(0.059)	(0.081)
change, share free/red lunch elig	-4.278	-8.651***	-5.397*	-9.775***	-5.272***	-6.319**
	(2.664)	(2.756)	(2.964)	(2.878)	(1.951)	(2.792)
constant	5.141***	7.868***	3.309***	5.153***	2.552***	3.129***
	(0.513)	(0.461)	(0.507)	(0.477)	(0.487)	(0.579)
effect for mean change in share bilingual (-0.12)	-1.889***	1.803***	-1.144***	2.715***	-0.331	4.064***
MATH (dependent variable is mean percentile rank in math)						
change, share bilingual	18.935***	-4.058	25.609***	-7.670*	17.353***	-10.527**
	(4.714)	(4.575)	(4.336)	(4.224)	(4.684)	(5.180)
change, mean class size in grade	-0.337***	-0.275***	-0.190**	-0.149*	-0.238**	-0.09
	(0.066)	(0.065)	(0.082)	(0.083)	(0.099)	(0.099)
change, share free/red lunch elig	-4.239	-9.866***	-8.827**	-11.576***	-8.824***	-9.544***
	(3.648)	(3.771)	(3.891)	(3.791)	(3.284)	(3.261)
constant	5.064***	6.930***	5.701***	6.789***	4.126***	5.510***
	(0.722)	(0.627)	(0.643)	(0.586)	(0.733)	(0.692)
effect for mean change in share bilingual (-0.12)	-2.215***	0.475	-3.124***	0.936*	-2.169***	1.316**
LANGUAGE (dependent variable is mean percentile rank in language)						
change, share bilingual	15.134***	-19.432***	8.642**	-25.024***	-1.655	-37.102***
	(3.935)	(3.951)	(3.715)	(3.951)	(3.861)	(5.230)
change, mean class size in grade	-0.113**	-0.064	-0.037	-0.049	-0.031	0.069
	(0.053)	(0.054)	(0.068)	(0.079)	(0.083)	(0.087)
change, share free/red lunch elig	-6.125*	-6.937**	-8.513**	-8.747***	-7.264***	-4.726
	(3.191)	(3.352)	(3.489)	(3.276)	(2.690)	(3.425)
constant	3.311***	4.709***	3.535***	3.021***	1.043*	0.822
	(0.614)	(0.527)	(0.575)	(0.505)	(0.627)	(0.645)
effect for mean change in share bilingual (-0.12)	-1.771***	2.274***	-1.054**	3.053***	0.207	4.638***
Observations	1335	1320	1248	1242	1140	1134

Robust standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 9
Effect of Bilingual Education on Sixth, Seventh, and Eighth Graders

	Sixth Graders		**Seventh Graders**		**Eighth Graders**	
	LEP	Non-LEP	LEP	Non-LEP	LEP	Non-LEP
Regression type	IV	IV	IV	IV	IV	IV
Habit. effect purged	yes	yes	yes	yes	yes	yes
READING (dependent variable is mean percentile rank in reading)						
change, share bilingual	11.668	-35.267**	9.710	-34.808***	7.541	-14.310
	(9.738)	(18.579)	(9.692)	(9.690)	(9.287)	(19.166)
change, share non-LEP students experienced	2.938	-22.701**	-14.108*	2.153	-5.281	10.680
	(10.365)	(10.160)	(8.500)	(8.270)	(8.402)	(8.722)
change, mean class size in grade	0.048	0.071				
	(0.065)	(0.086)				
constant	3.445***	0.971	4.234***	6.380***	3.830***	6.356***
	(0.883)	(1.198)	(0.569)	(0.561)	(0.529)	(0.578)
effect for mean change in share bilingual (-0.056 for 6th, -0.037 for 7th and 8th)	-0.653	1.975**	-0.359	1.288***	-0.279	0.529
MATH (dependent variable is mean percentile rank in math)						
change, share bilingual	32.506**	-22.806	-3.663	-22.536**	11.777	22.513
	(16.584)	(22.556)	(10.105)	(9.416)	(9.726)	(22.769)
change, share non-LEP students experienced	-2.16	-23.329**	-16.115**	1.608	-11.934	2.418
	(15.938)	(11.714)	(8.072)	(7.560)	(8.111)	(10.259)
change, mean class size in grade	0.114	0.106				
	(0.104)	(0.118)				
constant	5.232***	0.925	-0.424	1.998***	-1.354	1.456**
	(1.359)	(1.440)	(0.609)	(0.481)	(0.616)	(0.680)
effect for mean change in share bilingual (-0.056 for 6th, -0.037 for 7th and 8th)	-1.820**	1.278	0.136	0.834**	-0.436	-0.833
LANGUAGE (dependent variable is mean percentile rank in language)						
change, share bilingual	-0.084	-51.695**	28.364**	-17.686	8.241	6.553
	(12.932)	(23.478)	(13.395)	(11.475)	(11.478)	(21.911)
change, share non-LEP students experienced	2.705	-32.19***	-13.394	5.880	-16.201**	14.347
	(13.527)	(12.520)	(12.138)	(9.032)	(7.898)	(10.583)
change, mean class size in grade	0.146	0.071				
	(0.081)	(0.104)				
constant	2.748**	-2.124	2.257***	4.436***	-0.823	2.599***
	(1.213)	(1.437)	(0.746)	(0.674)	(0.653)	(0.681)
effect for mean change in share bilingual (-0.056 for 6th, -0.037 for 7th and 8th)	0.005	2.894**	-1.049**	0.654	-0.305	0.242
Observations	522	515	558	553	546	546

Robust standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 10

Effect of Bilingual Education on LEP Students throughout the Test Score Distribution: Three-Year Changes

	Reading			Math			Language		
	mean NPR score	% scoring above 25th %ile NPR	% scoring above 75th %ile NPR	mean NPR score	% scoring above 25th %ile NPR	% scoring above 75th %ile NPR	mean NPR score	% scoring above 25th %ile NPR	% scoring above 75th %ile NPR
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Grade 3									
mean in 1998	16.119	29.565	1.499	27.161	46.628	6.828	21.245	36.700	3.500
	8.584	17.298	3.381	13.209	19.174	9.106	10.866	19.398	6.460
chg share bilingual	-0.697	-11.071*	7.223***	3.429	-24.054*	21.372**	-3.105	-35.230*	11.551**
(coef.)	(2.875)	(5.586)	(1.982)	(4.470)	(6.566)	(4.499)	(3.648)	(6.266)	(3.146)
est. effect at mean (% pts.)	0.081	1.286	-0.839	-0.398	2.795	-2.483	0.361	4.093	-1.342
est. effect at mean (std. dev.)	0.009	0.074	-0.248	-0.030	0.146	-0.273	0.033	0.211	-0.208
Obs.	1352	1352	1352	1352	1352	1352	1352	1352	1352
Grade 4									
mean in 1998	16.053	29.136	2.136	22.161	38.960	5.348	24.417	40.807	4.015
	8.290	16.487	4.482	10.934	17.485	7.963	9.879	17.506	6.155
chg share bilingual	6.256**	4.607	4.959**	14.549**	3.559	16.928**	0.995	-11.778*	5.325*
(coef.)	(2.918)	(5.597)	(2.273)	(4.005)	(6.357)	(3.877)	(3.407)	(5.949)	(3.221)
est. effect at mean (% pts.)	-0.757	-0.557	-0.600	-1.760	-0.431	-2.048	-0.120	1.425	-0.644
est. effect at mean (std. dev.)	-0.091	-0.034	-0.134	-0.161	-0.025	-0.257	-0.012	0.081	-0.105
Obs.	1266	1266	1266	1266	1266	1266	1266	1266	1266
Grade 5									
mean in 1998	14.835	26.039	1.470	21.477	35.227	4.142	21.915	38.907	4.015
	6.749	14.749	3.440	10.558	17.225	6.888	9.207	16.750	5.817
chg share bilingual	3.685	9.691	0.508	13.149**	6.752	12.450**	1.321	-11.616*	2.841
(coef.)	(2.795)	(6.525)	(1.883)	(4.448)	(7.260)	(3.568)	(3.789)	(6.882)	(3.183)
est. effect at mean (% pts.)	-0.459	-1.207	-0.063	-1.637	-0.841	-1.550	-0.164	1.446	-0.354
est. effect at mean (std. dev.)	-0.068	-0.082	-0.018	-0.155	-0.049	-0.225	-0.018	0.086	-0.061
Obs.	1159	1159	1159	1159	1159	1159	1159	1159	1159

All regressions instrument for the change in bilingual enrollments with the change implied by full compliance and control for changes in mean class size and free lunch eligibility. Robust standard errors in parentheses.
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 11
Effect of Bilingual Education on non-LEP Students throughout the Test Score Distribution: Three-Year Changes

	Reading			Math			Language		
	mean NPR score	% scoring above 25th %ile NPR	% scoring above 75th %ile NPR	mean NPR score	% scoring above 25th %ile NPR	% scoring above 75th %ile NPR	mean NPR score	% scoring above 25th %ile NPR	% scoring above 75th %ile NPR
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Grade 3									
mean in 1998	31.111	55.968	11.057	36.346	58.485	14.338	33.890	56.472	11.743
	14.330	18.896	10.215	16.423	18.785	12.872	14.867	18.778	10.955
chg share bilingual	-32.422*	-60.047*	6.171*	-20.036*	-43.552*	11.610**	-37.642*	-65.696*	-0.631
(coef.)	(3.446)	(6.006)	(3.195)	(4.487)	(5.757)	(4.740)	(4.122)	(6.174)	(3.742)
est. effect at mean (% pts.)	3.767	6.977	-0.717	2.328	5.060	-1.349	4.374	7.633	0.073
est. effect at mean (std. dev.)	0.263	0.369	-0.070	0.142	0.269	-0.105	0.294	0.406	0.007
Obs.	1337	1337	1337	1338	1338	1338	1337	1337	1337
Grade 4									
mean in 1998	33.161	58.251	13.597	33.844	56.551	14.639	39.138	62.170	14.738
	14.675	18.489	11.704	15.249	18.398	12.568	14.123	17.332	11.810
chg share bilingual	-25.803*	-41.711*	6.742**	-18.929*	-40.765*	3.883	-33.208*	-52.401*	-7.327**
(coef.)	(3.558)	(5.947)	(3.405)	(4.311)	(5.993)	(4.508)	(4.014)	(6.071)	(3.733)
est. effect at mean (% pts.)	3.122	5.047	-0.816	2.290	4.933	-0.470	4.018	6.341	0.887
est. effect at mean (std. dev.)	0.213	0.273	-0.070	0.150	0.268	-0.037	0.285	0.366	0.075
Obs.	1258	1258	1258	1259	1259	1259	1259	1259	1259
Grade 5									
mean in 1998	33.686	60.023	13.286	35.601	56.383	14.095	38.878	63.110	17.301
	13.410	17.046	10.617	14.936	18.003	11.947	14.412	16.480	12.113
chg share bilingual	-32.418*	-40.865*	2.112	-15.570*	-33.578*	8.804*	-34.938*	-37.548*	-12.358*
(coef.)	(4.537)	(7.075)	(3.824)	(5.157)	(7.333)	(4.787)	(5.121)	(6.630)	(4.595)
est. effect at mean (% pts.)	4.037	-5.089	-0.263	1.939	4.181	-1.096	4.351	4.676	1.539
est. effect at mean (std. dev.)	0.301	-0.299	-0.025	0.130	0.232	-0.092	0.302	0.284	0.127
Obs.	1153	1153	1153	1153	1153	1153	1152	1152	1152

All regressions instrument for the change in bilingual enrollments with the change implied by full compliance and control for changes in mean class size and free lunch eligibility. Robust standard errors in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%

Table 12

The Effect of the Reduction in Bilingual Education Experienced by the Average California School

	Third Graders		**Fourth Graders**		**Fifth Graders**		**SixthGraders**		**Seventh Graders**		**Eighth Graders**	
	LEP	**Non-LEP**	**LEP**	**Non-LEP**	**LEP**	**Non-LEP**	**LEP**	**Non-LEP**	**LEP**	**Non-LEP**	**LEP**	**Non-LEP**
READING												
effect of the mean change in share bilingual, measured in percentile rank points	-1.889***	1.803***	-1.144***	2.715***	-0.331	4.064***	-0.653	1.975**	-0.359	1.288***	-0.279	0.529
effect of the mean change in share bilingual, measured in standard deviations	-0.220***	0.126***	-0.139***	0.185***	-0.049	0.303***	-0.047	0.144**	-0.023	0.081***	-0.019	0.036
MATH												
effect of the mean change in share bilingual, measured in percentile rank points	-2.215***	0.475***	-3.124***	0.936*	-2.169***	1.316**	-1.820**	1.278	0.136	0.834**	-0.436	-0.833
effect of the mean change in share bilingual, measured in standard deviations	-0.168***	0.029***	-0.287***	0.061*	-0.237***	0.088**	-0.118**	0.083	0.008	0.051**	-0.027	-0.052
LANGUAGE												
effect of the mean change in share bilingual, measured in percentile rank points	-1.771***	2.274***	-1.054**	3.053***	0.207	4.638***	0.005	2.894**	-1.049**	0.654	-0.305	0.242
effect of the mean change in share bilingual, measured in standard deviations	-0.163***	0.153***	-0.107**	0.216***	0.020	0.322***	0.001	0.208*	-0.064**	0.040	-0.020	0.016

Notes to Table 12: The table shows instrumental variables (IV) regression estimates for all California schools with at least 10 LEP students in the relevant grade. The estimates are from the regressions shown in Tables 8 and 9.

Table A-1

First stage results

Dependent variable: actual change in percent bilingual education enrollment						
Population:	LEP	LEP	non-LEP	LEP	LEP	non-LEP
Time span:	1-year	1-year	1-year	3-year	3-year	3-year
change % bilingual eligible	0.311	0.313	0.309	0.310	0.313	0.315
	(0.016)	(0.016)	(0.016)	(0.015)	(0.016)	(0.016)
predicted change % non-LEP students in classes experienced		0.010	0.010		-0.022	-0.021
		(0.003)	(0.003)		(0.018)	(0.018)
change mean class size in grade	-0.001	-0.001	-0.001	0.0002	0.0003	0.0001
	(0.001)	(0.001)	(0.001)	(0.0010)	(0.0010)	(0.0010)
change share free/red lunch eligible in school	-0.028	-0.013	-0.038	-0.021	-0.021	-0.016
	(0.050)	(0.047)	(0.045)	(0.042)	(0.042)	(0.042)
F	122.72	101.05	103.20	136.29	102.64	105.09
N (schools)	1193	1193	1193	1241	1241	1241

Table A-2

First stage results

Dependent variable: actual change in % non-LEP students in classes experienced				
Population:	LEP	non-LEP	LEP	non-LEP
Time span:	1-year	1-year	3-year	3-year
change % bilingual eligible	0.029	0.017	-0.064	-0.064
	(0.016)	(0.016)	(0.015)	(0.015)
predicted change % non-LEP students in classes experienced	0.007	0.007	0.262	0.260
	(0.003)	(0.003)	(0.017)	(0.017)
change mean class size in grade	0.001	0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)
change share free/red lunch eligible in school	-0.012	0.010	-0.154	-0.155
	(0.046)	(0.046)	(0.041)	(0.042)
F	2.740	2.190	60.610	59.260
N (schools)	1228	1228	1242	1242

Table A-3						
Three-Year OLS Estimates without Purging Habituation Effects						
	Third Graders		**Fourth Graders**		**Fifth Graders**	
	LEP	Non-LEP	LEP	Non-LEP	LEP	Non-LEP
Regression type	OLS	OLS	OLS	OLS	OLS	OLS
Habit. effect purged	no	no	no	no	no	no
READING (dependent variable is median percentile rank in reading)						
change, share bilingual	-5.765***	-16.576***	-0.095	-10.703***	1.078	-9.386***
	(1.233)	(1.762)	(1.265)	(1.737)	(1.226)	(1.709)
change, mean class size in grade	-0.075*	-0.057	-0.052	-0.044	-0.084	-0.044
	(0.041)	(0.047)	(0.057)	(0.067)	(0.057)	(0.074)
change, share free/red lunch elig	-1.75	-6.306**	-2.368	-7.465***	-3.619**	-5.858**
	(2.298)	(2.657)	(2.727)	(2.757)	(1.796)	(2.699)
constant	7.494***	12.454***	6.718***	11.127***	4.479***	8.150***
	(0.337)	(0.357)	(0.283)	(0.331)	(0.274)	(0.341)
effect for mean change in share bilingual (-0.12)	0.675	1.939	0.012	1.306	-0.135	1.173
MATH (dependent variable is median percentile rank in reading)						
change, share bilingual	-2.464	-14.730***	2.49	-8.729***	5.037**	-4.806**
	(2.123)	(2.383)	(1.892)	(2.190)	(2.167)	(2.203)
change, mean class size in grade	-0.374***	-0.335***	-0.199***	-0.152*	-0.275***	-0.136
	(0.064)	(0.064)	(0.075)	(0.082)	(0.097)	(0.097)
change, share free/red lunch elig	-1.767	-7.582**	-6.552*	-9.996***	-7.901**	-9.497***
	(3.339)	(3.728)	(3.689)	(3.674)	(3.084)	(3.293)
constant	15.286***	18.316***	12.637***	16.403***	11.632***	15.281***
	(0.500)	(0.489)	(0.387)	(0.414)	(0.432)	(0.432)
effect for mean change in share bilingual (-0.12)	0.288	1.723	-0.304	1.065	-0.630	0.601
LANGUAGE (dependent variable is median percentile rank in reading)						
change, share bilingual	-7.225***	-19.397***	-2.147	-14.110***	0.787	-11.920***
	(1.647)	(2.052)	(1.468)	(1.986)	(1.699)	(2.171)
change, mean class size in grade	-0.154***	-0.149***	-0.074	-0.08	-0.109	-0.024
	(0.051)	(0.053)	(0.066)	(0.074)	(0.081)	(0.081)
change, share free/red lunch elig	-3.001	-4.043	-5.232	-6.246**	-5.455**	-4.02
	(2.872)	(3.297)	(3.265)	(3.153)	(2.642)	(3.284)
constant	10.817***	14.627***	8.131***	10.264***	6.758***	9.382***
	(0.411)	(0.408)	(0.324)	(0.352)	(0.369)	(0.392)
effect for mean change in share bilingual (-0.12)	0.845	2.269	0.262	1.721	-0.098	1.490
Observations	1336	1321	1250	1244	1140	1134
Robust standard errors in parentheses.						
* significant at 10%; ** significant at 5%; *** significant at 1%.						

Table A-4

Three-Year IV Estimates without Purging Habituation Effects

	Third Graders		Fourth Graders		Fifth Graders	
	LEP	Non-LEP	LEP	Non-LEP	LEP	Non-LEP
Regression type	IV	IV	IV	IV	IV	IV
Habit. effect purged	no	no	no	no	no	no
READING (dependent variable is median percentile rank in reading)						
change, share bilingual	-0.637	-32.207***	5.706*	-26.005***	3.231	-32.222***
	(2.888)	(3.404)	(2.932)	(3.569)	(2.782)	(4.501)
change, mean class size in grade	-0.086**	-0.029	-0.055	-0.051	-0.086	-0.023
	(0.041)	(0.048)	(0.057)	(0.070)	(0.057)	(0.079)
change, share free/red lunch elig	-1.828	-6.237**	-2.572	-6.919**	-3.719**	-4.797*
	(2.321)	(2.665)	(2.710)	(2.835)	(1.807)	(2.730)
constant	8.048***	10.783***	7.422***	9.266***	4.749***	5.298***
	(0.487)	(0.439)	(0.477)	(0.457)	(0.459)	(0.553)
effect for mean change in share bilingual (-0.12)	0.075	3.768	-0.696	3.173	-0.404	4.028
MATH (dependent variable is median percentile rank in reading)						
change, share bilingual	2.848	-20.189***	14.161***	-19.128***	12.912***	-15.223***
	(4.461)	(4.477)	(4.022)	(4.323)	(4.429)	(5.111)
change, mean class size in grade	-0.385***	-0.321***	-0.197**	-0.153*	-0.282***	-0.126
	(0.064)	(0.064)	(0.077)	(0.083)	(0.097)	(0.097)
change, share free/red lunch elig	-1.845	-7.518**	-6.931*	-9.607***	-8.266***	-9.013***
	(3.376)	(3.699)	(3.671)	(3.624)	(3.126)	(3.304)
constant	15.860***	17.733***	14.037***	15.127***	12.621***	13.980***
	(0.703)	(0.618)	(0.615)	(0.585)	(0.706)	(0.681)
effect for mean change in share bilingual (-0.12)	-0.333	2.362	-1.728	2.334	-1.614	1.903
LANGUAGE (dependent variable is median percentile rank in reading)						
change, share bilingual	-3.246	-37.909***	0.354	-33.389***	0.678	-34.945***
	(3.660)	(4.116)	(3.426)	(4.030)	(3.762)	(5.090)
change, mean class size in grade	-0.163***	-0.113**	-0.079	-0.085	-0.109	-0.003
	(0.051)	(0.056)	(0.066)	(0.079)	(0.081)	(0.085)
change, share free/red lunch elig	-3.064	-3.947	-5.34	-5.533*	-5.450**	-2.949
	(2.889)	(3.355)	(3.265)	(3.215)	(2.646)	(3.384)
constant	11.247***	12.647***	8.432***	7.918***	6.744***	6.506***
	(0.591)	(0.522)	(0.544)	(0.495)	(0.610)	(0.626)
effect for mean change in share bilingual (-0.12)	0.380	4.435	-0.043	4.073	-0.085	4.368
Observations	1335	1320	1248	1242	1140	1134

Robust standard errors in parentheses.

* significant at 10%; ** significant at 5%; *** significant at 1%.

Table A-5						
Effect of Bilingual Education on Third, Fourth, and Fifth Graders: Three-Year Changes Weighted by Enrollment						
	Third Graders		**Fourth Graders**		**Fifth Graders**	
	LEP	Non-LEP	LEP	Non-LEP	LEP	Non-LEP
Regression type	IV	IV	IV	IV	IV	IV
Habit. effect purged	yes	yes	yes	yes	yes	yes
READING (dependent variable is median percentile rank in reading)						
change, share bilingual	16.560***	-11.858***	13.212***	-19.100***	4.866*	-30.632***
	(3.018)	(3.191)	(2.855)	(3.479)	(2.588)	(4.508)
change, mean class size in grade	-0.056	0.004	-0.043	-0.044	-0.017	0.055
	(0.045)	(0.050)	(0.061)	(0.072)	(0.057)	(0.116)
change, share free/red lunch elig	-3.444	-9.158***	-4.286	-9.788***	-5.805***	-8.466***
	(2.767)	(2.988)	(2.729)	(2.999)	(1.915)	(3.173)
constant	4.936***	8.038***	3.217***	5.062***	2.721***	3.218***
	(0.521)	(0.464)	(0.477)	(0.476)	(0.459)	(0.568)
effect for mean change in share bilingual (-0.12)	-1.938	1.387	-1.612	2.330	-0.608	3.829
MATH (dependent variable is median percentile rank in reading)						
change, share bilingual	21.208***	0.036	29.492***	-5.888	20.390***	-8.853*
	(4.510)	(4.240)	(4.064)	(3.798)	(4.090)	(4.991)
change, mean class size in grade	-0.358***	-0.295***	-0.225**	-0.166**	-0.240**	-0.066
	(0.067)	(0.068)	(0.088)	(0.084)	(0.099)	(0.146)
change, share free/red lunch elig	-3.409	-10.131**	-7.307**	-9.862***	-8.721***	-11.819***
	(3.882)	(4.213)	(3.617)	(3.685)	(3.253)	(3.563)
constant	4.968***	7.214***	5.761***	6.745***	4.371***	5.593***
	(0.747)	(0.637)	(0.628)	(0.571)	(0.701)	(0.670)
effect for mean change in share bilingual (-0.12)	-2.481	-0.004	-3.598	0.718	-2.549	1.107
LANGUAGE (dependent variable is median percentile rank in reading)						
change, share bilingual	16.075***	-16.382***	11.805***	-23.854***	1.22	-37.840***
	(3.816)	(3.784)	(3.261)	(3.718)	(3.355)	(5.204)
change, mean class size in grade	-0.133**	-0.084	-0.05	-0.062	-0.037	0.111
	(0.055)	(0.057)	(0.069)	(0.078)	(0.083)	(0.142)
change, share free/red lunch elig	-5.803*	-7.765**	-6.846**	-7.731**	-7.272***	-7.143*
	(3.415)	(3.503)	(3.036)	(3.344)	(2.605)	(4.005)
constant	3.028***	4.741***	3.357***	2.736***	1.166*	0.615
	(0.624)	(0.537)	(0.541)	(0.498)	(0.599)	(0.639)
effect for mean change in share bilingual (-0.12)	-1.881	1.917	-1.440	2.910	-0.153	4.730
Observations	1335	1320	1248	1242	1140	1134
Robust standard errors in parentheses						
* significant at 10%; ** significant at 5%; *** significant at 1%						

Table A-6						
Effect of Bilingual Education on Third, Fourth, and Fifth Graders: One-Year Changes						
	Third Graders		**Fourth Graders**		**Fifth Graders**	
	LEP	Non-LEP	LEP	Non-LEP	LEP	Non-LEP
Regression type	IV	IV	IV	IV	IV	IV
Habit. effect purged	yes	yes	yes	yes	yes	yes
READING (dependent variable is median percentile rank in reading)						
change, share bilingual	13.806***	-9.201***	3.951	-17.240***	2.358	-24.056***
	(2.651)	(3.132)	(2.781)	(3.511)	(2.797)	(3.845)
change, mean class size in	-0.084**	-0.132***	0.037	-0.091	-0.061	-0.057
grade	(0.042)	(0.045)	(0.058)	(0.071)	(0.046)	(0.066)
change, share free/red lunch	-1.117	-5.171**	-3.236	-4.717	-0.933	-1.305
elig	(2.045)	(2.545)	(2.575)	(3.157)	(1.826)	(3.018)
constant	2.025***	2.229***	1.493***	1.743***	-0.102	0.163
	(0.433)	(0.400)	(0.468)	(0.421)	(0.481)	(0.470)
effect for mean change in						
share bilingual (-0.12)	-1.588	1.058	-0.474	2.069	-0.288	2.935
MATH (dependent variable is median percentile rank in reading)						
change, share bilingual	13.784***	-2.405	8.311**	-12.685***	-1.653	-18.342***
	(3.880)	(3.819)	(3.349)	(4.068)	(3.713)	(4.077)
change, mean class size in	-0.243***	-0.258***	0.029	-0.045	-0.08	-0.086
grade	(0.062)	(0.061)	(0.077)	(0.089)	(0.067)	(0.072)
change, share free/red lunch	-1.757	-5.026	-6.895**	-7.999**	-3.975	-3.74
elig	(2.916)	(3.283)	(3.179)	(3.393)	(2.679)	(3.297)
constant	1.356**	1.285**	2.551***	1.712***	-1.687***	-1.102**
	(0.611)	(0.501)	(0.554)	(0.506)	(0.631)	(0.528)
effect for mean change in						
share bilingual (-0.12)	-1.585	0.277	-0.997	1.522	0.202	2.238
LANGUAGE (dependent variable is median percentile rank in reading)						
change, share bilingual	9.705***	-13.339***	10.003***	-11.600***	-3.79	-25.653***
	(3.224)	(3.504)	(2.997)	(3.690)	(3.772)	(4.616)
change, mean class size in	-0.127**	-0.192***	-0.011	-0.094	-0.031	-0.06
grade	(0.050)	(0.053)	(0.072)	(0.077)	(0.074)	(0.077)
change, share free/red lunch	-2.449	-3.351	-4.897*	-8.413**	-1.083	-1.2
elig	(2.427)	(2.758)	(2.541)	(3.390)	(2.150)	(3.244)
constant	1.368**	1.199***	1.486***	0.066	-1.442**	-1.257**
	(0.531)	(0.430)	(0.513)	(0.432)	(0.637)	(0.552)
effect for mean change in						
share bilingual (-0.12)	-1.116	1.534	-1.200	1.392	0.462	3.130
Observations	1294	1308	1204	1229	1079	1125
Robust standard errors in parentheses						
* significant at 10%; ** significant at 5%; *** significant at 1%						

Table A-7

Effect of Bilingual Education on Third, Fourth, and Fifth Graders: Two-Year Changes

	Third Graders		Fourth Graders		Fifth Graders	
	LEP	Non-LEP	LEP	Non-LEP	LEP	Non-LEP
Regression type	IV	IV	IV	IV	IV	IV
Habit. effect purged	yes	yes	yes	yes	yes	yes
READING (dependent variable is median percentile rank in reading)						
change, share bilingual	9.102***	-16.114***	11.413***	-18.263***	4.731	-30.648***
	(3.014)	(3.333)	(3.237)	(3.950)	(2.965)	(4.702)
change, mean class size in grade	-0.109***	-0.037	-0.064	-0.062	-0.057	-0.116
	(0.042)	(0.045)	(0.049)	(0.073)	(0.058)	(0.076)
change, share free/red lunch elig	0.228	-3.865*	0.321	-5.099**	-3.078*	-2.182
	(1.812)	(2.133)	(2.166)	(2.264)	(1.725)	(2.628)
constant	3.015***	4.941***	2.379***	3.415***	1.562***	1.246**
	(0.483)	(0.434)	(0.497)	(0.469)	(0.470)	(0.562)
effect for mean change in share bilingual (-0.12)	-1.029	1.821	-1.347	2.155	-0.572	3.708
MATH (dependent variable is median percentile rank in reading)						
change, share bilingual	11.579**	-6.589	24.912***	0.437	13.137***	-10.944**
	(4.686)	(4.756)	(4.540)	(4.286)	(4.442)	(4.859)
change, mean class size in grade	-0.309***	-0.245***	-0.068	-0.029	-0.084	-0.148*
	(0.067)	(0.063)	(0.072)	(0.078)	(0.085)	(0.088)
change, share free/red lunch elig	-2.832	-1.942	0.343	0.026	-1.101	-0.317
	(3.120)	(3.169)	(3.470)	(3.016)	(3.038)	(3.033)
constant	4.683***	6.000***	5.554***	6.276***	1.599**	2.231***
	(0.716)	(0.591)	(0.666)	(0.548)	(0.696)	(0.626)
effect for mean change in share bilingual (-0.12)	-1.308	0.745	-2.940	-0.052	-1.590	1.324
LANGUAGE (dependent variable is median percentile rank in reading)						
change, share bilingual	8.192**	-15.644***	13.819***	-14.207***	5.116	-33.894***
	(3.831)	(3.845)	(3.762)	(4.002)	(3.886)	(5.365)
change, mean class size in grade	-0.205***	-0.160***	-0.035	-0.029	0.002	-0.102
	(0.052)	(0.053)	(0.059)	(0.079)	(0.073)	(0.075)
change, share free/red lunch elig	-0.578	-2.05	-1.873	-2.426	-2.471	-1.682
	(2.451)	(2.295)	(2.499)	(2.355)	(2.322)	(2.848)
constant	2.833***	4.371***	2.523***	1.665***	1.384**	-0.303
	(0.609)	(0.497)	(0.570)	(0.472)	(0.610)	(0.633)
effect for mean change in share bilingual (-0.12)	-0.926	1.768	-1.631	1.676	-0.619	4.101
Observations	1337	1352	1243	1262	1118	1151
Robust standard errors in parentheses						
* significant at 10%; ** significant at 5%; *** significant at 1%						