Sinking or Swimming?

Evaluating the Impact of English Immersion
versus Bilingual Education on Student Achievement

JORDAN D. MATSUDAIRA[1]

Department of Economics and
Gerald R. Ford School of Public Policy
University of Michigan
jmatsuda@umich.edu

November 13, 2004

Earlier Drafts Available on Request

[1] I thank John DiNardo, Rebecca Blank, Sheldon Danziger, Julie Cullen, Mary Corcoran and Justin McCrary for the continuity and richness of their support throughout my graduate studies at Michigan. I also thank John Bound, Gary Solon, Charlie Brown and other participants at the University of Michigan labor seminar for helpful comments on this paper. I also gratefully acknowledge the National Poverty Center at the University of Michigan and the Spencer Foundation for financial support.

Abstract

In recent years, the role of bilingual education as the dominant pedagogy for teaching immigrant students has been challenged by federal and state policy reforms. Critics of bilingual education contend that bilingual programs hinder achievement by reducing incentives to learn English and by trapping students in classrooms with low-performing peers. Proponents argue that learning occurs most rapidly when students master concepts in their native language first, and therefore that bilingual education programs promote achievement. To distinguish between these competing views, I exploit quasi-random assignment of students to bilingual and mainstream (English-immersion) classes generated by discontinuous program eligibility rules in a large urban school district. In the District, eligibility for bilingual programs is determined by a test of English proficiency: students scoring below a preset threshold level are eligible for bilingual classes and students scoring above this threshold are not eligible. Using information on achievement and program participation from a large administrative data set, I compare students scoring just below and just above this threshold for several years following the proficiency test. Compared to students scoring just above the threshold, students scoring just below it are nearly 90 percent more likely to participate in a bilingual program and are surrounded by peers with significantly lower average achievement scores. Despite these striking differences in classroom environments, however, I find negligible differences in achievement in both reading and math (measured by tests given in English) across these two groups of students. These results speak directly to current policy debates over the most effective methods to educate immigrant children. In addition, the results inform a broader debate regarding the importance of educational peer effects, as bilingual programs involve learning in classes surrounded by a group of peers who have markedly lower average achievement scores and are more similar in ethnic background compared to mainstream English-immersion classes.

1 Introduction

This paper examines whether school-aged immigrant children learn more when they are placed in classes where instruction occurs partially in their native language surrounded by other English learners than when they are placed in mainstream classes conducted only in English. In light of secular declines in fertility and changes in the age structure of the population, immigration figures to be an increasingly important source of labor power and hence economic growth in the United States. Indeed, immigration increased dramatically in the postwar period and continues to rise: more immigrants arrived in the U.S. in the last decade than in any other decade in the nation's history (Migration Policy Institute, 2004). As the number of immigrants has risen, however, so too has concern over whether these newcomers posess the skills necessary to participate effectively in the economy. In a number of articles, George Borjas has pointed out that newly arrived immigrants have fared increasingly poorly in terms of their wages and poverty rates relative to natives.[1] Immigrants with limited English proficiency—and recent immigrants are increasingly likely not to speak English upon arrival—struggle disproportionately. In Los Angeles and New York City, for example, more than one in three immigrants with limited English proficiency have incomes below the poverty line (Capps et al. 2002).[2]

These considerations have animated debate about how best to educate immigrant school children with limited English proficiency (LEP). In particular, the pedagogical practice of bilingual education, where students are taught partially in their native language and often in classrooms apart from non-LEP students, has become a lightning rod for controversy. Persistent socio-economic disparities between linguistic minorities—primarily Hispanics—and other ethnic groups have led many commentators to blame these programs that were just fifteen years earlier thought to be the remedy for such ills. Introduced in the wake of the Civil Rights Act, bilingual education was once touted as a way to ensure educational opportunity for linguistic minorities. An increasingly

[1]See, for example, Borjas (1995).

[2]This cross-sectional correlation holds up when other observable differences are accounted for: Chiswick (1991) and Kossoudji (1988) both find that English deficiencies are associated with reduced earnings even after controlling for schooling and occupational differences. Further, Trejo (1997) estimates that differences in English proficiency may account for nearly one-quarter of the wage-gap between Mexican-Americans and non-Hispanic white male workers. According to Trejo, among third- and higher-generation Mexican men, this gap in average wages was 21 percent in 1989.

vocal opposition, however, has insisted instead that bilingual education programs trap immigrant children in substandard learning environments, foreclosing their chances to economically assimilate in U.S. society. Giving early voice to such concerns in 1981, then President Reagan stated: "It is absolutely wrong and against American concepts to have a bilingual education program that is now openly, admittedly dedicated to preserving their native language and never getting them adequate in English so they can go out into the job market and participate."[3]

Today, the notion that immigrant children would fare better if placed in mainstream classes taught only in English is gathering momentum. Voters in several states, including California, Arizona, and Massachusetts have passed laws mandating that the LEP students living in those states be taught primarily in English. Federal policies, once strongly supporting bilingualism, have similarly begun to emphasize rapid English language acquisition and English-only immersion programs. Despite the hundreds of studies claiming to evaluate the relative merits of bilingual versus English-only programs, however, these policy changes have been enacted largely in the dark. Students enrolled in bilingual classes differ from students enrolled in English-only classes in fundamental ways that affect their relative performance on achievement tests. English language ability is only the most obvious difference. Bilingual students also tend to come from poorer families, have parents with lower educational attainment, and have lived in the country for a shorter period of time. Since nearly all studies of bilingual education and achievement fail to adequately control for the negative selection of students into bilingual programs, their results are likely biased against finding that the programs are effective.

To obtain an unbiased estimate of the effectiveness of bilingual education relative to mainstream English-only classes, I exploit quasi-random assignment of students to classes generated by discontinuous program eligibility rules in a Large Urban School District in the Northeast (hereafter LUSDiNE[4]). In the District, eligibility for bilingual programs is determined by a test of English proficiency: students scoring below a preset threshold level are classified as Limited English Proficient and are thus eligible to participate in bilingual classes, while students scoring above this

[3] Cited in Crawford (1999).

[4] The name of the school district is kept anonymous per an agreement with the district's administration allowing use of the data.

threshold are not eligible. If there is some randomness in student test scores, then students scoring just above and just below the threshold on the proficiency test are likely to be similar in terms of other determinants of achievement. Thus, analogous to a randomized controlled trial, I assess the effects of the programs on math and reading achievement by comparing the subsequent test scores of students scoring just below and just above this threshold for several years following the proficiency test.[5]

To preview the results, I find very little difference in achievement among students based on their eligibility for bilingual programs. Nearly 90 percent of students classified as LEP enroll in bilingual education programs in the year following assessment, regardless of how close they were to the language proficiency test cutoff score, while close to zero percent of non-LEP students do so. Moreover, students scoring just below the cutoff score on the language proficiency test tend to remain in these programs for several years: nearly 30 percent remain in bilingual programs two years afterwards, and about fifteen percent remain three and even four years later. On the other hand, the achievement effects of this accumulated experience appear quite muted. There are very small differences in reading or math achievement (measured on tests given in English) by program status for students measured in any of the four years following assignment to ELS or mainstream classes. While there is some heterogeneity in the measured effects across different grade levels and years, no systematic patterns are evident. Overall, the average of different estimates of the achievement impacts of bilingual education relative to English immersion for the subset of students near the threshold level of proficiency is just above zero.

While the results of this paper pertain directly to the current policy debate on the relative effectiveness of bilingual education programs, they also inform a broader literature on endogenous social effects and, in particular, educational peer effects. Differences in the extent to which students' native languages are used in instruction delineate bilingual programs from English-only programs, but pedagogical and curricular differences are not the only differences between programs. Students in bilingual programs also tend to be taught in classrooms with other English learning immigrants from similar linguistic backgrounds. Relative to students in mainstream English-only classes, then,

[5]This is the regression discontinuity research design introduced by Thistlewaite and Campbell (1960).

students in bilingual classes are surrounded by peers who have lower English ability, have lower achievement test scores in both math and reading, and are from similar ethnic backgrounds. Recent studies by Hoxby (2000) and Hanushek et al. (2001) have found that such differences in the characteristics of peer groups significantly affect student achievement.[6] The evidence presented in this paper suggests to the contrary that—at least in the context of immigrant school children and bilingual education—differences in peer quality and racial composition do not noticeably affect achievement outcomes.[7]

The remainder of this paper is organized as follows: Section 2 provides a brief statistical overview of LEP children in the U.S., as well as a summary of trends in policies towards the education of immigrant children, and previous research on the the effectiveness of different models. Section 3 provides an overview of the bilingual program eligibility policy in LUSDiNE and how its features are exploited to create the regression discontinuity design used in the paper, as well as details about the data, estimation, and inference procedures. Section 4 through section 7 present the main results of the paper. Finally, section 8 concludes with a discussion of the results and their implications for policy in this area.

2 Background

There were approximately 3.7 million school-aged LEP children in U.S. schools in the school year ending in Spring 2000 (Kindler 2002). While this accounted for about 8 percent of total public school enrollment, LEP students are heavily concentrated in a handful of states. California has the lion's share with nearly 1.5 million, followed by Texas, Florida, New York, and Illinois (Kindler 2002). And within these states, students are concentrated further in the large urban areas that tend to draw the highest fractions of immigrants. Thus, Los Angeles' and New York City's school districts have by far the largest concentrations of LEP students: in 2001, Los Angeles had 311,958

[6] On the other hand, Angrist and Lang (2002) and Lefgren (2004) find little to no evidence for peer effects on achievement among school-aged children and Sacerdote (2000) finds small effects in a study of randomly matched college students.

[7] Another possibility, discussed below, is that the (presumably negative) effects of peer quality for students in bilingual programs are just offset by positive effects due to other programmatic differences such as language of instruction or teacher quality.

accounting for 43 percent of total enrollment, while New York had 180,440 students accounting for 17 percent of total enrollment; Chicago was a distant third with 57,767 (13 percent) (Antunez 2003). While the majority—about three-fourths (Kindler 2002)—of LEP students come from Spanish speaking backgrounds, this fraction has been declining with increasing immigration from other regions, particularly Asia.

School districts have implemented a variety of different programs to cater to the educational needs of LEP students. There are two main classes of program: 'bilingual education' and 'English-immersion'. Broadly speaking 'bilingual education' programs use the student's native language for a non-negligible fraction of instruction time. These programs generally teach content in math, reading, etc. in a student's native language and generally dedicate time to improving literacy in that language as well. Moreover, students are generally taught in sheltered classrooms only with other LEP students from similar linguistic backgrounds. In contrast, 'English immersion' programs use little to none of the student's native language. 'English as a Second Language (ESL)' classes focus on teaching the English language to LEP students, and are often partly included in both 'bilingual' and 'immersion' programs. In LUSDiNE, ESL students are generally 'pulled-out' of their regular classes for 180 minutes of English instruction per week. For the purposes of this paper, I refer to all programs in which some of a student's native language is used—that is, either bilingual or ESL programs—using the catch-all term 'English Language Services (ELS)'.[8]

Policy Background

Bilingual education has a long history in the United States, as does the opposition to the use of non-English languages in schools. In the late 1800s, bilingual education was relatively common as waves of European immigrants formed their own schools: according to one estimate, in 1900 approximately 600,000 school children (4 percent of the elementary school population) received instruction in German (Kloss 1998). With increased immigration in the early 1900s and

[8]LUSDiNE officials do not keep data on the nature of the curriculum of ELS classes for individual students. It is possible that such classes are similar to those that would be received by students enrolled in programs referred to as 'structured immersion'. It should also be noted that in this paper, ELS will be contrasted to a state where students are offered no services. This is referred to as immersion in this paper, but it should be noted that, nationwide, many programs referred to as immersion provide ESL classes or other services to students. Thus, immersion in the context of this paper might better correspond to what has been referred to as 'submersion': putting LEP children in mainstream classes without additional support.

especially with the beginning of World War I, however, popular opinion mobilized against allowing foreign languages to be spoken in schools. By the mid-1920s, 34 states had mandated English-only instruction in all elementary schools (Kloss 1998).

After a long period of relative dormancy, significant efforts to reintroduce bilingual education programs in schools appeared in the 1960s as part of the Civil Rights movement. The birth of modern bilingual education programs backed by the federal government came with the passage of the Bilingual Education Act, or Title VII of the Elementary and Secondary Education Act, passed in 1968.[9] This Act provided direct funding to local education agencies with the goal of promoting and developing innovative approaches to the education of LEP students. A significant aspect of the Act was that it recognized promotion of literacy in a student's native language as a valid goal of funded programs, and indeed bilingual programs received the large majority of funding since (Crawford 2003).

Bilingual education received further support with the 1974 Supreme Court ruling in *Lau v. Nichols* in 1974, and related regulations promulgated in 1975 by the Office for Civil Rights known as the *Lau* Remedies. In *Lau*, a group of 1,800 Chinese immigrant students sued local educational authorities claiming they were discriminated against since they could not understand the English-based instruction in their schools. Relying on the 1964 Civil Rights Act, the Court ruled that "There is no equality of treatment merely by providing students with the same facilities, textbooks, teachers, and curriculum; for students who do not understand English are effectively foreclosed from any meaningful education."[10] While the Court did not specify specific curricular remedies, it required educational authorities to take affirmative steps to ensure that LEP students had equal opportunities to learn. The *Lau* Remedies strengthened the Court ruling by requiring all school districts with at least 20 LEP students in a given language group to establish bilingual programs, specifying acceptable pedagogical strategies, and requiring districts to provide evidence they had programs to meet the needs of LEP students. Non-compliance could be punished by loss of federal funds (Ovando 2003).

While the trend over the 1960s and 70s of increased support for teaching LEP students in their

[9]This discussion is based on Ovando (2003).

[10]From the *Lau* opinion written by Justice Douglas, cited in Ovando (2003).

native language was never without opposition, the tide began to shift against bilingual education in the early 1980s. The Reagan administration, as evident from the quote in the introduction, was not a supporter of bilingual education and began efforts to redirect government funds to English only programs for immigrant education. Broadly speaking, federal support for bilingual education programs has been in decline ever since. While Title VII funding had been largely earmarked for bilingual programs only, a series of modifications to the law have allowed English-only instructional programs to receive an increasing share of appropriations. Congressional opponents of native-language instruction have also advocated reducing the role of bilingual programs by imposing a time-limit on the length of time that LEP students may receive English language services (Ovando 2003).

The No Child Left Behind Act (NCLB), passed in 2002, further shifts federal support from bilingual education to English immersion programs. While overall funding for LEP programs was increased by the Act, funds are now given as block grants to states rather than awarded competitively to local educational providers. This effectively eliminates federal control over pedagogy, and the preference for native language programs that existed under the Bilingual Education Act. Indeed, under NCLB that title is replaced with the "English Language Acquisition, Language Enhancement, and Academic Achievement Act (Title III of NCLB)," and promoting literacy in a student's native language is no longer listed as an approved purpose of programs funded under the Act. While language in an earlier version of the Act imposed a two-year time-limit on the receipt of ELS, this was ultimately not passed. The law did specify, however, that districts must show improvement among LEP students in 1) the fraction who are reclassified as fluent each year, 2) average test scores on English-language achievement tests (e.g., in math and reading). The law also mandates that all LEP students who have been enrolled in U.S. schools and in the country for longer than three years will be required to have their achievement tested in English (Crawford 2003). Taken together, these changes create incentives for schools to reclassify and move LEP students to mainstream classes as quickly as possible.

In addition to the changes in federal policy described above, several states have enacted policies that dramatically curtail the use of bilingual education in schools. Most notably, in 1998 California passed a referendum (Proposition 227) to make English the primary mode of instruction for LEP students. Under the new law in California, LEP students are placed in English-only immersion classes unless their parents petition to allow them to participate in a bilingual program. Similar initiatives have been adopted by other states such as Arizona in 2000 and Massachusetts in 2002.

Previous Research

Language in the NCLB Act requires that the design of all programs funded under the new law be based on scientific research. Unfortunately, there are very few studies on the effectiveness of different pedagogical strategies that satisfy this criterion. There have been three national evaluations of programs for LEP students, two of which were commissioned by the U.S. Department of Education: Danoff (1978), Burkheimer (1989), and Ramírez et al. (1991). All of these studies suffer from serious flaws in their research designs. A National Academy of Science Review of the latter two evaluations (Meyer and Fienberg 1992) concludes that there is essentially no useful information from these studies to discern which type of instructional program contributes most to student achievement.

Hundreds of smaller-scale studies have also been conducted, and several researchers have used meta-analytic techniques to produce average estimates of program effectiveness across subsets of these studies. Baker and de Kanter (1981) was the first such review, and identified about 150 studies claiming to evaluate the effectiveness of programs using native language instruction compared to English-immersion. After imposing the criteria that a study must either have random assignment across programs or include statistical controls to account for differences across students, only 28 of these studies remained in their analysis. Reviewing these studies, the authors concluded that there was no sufficient research basis to support exclusive reliance on bilingual education.

Since Baker and de Kanter, a string of similar papers have appeared. Willig (1985) reviewed the same list of studies, but eliminated several of the 28 that had made the cut and estimated average effect sizes using meta-analytic techniques. She concluded that in fact, there was evidence supporting bilingual approaches over immersion. More recently, reviews have been authored by Rossell and Baker (1996) and Greene (1998), finding, respectively, a lack of support and support for the effectiveness of bilingual education. While the relative merits of all of these reviews are arguable, a more striking feature is that the list of studies in each review has changed by fewer than 5 over the last 25 years. Indeed, 4 of the 5 'highest quality' studies using random assignment that are identified by Greene are more than 30 years old and none are fewer than 15 years old. Only one of these studies was subject to peer review. More than contributing to the current debate over the relative effectiveness of immersion and bilingual strategies, a review of the past research begs for new studies to be conducted that effectively deal with selection biases.

3 Assignment to English Language Services in LUSDiNE and Empirical Strategy

Sample Overview and Institutional Background

The main challenge in estimating the causal effect of bilingual education on student achievement is made clear by Table 1. The Table depicts the means of a range of characteristics for certain subgroups of students in LUSDiNE whose English language proficiency was formally assessed by the local education authority, and are thus 'at risk' for participating in a bilingual education program. As shown in the first three rows of the Table, students participating in ELS significantly underperform those students not participating in ELS. The test scores in math and reading are between .54 and .64 of a standard deviation[11] lower for students enrolled in ELS in the previous year, and about 4.5 percent fewer ELS students are promoted to the next grade at the end of the year. These types of differences under-gird the critiques that ELS programs fail to improve the outcomes of LEP students.

The remainder of Table 1, however, provides evidence suggesting these raw differences in student performance may not reflect causal effects. The second three columns of Table 1 show that students who receive English Language Services differ in several observable dimensions that may affect achievement from those who do not, even among this population already selected as 'at risk'. ELS participants tend to be older, and are more likely than non-participants to be Hispanic, be born outside of the U.S., have Spanish as the primary language spoken in their home, and (slightly)

[11] Math and reading achievement outcomes for this study are expressed in terms of z-scores: scores are normalized by the grade and year specific mean and standard deviation for the distribution of *all* students taking the outcome test (including students who speak English fluently and did not take the LAT test) – not only the target population studied in this paper.

to qualify for free or reduced lunch. ELS participants also attend schools in different parts of the District—they are more likely to attend school in the Northern part of the District, and less likely to to attend school in the Northeast. Given the differences in this short list of observable characteristics, it is likely that the raw differences in achievement and promotion depicted in the top rows of Table 1 are confounded by both the differences in the baseline characteristics shown in the lower rows of Table 1 as well as differences in unobserved characteristics like parental education or the amount of time a student has lived in the U.S. Since program participation is not randomly assigned, treatment effect estimates based on naive comparisons of average test scores by program type will be biased by systematic differences between those in ELS and those not.

To overcome these selection problems, I exploit a particular feature in the rules governing eligibility for bilingual and ESL services in LUSDiNE that generates quasi-random assignment to ELS among a subgroup of students. There is a three step process involved in enrolling in English Language Services in LUSDiNE. First, when a new student enrolls in a school, his parents are required to fill out a home language survey indicating whether the student's native language or the primary language spoken at home is other than English. If either of these things are true the student is assessed with a Language Assessment Test (LAT) designed to measure students' abilities in listening, reading, writing, and (for Grades 3 and below, only) speaking English. Students' scores are normed and reported in terms of where the student would rank in a distribution of scores of native English speakers at the same grade level on a similar test. A score at the $40th$ percentile or below results in being categorized as Limited English Proficient (LEP) and eligible by state law for English Language Services; students scoring above this level are ineligible.

If a student qualifies as LEP, placement into bilingual education or ESL classes is determined by parental or teacher choice subject to program availability at the student's school.[12] LEP students are reevaluated through a similar LAT test each Spring and remain eligible for ELS until they score above the 40th percentile. Appendix Figure 1 provides an overview of this process.

[12]When a student initially registers, if he qualifies as LEP and his parents' first choice program is unavailable, the student is offered a transfer to a school where the desired program is available. Subsequent changes between programs are highly discouraged by the District, and are unlikely to be a factor for the students in this analysis since they are in higher grades (at least the third grade) and have nearly all been in bilingual programs before they enter the sample time frame.

I rely on the sharp discontinuity in eligibility status as a function of a student's LAT score at the 40th percentile to identify the causal impact of bilingual education and ESL services. Figure 1 depicts the relationship between LAT scores and the probability of participating in English Language Services created by LUSDiNE's eligibility rules. Nearly 90 percent of those students scoring just below the 40th percentile were enrolled in some sort of ELS program in the subsequent year, while a negligible fraction of those scoring just a few percentiles above this cutoff were enrolled. The crux of the identification strategy used in this paper is to compare the reading and math achievement scores of students scoring just below the 40th percentile cutoff to the scores of those scoring just above the cutoff. If the other determinants of student achievement do not change sharply near the 40th percentile cutoff, then differences in achievement outcomes for students in this neighborhood can be attributed to program participation.

Econometric Framework

To make these arguments more precise, suppose that the relationship between a student's achievement outcome and their participation in a bilingual education program is given by the constant treatment effects model:

$$(1) \quad Y_i = \alpha + T_i\theta + \nu_i$$

where Y_i is student i's achievement test score at the end of the year; α is a constant term; T_i is an indicator function equal to one if student i is enrolled in an ELS program in the year prior to the test; and ν_i is a mean zero error term representing all other determinants of achievement. Letting S_i equal student i's LAT score normalized so that zero corresponds to the 40th percentile, also define an indicator $D_i = 1\{S_i \leq 0\}$ which equals one if the student is eligible for ELS. As discussed above, the primary challenge to estimating the effect of ELS, θ, is the endogeneity of T_i and ν_i.

If $D_i = T_i$, that is if all eligible students enrolled in some form of ELS, then θ could be estimated in a 'sharp' regression discontinuity research design by[13]

$$(2) \quad \theta^{sharp} = \lim_{s\uparrow 0} E[Y_i|S_i = s] - \lim_{s\downarrow 0} E[Y_i|S_i = s],$$

provided that the conditional expectations of the other determinants of achievement, $E[\nu_i|S_i = s]$, are continuous at the 40th percentile cutoff. This identification condition requires that all predetermined characteristics affecting achievement—such as SES, nativity, parental background, etc.—are similar on average between those students just above and just below the cutoff. This creates conditions analogous to a randomized trial, where local to the cutoff the only difference between treatment ($T_i = 1$) and control groups ($T_i = 0$) is ELS participation status.[14]

Figure 2 provides evidence supporting the validity of this key identification assumption. As can be seen from the Figure, the conditional expectation functions of indicator variables for speaking Spanish at home, being eligible for free or reduced lunch, attending school in the Northeast region of the District, and being born abroad are all smooth through the 40th percentile score on the LAT. Similar analyses not reported show that there are no differences in any of the baseline characteristics reported in Table 1. This analysis is similar to a test for randomization in the randomized control trial (Lee 2003). The set of covariates analyzed here is quite sparse, and it is ultimately impossible to know whether other unobserved factors affecting achievement differ markedly for students scoring just above relative to those just below the cutoff. The fact that no discontinuities are evident in any of the available covariates, however, buttresses the supposition that achievement differences among these students near the cutoff are causally attributable to differences in ELS participation.[15]

In this paper I will focus primarily on estimating the treatment effect of ELS *eligibility* on

[13]The notation $\lim_{s\uparrow 0} E[Y_i|S_i = s]$ refers to the limit in the expected outcome score as LAT scores are arbitrarily increased upwards to the 40th percentile. This corresponds to what is referred to in the text as the average outcome of test scores for students just below the 40th percentile. Analogously, $\lim_{s\downarrow 0} E[Y_i|S_i = s]$ refers to the limit in the expected outcome score as LAT scores are decreased downwards to the 40th percentile. This is referred to in the text as the average outcome score for students scoring just above the 40th percentile.

[14]See Hahn, Todd, and van der Klaauw (2001) for formal derivations.

[15]I note that the data fail an additional test for over-identification often employed in the regression discontinuity literature: the density of the LAT score appears to be discontinuous in the neighborhood of the 40th percentile cutoff score. While the sparse nature of support makes accurate inference difficult, there appears to a 'lump' in the density of LAT scores just above the 40th percentile (this is perceptible with some difficulty in Appendix Table 2). Since observable covariates seem to balance, however, there is no evidence of manipulation of LAT scores that might compromise the results presented here.

achievement outcomes using the sharp regression discontinuity design, which is given by equation (2). While nearly all eligible students participate in ELS (about 90 percent), there is some slippage between eligibility and participation. A simple modification to (2) above allows computation of a treatment effect for ELS *participation* by

$$(3) \quad \theta = \frac{\lim_{s\uparrow 0} E[Y|S=s] - \lim_{s\downarrow 0} E[Y|S=s]}{\lim_{s\uparrow 0} E[T|S=s] - \lim_{s\downarrow 0} E[T|S=s]}.$$

Intuitively, this estimator simply magnifies the estimate of the discontinuity in achievement scores given by (2) by the inverse of the difference in proportions of students who enroll in ELS between ELS-eligibles and ineligibles. Technically, this approach uses program eligibility as an instrument to find the treatment effect of participation in the neighborhood of the 40th percentile cutoff. Since the denominator of (3) is near .90 in nearly all cases, eligibility treatment effects can be converted to approximate participation treatment effects by multiplying by 1.1.

Data

Since the nature of the available data dictates the particulars of my estimation and inference procedures, I briefly describe the data used in this analysis here. The data is a subset of an administrative data set I obtained from LUSDiNE officials containing individual level information for all students in grades three through eight between 1998 and 2002. For the purposes of this paper, I restrict attention to those students who were previously identified as LEP and thus had their English proficiency reassessed on the Language Assessment Test between 1998 and 2001. Student outcomes are measured on Spring achievement tests (given in English) in math and reading between 1999 and 2002. The baseline sample is comprised of 183,254 students who took a LAT test in this time frame and did not leave the school system before the achievement test was administered the following year. In analyses presented here, I further limited the sample to those students scoring between the 15th and 98th percentile on the LAT. The details of further sample selection criteria are described in a data appendix. The conditional expectations of indicator variables for all selection criteria are smooth through the 40th percentile threshold.

Throughout this paper, analyses will involve estimating the conditional expectation functions of several outcome variables as a function of the percentile score of a student on the Language Assessment Test. Appendix Table 2 shows the density of student LAT scores by grade level, pooling across the 4 years of data. For the discussion below, several points about the nature of the LAT test are important. First, students are scored in terms of percentiles, and so normalizing the 40th percentile to zero, the support of the test is discrete including the integers between -25 to 58. Second, the empirical distribution of test scores has sparse support. That is, not every percentile score of the test has a positive density of students. For example, no 4th graders received a LAT score of either -1 or 0 (the 39th or 40th percentiles). In what follows, I model the conditional expectations over the entire range of 84 possible scores—not just the points of support of the empirical distribution.

A related point is that for certain grades, the number of points of support in the test distribution are exceedingly small. At worst, there are only 6 points of support for the distribution of LAT scores above the 40th percentile. As I explain in the next section, the discrete and sparse nature of the support of the LAT score distribution require special consideration in estimating the discontinuities in equations (2) and (3) that are used to measure the effect of ELS.

Estimation and Inference

The literature on regression discontinuity designs has yet to converge on a preferred estimator for the discontinuities in the conditional expectations in (3) that are used to form estimates of treatment effects. Adapting the notation of Porter (2003) to account for the discrete support of the LAT score, rewrite the model for student achievement scores as a function of ELS eligibility and other factors as

$$(4) \quad Y_{ij} = m(S_j) + D_j\beta + \nu_{ij}.$$

Here, j is an index of the points of support in the LAT score distribution. Thus, Y_{ij} is the outcome score for student i with the jth value of S (the score on the LAT). Adopting a potential outcomes framework where Y_{ij}^d is the outcome of a student given that he received treatment $D \in \{0, 1\}$, 1 indicating eligibility for ELS, the parameter of interest is the treatment effect of ELS eligibility, or

$$(5) \quad E[Y_{ij}^1 - Y_{ij}^0 | S_j = 0].$$

In words, this is the average outcome for ELS eligible students with LAT scores at the 40th percentile minus the (counterfactual) outcome of the same group of students had they not been eligible.

Identification of β requires that $m(s)$ is continuous at $s = 0$. The challenge in implementing the regression discontinuity design is thus to find an appropriate estimator for $m(\cdot)$. Porter (2003) has demonstrated that under certain assumptions the difference in local polynomial fits to the left and right of the discontinuity converges at an optimal rate, and several researchers have followed this strategy in practice.[16]

In this paper, however, I follow a separate strain of the literature[17] and specify a flexible parametric model for $m(\cdot)$. Specifically, I estimate equation (4) by least squares including a 3rd degree polynomial in S, fully interacted with the indicator $D_j = 1\{S_j \leq 0\}$, allowing the parameters of each term of the polynomial to vary on either side of the cutoff. For clarity, the estimating equation is

$$(6) \quad Y_{ij} = \gamma_0 + D_j \beta + D_j \sum_{p=1}^{3} \gamma_p (S_j)^p + (1 - D_j) \sum_{p=1}^{3} \gamma_p' (S_j)^p + X_i' \Gamma + v_{ij}.$$

where X_i is a vector of student characteristics. In principle and in practice, inclusion of covariates has a negligible effect on treatment effect estimates since the covariates are smooth functions of the LAT score at the 40th percentile (see Figure 2). They are included in the estimating equation for variance reduction.

This parametric approach is motivated by the discrete support of the LAT score distribution. If the LAT score were measured in continuous units, it would be possible (asymptotically) to average the outcome scores of students whose LAT scores were arbitrarily close to the 40th percentile. In the present application, however, it is not possible to observe ineligible students with LAT scores any

[16]For example, see McCrary and Royer (2003)

[17]See DiNardo and Lee (2002), Card, Dobkin, and Maestas (2004), or Lemieux and Milligan (2004) for examples.

closer than the 41st percentile. Further, due to the sparse support of the LAT score distribution, in some cases no students scored within several percentiles of the 40th percentile. For example, no 5th graders scored between the 36th and the 41st percentile on the LAT. Under these circumstances, β is not identified without some parametric assumption.[18] The 3rd degree polynomial was the modal choice from a procedure fitting (4) to reading and math score outcomes separately for all grade-years using degrees between 1 and 8 and applying the Schwarz criterion.[19]

If $E[\nu_{ij}|S_j = s]$ is continuous at $s = 0$, as suggested by Figure 2, then $\hat{\beta}$ will consistently estimate the causal effect of ELS eligibility on Y. The discrete support of the LAT score distribution, however, raises some important considerations for inference. Referring back to Figure 1, note that the parametric fit for the conditional expectation of ELS as a function of LAT score drawn with a solid line does not perfectly fit the local averages $\bar{Y}_j$, shown as hollow circles. This specification error introduces a common variance component for all students with the same LAT score, and so I rewrite the error term in (6) $\nu_{ij} = a_j + \epsilon_{ij}$, where the error components are independent of each other and the regressors in (4). Accordingly, I adjust my standard errors to account for these specification errors.[20]

4 The Relationship between LAT Scores, Program Eligibility and Participation

To reduce the length of the discussion, I fully present the results of the analysis described above only for those students who took the LAT in the 4th grade. The results for the remaining grades

[18] Local methods could also perform this smoothing function if a sufficiently large bandwidth is used for the kernel. Choosing this bandwidth parameter, however, is also a delicate art over which little consensus has appeared in the literature. Ultimately, computational ease may be the largest factor arguing for the use of a global polynomial approach.

[19] See Schwarz (1978). This procedure is akin to choosing a model based adjusted R-squared.

[20] Applied papers in this field have typically addressed this feature by using clustered standard errors for inference. As Lee and Card (2004) explain, however, this approach implicitly assumes that the specification error at each point of the test score distribution is identical across counterfactual states (that is, $a_j^1 = a_j^0$). Lee and Card propose a more conservative approach to inference that accounts for the variance created by the specification error, a_j. Standard errors calculated in this matter are strictly larger than those estimated using cluster methods, with the difference driven by the deviations between the parametric fit of the conditional expectation function and the local averages (i.e., in figures 1-7, the difference between the scatter points and the fitted lines). On average, the standard errors used here are about 2-3% larger than the 'traditional' clustered standard errors.

are briefly summarized in section 7, highlighting only major differences.

As already demonstrated in Figure 1, scoring below the 40th percentile and thus being classified as LEP has a dramatic effect on program assignment. This is to be expected as state law requires that children identified as LEP receive at minimum some sort of ELS program. This rule appears to be implemented quite faithfully: averaged over the 4 years of the sample, students scoring at the 40th percentile are 89.7 percentage points more likely to be in an ELS program than students scoring just above this cutoff.

Figure 1 shows the probability of program participation separately for the two components of ELS—bilingual education programs and ESL. As can be seen, the probability of participation in both of these programs is a discontinuous function of the LAT at the 40th percentile - for students at the 40th percentile, about 64 percent enroll in ELS classes while about 26 percent are enrolled in bilingual classes. A neglibile fraction of students scoring above the threshold are enrolled in an ELS program of any kind. In the absence of additional assumptions it is not possible to separately identify the effects of bilingual education and ELS relative to immersion, since parental and teacher choice partially determine the program in which eligible students enroll.[21] Thus, the estimates below should be interpreted as measuring a mixture of the achievement effects of bilingual and ESL programs, both relative to immersion.

Figure 3 reveals an interesting fact about the subsequent experiences of students who barely score below the 40th percentile relative to those scoring just above this mark. As we have seen above, students below the cutoff are dramatically more likely to be enrolled in ELS services in the year following the LAT. Figure 3 shows, however, that they remain more likely to be enrolled two, three, and even four years following this test. Students scoring at the 40th percentile are still nearly 33 percentage points more likely to be enrolled in ELS services two years after the LAT, and roughly 15 percentage points more likely to be enrolled three and four years later.[22]

[21] If students do not sort across schools based on the presence of bilingual vs. ESL programs then it may be possible to estimate separate treatment effects for the components of ELS. One potential method is to combine IV and RD approaches by instrumenting for participation in ESL local to the cutoff with an indicator for (the absence of) bilingual program availability.

[22] The conditional expectation functions are estimated using the maximum amount of data available in each case. This involves changing sample composition for each estimate. In practice, this has little effect on the estimates discussed here: they are essentially the same if I focus only on 4th graders taking the LAT in 1998, for whom all 4 years of data are available.

This fact reflects the policy in LUSDiNE that LEP students remain eligible for ELS until they test above the 40th percentile, whereupon they become permanently ineligible. In the absence of any English language learning but with test-retest error symmetrically distributed around a zero mean we would expect the fraction of students enrolled in ELS who originally scored at the 40th percentile to fall by roughly half per year. The fact that the fraction falls from nearly 90 percent to 33 percent between year 1 and year 2 suggests that some English learning is taking place. Since students above the 40th percentile threshold are not given the LAT, however, it is impossible to judge whether ELS programs have a causal impact on English language ability as measured with the LAT.

The results in Figure 3 must be kept in mind when interpreting the achievement effects of ELS estimated two, three, and four years after the LAT that will be presented below. These results are not simply longer run estimates of a single treatment, but rather represent the accumulated effects of multiple treatments. In other words, the treatment associated with scoring at the 40th percentile relative to just above that mark is a 90 percent chance of enrolling in ELS in the next year, a 33 percent chance of enrolling two years later, and a 15 percent chance of enrolling in each of the following years. While it is tempting to attempt to estimate treatment effects separately for students treated different numbers of times, variation in the number of times a student is treated is not randomly assigned and such estimates would likely reflect selection within the group of ELS eligibles.

To summarize, as expected, the eligibility rules for ELS in LUSDiNE sharply divide students with similar levels of language proficiency into two groups whose subsequent education experience is very different. Nearly all students scoring just below the 40th percentile on the LAT enroll in either a bilingual or an ESL program, whereas almost none of the students scoring even slightly above that threshold do so. Critics of bilingual education programs have contended that these programs 'trap' students—once a student is assigned to a program, they tend to stay in the program and segregated from mainstream classes for native English speakers. A strong form of this argument is clearly not supported by the data: students clearly do exit from ELS programs over time. On the other hand, it is notable that a non-negligible fraction of students who ostensibly were within one-question of being re-categorized as fluent on the LAT exam remain in ELS programs not just for the following year, but even up to four years afterwards. In the sections that follow, I turn to the achievement consequences of being categorized as a LEP student and thus being eligible for ELS.

5 ELS Eligibility and Reading Achievement

Nearly all the students scoring at or below the 40th percentile on the LAT enroll in ELS services in the year subsequent to the LAT, and significant fractions of these students are enrolled two years, and to a lesser extent, three and four years later. In stark contrast, practically no students scoring above this cutoff receive any form of services – they are placed in mainstream classes and "immersed" in English. To the extent that immersion is clearly superior to ELS programs in increasing student achievement, this fact should produce a discontinuity in the conditional expectation functions for math and reading outcome scores as a function of LAT scores. In particular, both functions should 'jump up' as they pass through the discontinuity point since students above the 40th percentile will be in immersion rather than in ELS.

Figure 4 shows that for reading scores 1 year after the LAT, no such relationship exists. In this Figure, the local averages of reading scores by LAT score are plotted with circles, along with the parametric fit given by equation (6). In fact, the discontinuity in the conditional expectation function (shown with a solid line) drops downward, indicating that for students at the cutoff, those eligible for ELS perform about one-tenth (.115) of a standard deviation in test scores better than those ineligible for ELS, a statistically significant difference. This implies that ELS *participation* effects a test score gain of 0.13 standard deviations (.115/.897). Interpreting this as a causal effect of ELS participation is strengthened by the lack of difference in observable characteristics shown above in Figure 2.

However, the dashed line plotted in Figure 4, which shows the probability that a student took the reading test exhibits a clear break at the 40th percentile suggesting that a causal interpretation may be inappropriate. While nearly all students in immersion take the reading exam, about 78 percent of LEP students near the cutoff take the exam. To the extent that differences in writing the exam among students with similar LAT scores are related to their likely performance, the estimate of the discontinuity in achievement scores may be biased by differential censoring of these achievement test scores.

There are several possible reasons to suspect that this may be true. The assessment of LEP students is a subject of controversy, and many educators feel that it is inappropriate to test students content knowledge in English because such tests will primarily test English proficiency. It is possible that even among students with the same LAT score, teachers exempt students from taking the reading exam who are progressing poorly in English and would have otherwise performed poorly on the exam. Accountability policies requiring schools to show improvements in test scores of LEP students would only reinforce this tendency. Further, various state and District memoranda in LUSDiNE suggest that students may be exempted if they performed worse on the reading component of the LAT score[23] or if they have been in ELS programs or in the U.S. for a shorter amount of time.

While all of the considerations listed above suggest that the achievement effect estimates may be positively biased in favor of ELS, I find mixed evidence on the sign and magnitude of the correlation between the fraction of students taking the test and effect size estimates. Using data from all grades and years, I find a positive correlation between the fraction taking the test and the estimated effects of ELS on reading achievement, and the opposite result for math, though both estimates are extremely noisy. This is documented at greater length below, and summarized in Figure 8.

The results of analyzing the effects of ELS on achievement several years after the initial LAT test suggest that selection over which students are tested may indeed be a non-issue. Figure 5 shows the expectations of reading scores two and three years following the LAT test (4 year effects are not available because 7th grade reading test scores were not made publicly available in 2002). In each panel the discontinuities in reading achievement scores are about as high or higher than the 1-year effect estimate. Students who were eligible for ELS (in the first year) performed .145 and .10 standard deviations better than ineligibles two and three years following the LAT test,

[23] This component of the test score is not separately available in the data.

though effect for the third year is statistically insignificant from zero. In contrast to the effects measured one year following the LAT, however, these results are not affected by selection as nearly all students on either side of the cutoff take the test – no significant discontinuity in the probability of writing the test is evident in either panel.

Taken together, Figures 4 and 5 provide no evidence that immersion is better for student achievement than ELS programs. If anything, there appears to be a small positive effect on reading outcomes of ELS eligibility on the order of one-ninth of a standard deviation.

6 ELS Eligibility and Math Achievement

The results for the effects of ELS eligibility on math achievement for 4th graders differ only slightly to those for reading achievement so I summarize them briefly here. Figure 6 shows there is essentially no difference—the point estimate is .016, with standard error .049—in math achievement one year after the LAT for students at the 40th percentile based on eligibility status. This remains true looking at outcomes measured two, three, and four years afterwards, as demonstrated in Figure 7. While the point estimates flip sign suggesting an advantage for immersion in years two and three, the sign flips back to favor ELS again in year four and the point estimates are both close to, and statistically indistinct from, zero.

For both the reading and math outcomes presented above, the estimates based on outcomes two, three, and four years after the LAT involve different samples due to the fact that outcome scores are only available to 2002. For example, 1 year effects are based on all 4th graders taking the LAT between 1998 and 2001, whereas 4 year effects are based only on students taking the exam in 1998. Table 2 shows the results equivalent to those presented above broken down by different samples corresponding to the number of years I observe student outcomes in the data. To be clear, the second row of the top panel estimates discontinuities in reading achievement scores estimated one and two years after the LAT for those students present in the data for at least two years (i.e., 4th graders taking the LAT between 1998 and 2000). The Table confirms that accounting for sample composition does not qualitatively alter the main findings above that there appears to be a small positive effect of ELS on reading achievement, but no detectable differences in math achievement, among students scoring at the 40th percentile and just above it.

7 Experiences of Other Grades

The results for grade 4 are not, in general, representative of the measured effects of ELS in all grades and in all years. Even within the fourth grade, there appears to be some heterogeneity in estimates across different years. Tables 3 through 5 show the discontinuities in the conditional expectation functions for ELS, and 1 year reading and math achievement scores for each available grade and year the LAT was administered. Aggregate effect sizes across grades and years, in the last row and column of each panel, are averaged across individual grade-year estimates using a minimum distance estimator.[24]

Overall there is a significant degree of heterogeneity in the estimates in Tables 3 through 5, but few clear patterns are evident across grade levels or across years within grades. There are slight differences across grade levels in the assignment to ELS caused by being classified as LEP. In general, in higher grades, students are slightly less likely to participate in ELS due to scoring below the threshold, and within ELS students are increasingly likely to be in ESL programs rather than bilingual education programs.

Table 4 displays results for the effect of ELS eligibility on reading score achievement. Here, the evidence is quite mixed: the estimated effects are positive and significant in grades 3 and 4, and are on the margin of statistical significance in grades 7 and 8. For grades 5 and 6, however, the estimates are negative.[25] Averaging across all estimates using a minimum distance estimator, I find an average effect of -.009 (.013) standard deviations—suggesting that overall there is little difference in achievement outcomes for students in ELS programs compared to those in mainstream,

[24]This differs from the discontinuity estimates in the Figures, which are estimated using equation (6) on data pooled across years. Results from a pooled regression are similar to the minimum distance estimates when pooling is done across years, but not across grades within a given year reflecting differences in the shape of $m(S_j)$ across grade levels. Note that standard errors of these estimates are computed assuming independence across grades and years. Since students in 4th grade in 1998 are likely in the 5th grade in 1999, this independence assumption may not be valid and the standard errors here may be too small.

[25]I note here that the estimates for grades 3,5, and 6 are the most sensitive to the specification of (6), due to the sparse support of the LAT score distribution in those grades (see Appendix table 2). Incorporating model uncertainty in the standard error calculations is left to future work, but would presumably increase the estimated variance most noticably for these grades.

English-only classes. Table 5 shows similar results for math achievement, and Table 6 confirms that there are only small, and largely insignificant, differences in achievement two years after assignment to ELS or English immersion.

8 Discussion

For 4th grade students in LUSDiNE, participation in ELS programs appears to have a slightly positive impact on student achievement in reading relative to mainstream English-only classes. Math achievement appears to be unaffected by program participation. Evidence from students in grades 3 and 5 through 8 yield a range of treatment effect estimates, ranging from about -.1 to .25 standard deviations. However, most estimates hover close to zero, and on average the results are broadly consistent with the finding that there are there is little significant difference in achievement due to participation in ELS rather than English-only classes.

Before drawing conclusions on the policy relevance of these findings, it is important to understand how the limitations of this study dictate their interpretation. First, as in all regression discontinuity research designs, the estimates correspond to the treatment effect of ELS on students scoring near the 40th percentile on the LAT. To the extent that the effect of ELS varies with English language proficiency then the estimates in this paper may not be adequate for thinking about how immersion would affect students with much lower initial proficiency. Since most commentators appear to believe that the benefits of bilingual are higher for those with lower initial language proficiency, the estimates presented in this paper may be a lower bound on the achievement effects for all ELS participants. Another important consideration is that all students in the sample employed here were previously in ELS. The estimates presented in this paper thus correspond to the effect of additional years of ELS, or immersion, conditional participation in ELS in the year prior to the LAT for students at the 40th percentile of English language proficiency.

In this light, the evidence presented above is perhaps most relevant for thinking about the effects of moving current LEP students (with LAT scores near the 40th percentile) into mainstream classes, relative to the counterfactual of letting them continue in ELS. This is, in fact, a central aim of recent policy changes introduced under the No Child Left Behind Act that attempt to incentivize schools to reclassify LEP children into mainstream classes at increasing rates over time. It is a more direct goal of recent proposals to put a lifetime limit on the number of years a student is eligible for ELS. If the effects of these programs on achievement are the primary consideration, the weight of the evidence in this paper provides no reason to proscribe either ELS or immersion strategies. At least for 4th graders, there is no clear evidence that English-only programs are more effective in promoting student achievement in reading or math. If anything, the data appear to suggest a slight advantage of ELS programs.

Of course, there are other considerations. Bilingual and ESL programs are not free, and standard welfare maximizing calculus demands that the benefits of enrolling the marginal student in an ELS program equal the costs. Although the achievement effects of ELS appear to be quite modest, it is possible that other benefits are created by these programs that justify their expense. In particular, proponents of the programs highlight their role in advancing cultural preservation and native language literacy. On the other hand, critics claim that such programs allow immigrants to remain isolated from mainstream society and thus weaken the social and political fabric of the U.S. Further research is necessary to determine the effects of ELS on these other outcomes and their importance.

Pedagogical and curricular differences aside, the classroom environments of students in ELS programs differ in other ways that past research has identified as important for student learning. In particular, students in bilingual education and ESL programs spend all or part of their days in classes with only other immigrant language learners, who tend to score much lower on achievement tests. Thus, students scoring just below the 40th percentile on the LAT score are assigned to classes with lower average peer quality than their counterparts scoring just a percentile higher on the LAT. The results of this paper suggest that being surrounded with lower performing peers may not significantly affect achievement for this group of students. A large caveat to this conclusion, however, is that I have limited data on the classroom level experiences of students. Assignment to ELS may involve other treatments that I cannot directly observe such as class size, teacher quality, and classroom resources such as books, computers, etc.[26] However, academic and public debate on

[26]More generally, the finding that the achievement levels of similar students assigned to ELS and English-only programs are roughly equal may be consistent with both positive and negative treatment effects of each of the

the achievement effects of bilingual programs has focused on the role of peer quality and language of instruction. For students near the ELS eligibility threshold, this paper shows that the combined effects of these separate channels has very little impact on student achievement in math or reading.

individual program attributes with the net effect summing to zero.

References

Angrist, J. D., and K. Lang (2002): "How Important are Classroom Peer Effects? Evidence from Boston's METCO Program," *National Bureau of Economics Research Working Paper*, w9263.

Antunez, B. (2003): *English Language Learners in the Great City Schools: Survey Results on Students, Languages and Programs*. Council of the Great City Schools, Washington, D.C.

Baker, K., and A. de Kanter (1981): *Effectiveness of Bilingual Education: A Review of the Literature*. U.S. Department of Education, Washington, D.C.

Borjas, G. J. (1995): "The Economic Benefits from Immigration," *Journal of Economic Perspectives*, 9, 3–22.

Card, D., C. Dobkin, and N. Maestas (2004): "The Impact of Nearly Universal Insurance Coverage on Health Care Utilization and Health: Evidence from Medicare," *National Bureau of Economics Research Working Paper*, w10365.

Chiswick, B. R. (1991): "Speaking, Reading, and Earnings among Low-skilled Immigrants," *Journal of Labor Economics*, 9, 149–170.

Crawford, J. W. (1999): *Bilingual Education: History, Politics, Theory and Practice, 4th ed.* Bilingual Education Services, Los Angeles.

——— (2003): *Section G: Programs for English Language Learners, ESEA Implementation Guide*. Small Axe Educational Communications, Inc., Washington, D.C.

Danoff, M. N., B. Arias, J. Gary, and Others. (1978): *Evaluation of the Impact of ESEA Title VII Spanish-English Bilingual Education Programs*. American Institutes for Research, Washington, D.C.

DiNardo, J., and D. S. Lee (2002): "The Impact of Unionization on Establishment Closure: A Regression Discontinuity Analysis of Representation Elections," *National Bureau of Economics Research Working Paper*, 34.

G.J. Burkheimer, J., A. conger, G. Dunteman, B. Elliot, and K. Mowbray (1989): *Effectiveness of Services for Language-minority Limited Enlish Proficient Students*. Research Triangle Institute, Research Triangle Park, NC.

Greene, J. P. (1998): "A meta-analysis of the effectiveness of bilingual education," unpublished mimeo.

Hahn, J., P. Todd, and W. van der Klaauw (2001): "Identification and Estimation of Treatment Effects with a Regression Discontinuity Design," *Econometrica*, 69, 201–209.

Hoxby, C. (2000): "Peer Effects in the Classroom: Learning from Gender and Race Variation," *National Bureau of Economics Research Working Paper*, w7867.

Kindler, A. L. (2002): *Survey of the States' Limited English Proficient Students and Available Educational Programs and Services: 1999-2000 Summary Report.* National Clearinghouse for English Language Acquisition and Language Instruction Educational Programs, Washington, D.C.

Kloss, H. (1998): *The American Bilingual Tradition.* Newbury House, Massachussetts.

Kossoudji, S. A. (1988): "English Language Ability and the Labor Market Opportunities of Hispanic and East Asian Immigrant Men," *Journal of Labor Economics*, 6, 205–228.

Lee, D. S. (2003): "Randomized Experiments from Non-random Selection in U.S. House Elections," mimeo, University of California, Berkeley.

Lee, D. S., and D. Card (2004): "Regression Discontinuity Inference with Specification Error," Center for Labor Economics WP74, University of California, Berkeley.

Lefgren, L. (2004): "Educational Peer Effects and the Chicago Public Schools," *Journal of Urban Economics*, 56, 169–191.

Lemieux, T., and K. S. Milligan (2004): "Incentive Effects of Social Assistance: A Regression-Discontinuity Approach," *National Bureau of Economics Research Working Paper*, w10541.

Lopez, M. H., and M. T. Mora (1998): "The Labor Market Effects of Bilingual Education Among Hispanic Workers," mimeo, University of Maryland.

McCrary, J., and H. Royer (2003): "Does Maternal Education Affect Infant Health? A Regression Discontinuity Approach Based on School Age Entry Laws," Mimeo, University of Michigan and University of California, Berkeley.

Meyer, M., and S. Fienberg (eds.) (1992): *Asessing Evaluation Studies: The Case of Bilingual Education Strategies.* Panel to Review Evaluation Studies of Bilingual Education, Committee on National Statistics, National Research Council. National Academy Press, Washington, D.C.

Migration Policy Research Institute (2004): "Immigration to the United States by Decade: Fiscal Years 1821 to 2000," Website: http://www.migrationinformation.org/GlobalData/charts/final.immig.shtml, Nov. 10, 2004.

Ovando, C. J. (2003): "Bilingual Education in the United States: Historical Development and Current Issues," *Bilingual Research Journal*, 17, 1–24.

Porter, J. (2003): "Estimation in the Regression Discontinuity Model," unpublished mimeo, Harvard University.

Ramírez, J., S. Yuen, D. Ramey, and D.Pasta (1991): *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 U.S. Department of Education. Aguirre International, San Mateo, CA.

Rossell, C., and K. Baker (1996): "The Educational Effectiveness of Bilingual Education," *Research in the Teaching of English*, 30, 7–74.

Sacerdote, B. (2000): "Peer Effects with Random Assignment: Results for Dartmouth Roommates," *National Bureau of Economics Research Working Paper*, w7469.

Schwarz, G. (1978): "Estimating the Dimension of a Model," *The Annals of Statistics*, 6, 497–511.

Thistlewaite, D., and D. Campbell (1960): "Regression-Discontinuity Analysis: An Alternative to the Ex Post Fact Experiment," *Journal of Educational Psychology*, 51, 309–317.

Trejo, S. J. (1997): "Why do Mexican Americans Earn Low Wages," *Journal of Political Economy*, 105, 1235–1268.

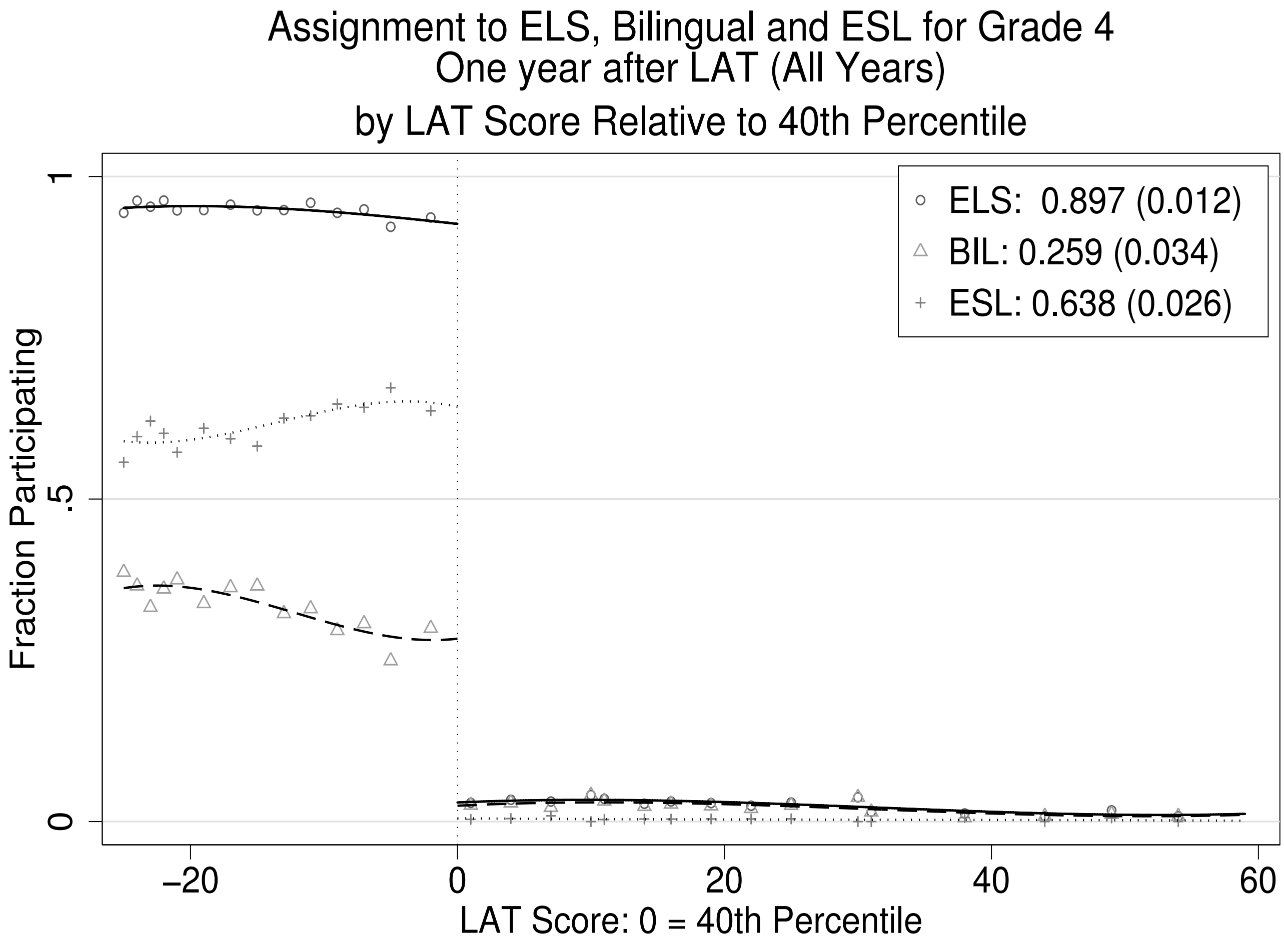
Willig, A. (1985): "A meta-analysis of selected studies on the effectiveness of bilingual education," *Review of Educational Research*, 55, 269–317.

Data Appendix

For the purposes of this paper, I first restrict attention to those students in grades three through eight who were previously identified as LEP and thus took the LAT between 1998 and 2001. While this population is fairly mobile, attrition in the sense of leaving the school district does not appear linked to ELS participation—an indicator function for whether a student disappeared from the data between the LAT and reading or math test is smooth through the 40th percentile on the LAT eligibility cutoff. Accordingly, I drop these students who leave the school system. A large proportion of those students taking the LAT score score in either the 1st (about 15 percent) or the 99th percentile (about 5 percent). Since I focus on differences in outcomes near the 40th percentile, I drop these students to keep them from unduly affecting the estimation of conditional expectation functions. I further drop the small subset of students who take a LAT test in the Spring, and are likely new entrants to the school system. The conditional expectations of indicator functions for each sample selection criterion are smooth through the 40th percentile. Together, these sample restrictions leave a base sample of 183,245 students who took the LAT during 1998 and 2001 and were still present in the school system one year later whether they had valid math and reading achievement scores or not.

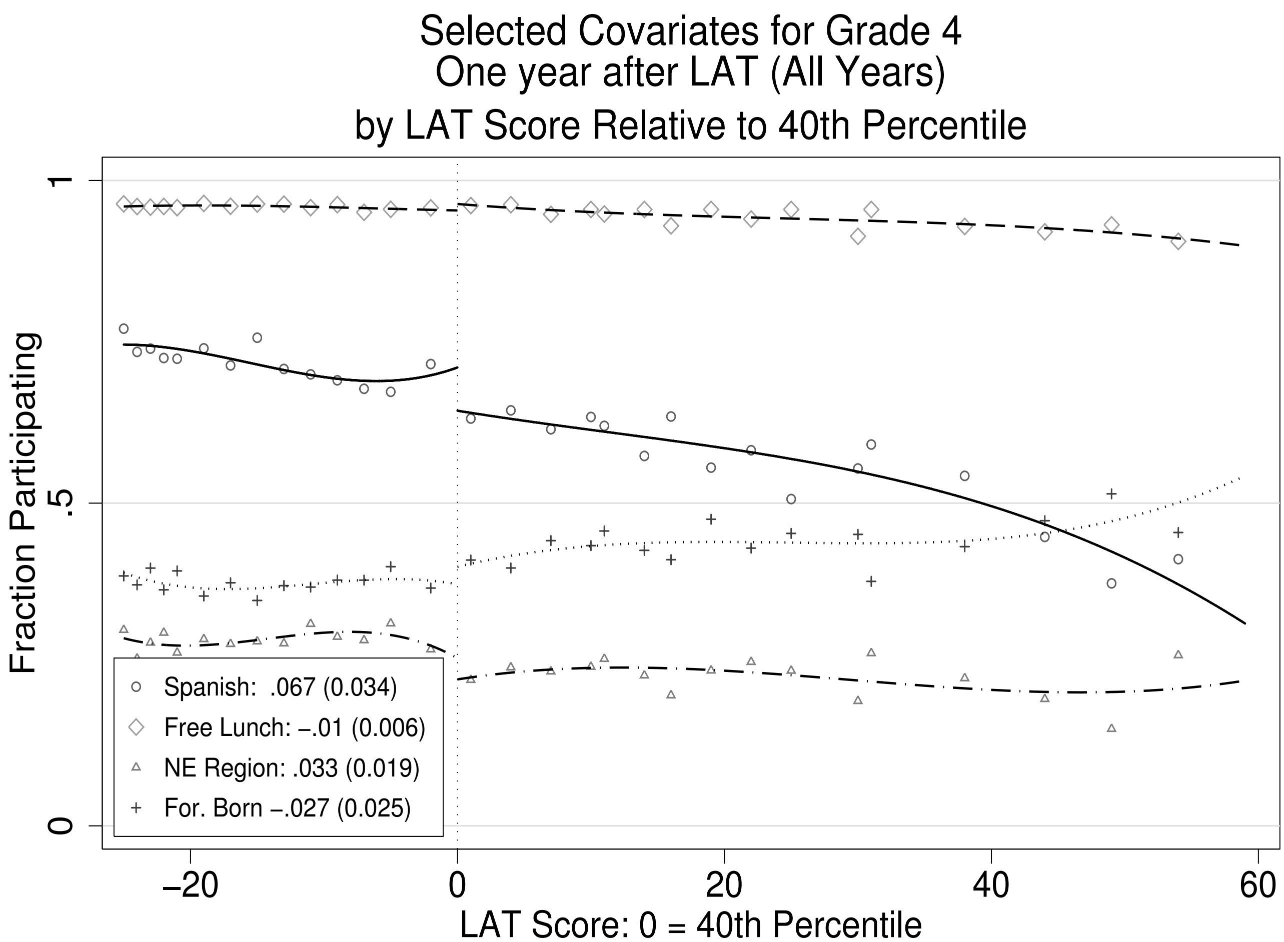
The math and reading outcomes are measured by achievement tests given in English in each Spring in the year following the LAT test. For example, outcomes the year following program enrollment for students taking the LAT in 1998 are measured in the Spring of 1999. Because I have multiple years of data covering several grade levels, I can also examine outcomes measured up to 4 years after student takes the LAT. Since outcome data is only available through 2002 and up to the 8th grade, however, analyzing outcomes 4 years after the LAT mechanically reduces the sample size to the subset of students in grades 3 through 4 who took the LAT in 1998 - reducing the overall sample to 22,082 students. Since the composition of the sample changes with the number of years of outcome data available, care is required in analyzing longer term effects. The breakdown of the baseline sample across years and grade levels, and how this sample is altered when looking at outcomes measured between 1 and 4 years after the LAT, is described in Appendix Table 1.

Figure 1:



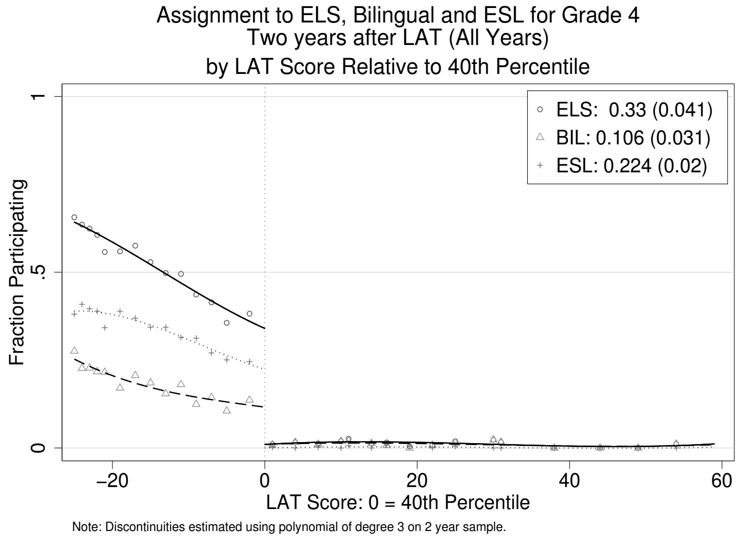
Note: Discontinuities estimated using polynomial of degree 3 on 1 year sample.

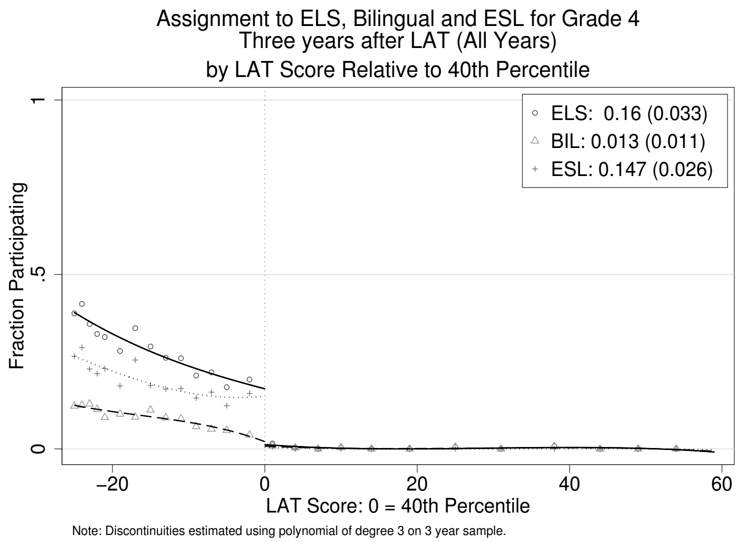
Figure 2:



Note: Discontinuities estimated using polynomial of degree 3 on 1 year sample.

Figure 3:


Assignment to ELS, Bilingual and ESL for Grade 4
Two years after LAT (All Years)
by LAT Score Relative to 40th Percentile
ELS: 0.33 (0.041)
BIL: 0.106 (0.031)
ESL: 0.224 (0.02)
Fraction Participating
0
.5
1
−20
0
20
40
60
LAT Score: 0 = 40th Percentile
Note: Discontinuities estimated using polynomial of degree 3 on 2 year sample.


Assignment to ELS, Bilingual and ESL for Grade 4
Three years after LAT (All Years)
by LAT Score Relative to 40th Percentile
ELS: 0.16 (0.033)
BIL: 0.013 (0.011)
ESL: 0.147 (0.026)
Fraction Participating
0
.5
1
−20
0
20
40
60
LAT Score: 0 = 40th Percentile
Note: Discontinuities estimated using polynomial of degree 3 on 3 year sample.

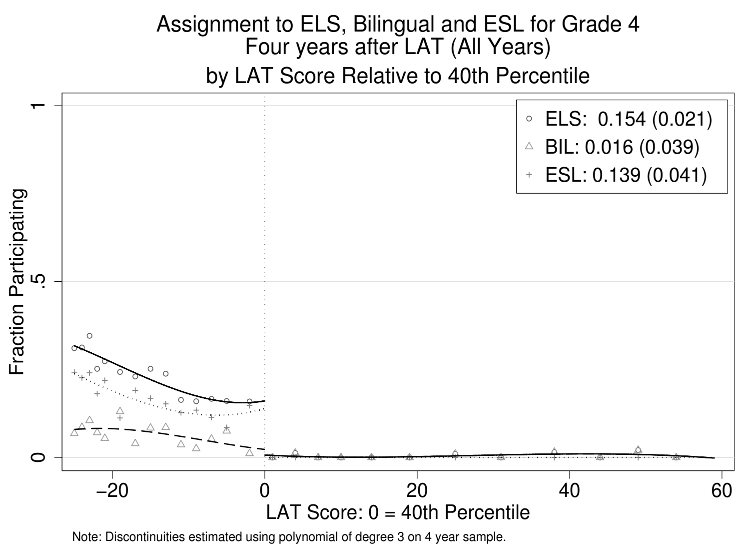

Assignment to ELS, Bilingual and ESL for Grade 4
Four years after LAT (All Years)
by LAT Score Relative to 40th Percentile
ELS: 0.154 (0.021)
BIL: 0.016 (0.039)
ESL: 0.139 (0.041)
Fraction Participating
0
.5
1
−20
0
20
40
60
LAT Score: 0 = 40th Percentile
Note: Discontinuities estimated using polynomial of degree 3 on 4 year sample.

Figure 4:

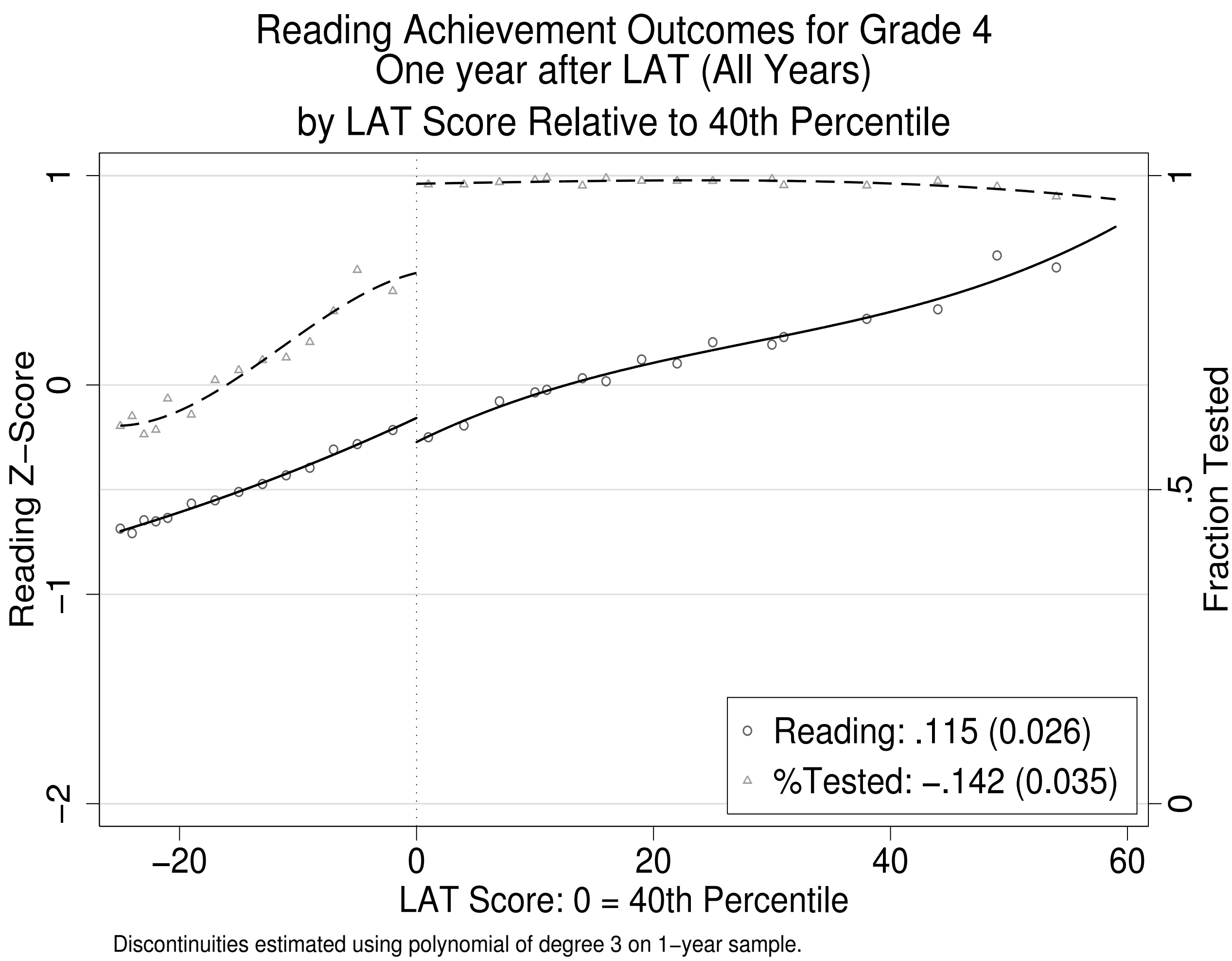

Reading Achievement Outcomes for Grade 4
One year after LAT (All Years)
by LAT Score Relative to 40th Percentile
Reading Z−Score
Fraction Tested
1
0
−1
−2
1
.5
0
−20
0
20
40
60
LAT Score: 0 = 40th Percentile
Reading: .115 (0.026)
%Tested: −.142 (0.035)
Discontinuities estimated using polynomial of degree 3 on 1−year sample.

Figure 5:

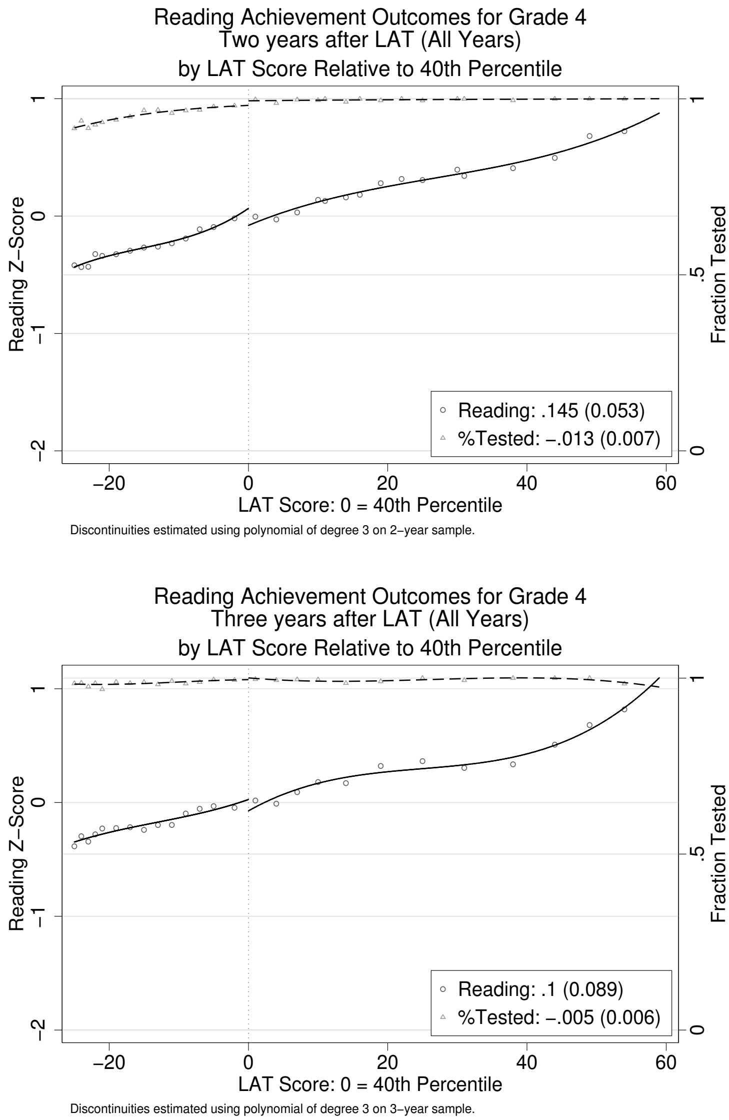
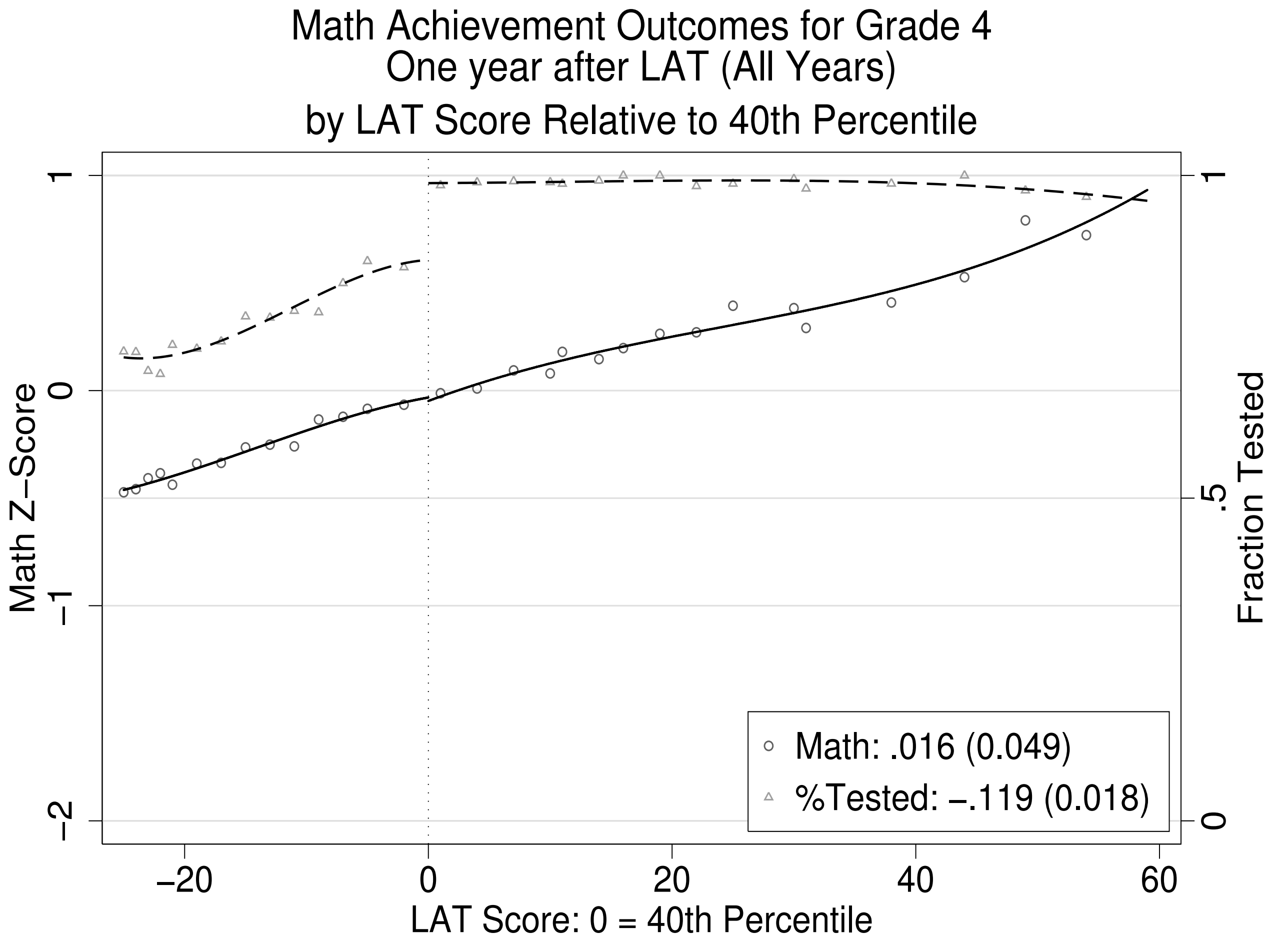
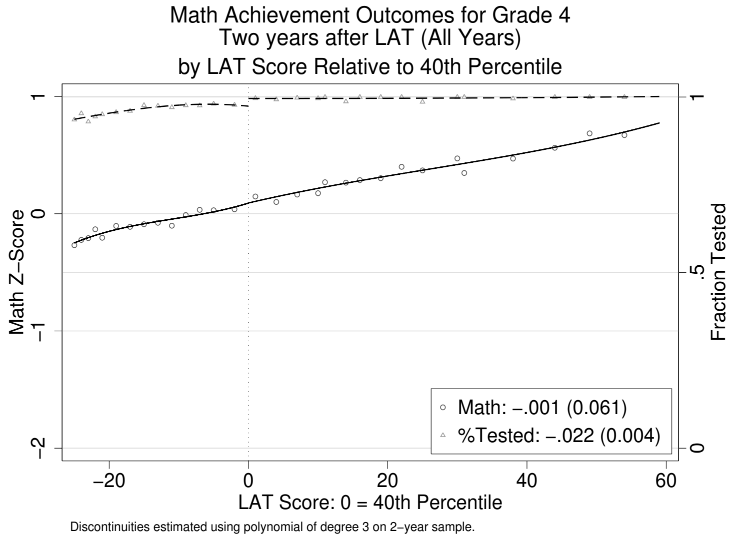

Reading Achievement Outcomes for Grade 4
Two years after LAT (All Years)
by LAT Score Relative to 40th Percentile
Reading Z−Score
Fraction Tested
Reading: .145 (0.053)
%Tested: −.013 (0.007)
LAT Score: 0 = 40th Percentile
Discontinuities estimated using polynomial of degree 3 on 2−year sample.
Reading Achievement Outcomes for Grade 4
Three years after LAT (All Years)
by LAT Score Relative to 40th Percentile
Reading Z−Score
Fraction Tested
Reading: .1 (0.089)
%Tested: −.005 (0.006)
LAT Score: 0 = 40th Percentile
Discontinuities estimated using polynomial of degree 3 on 3−year sample.

Figure 6:

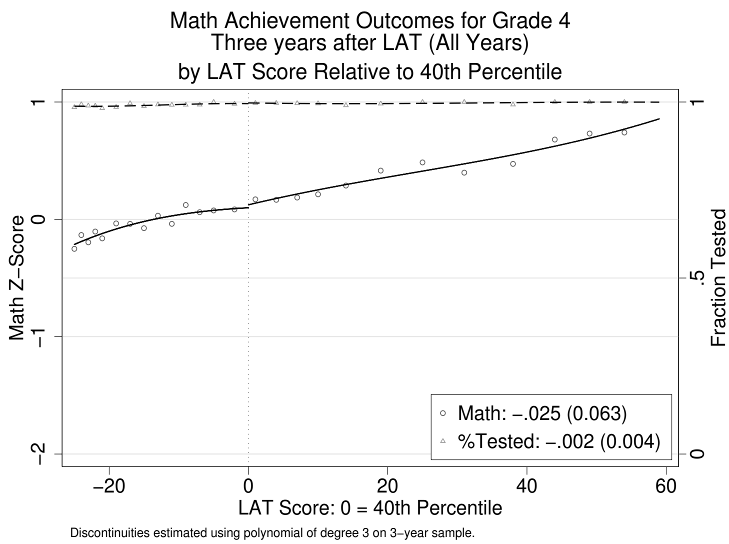


Discontinuities estimated using polynomial of degree 3 on 1−year sample.

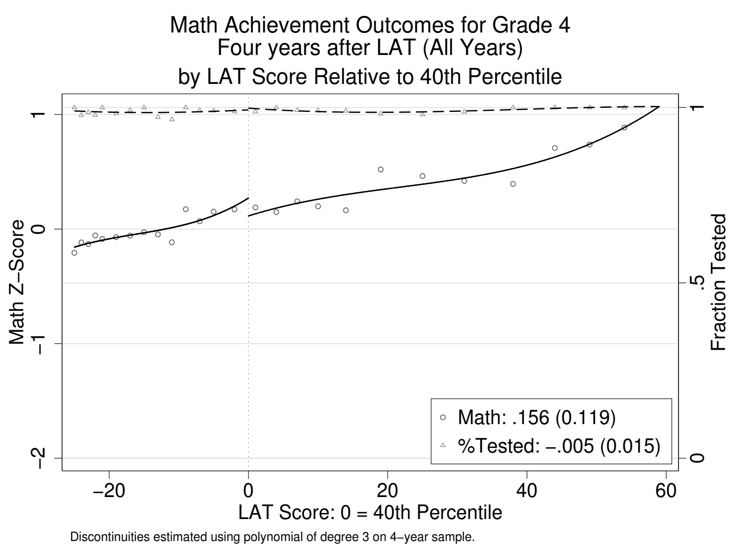
Figure 7:



Math Achievement Outcomes for Grade 4
Two years after LAT (All Years)
by LAT Score Relative to 40th Percentile
Math Z−Score
Fraction Tested
1
0
−1
−2
1
.5
0
−20
0
20
40
60
LAT Score: 0 = 40th Percentile
Math: −.001 (0.061)
%Tested: −.022 (0.004)
Discontinuities estimated using polynomial of degree 3 on 2−year sample.

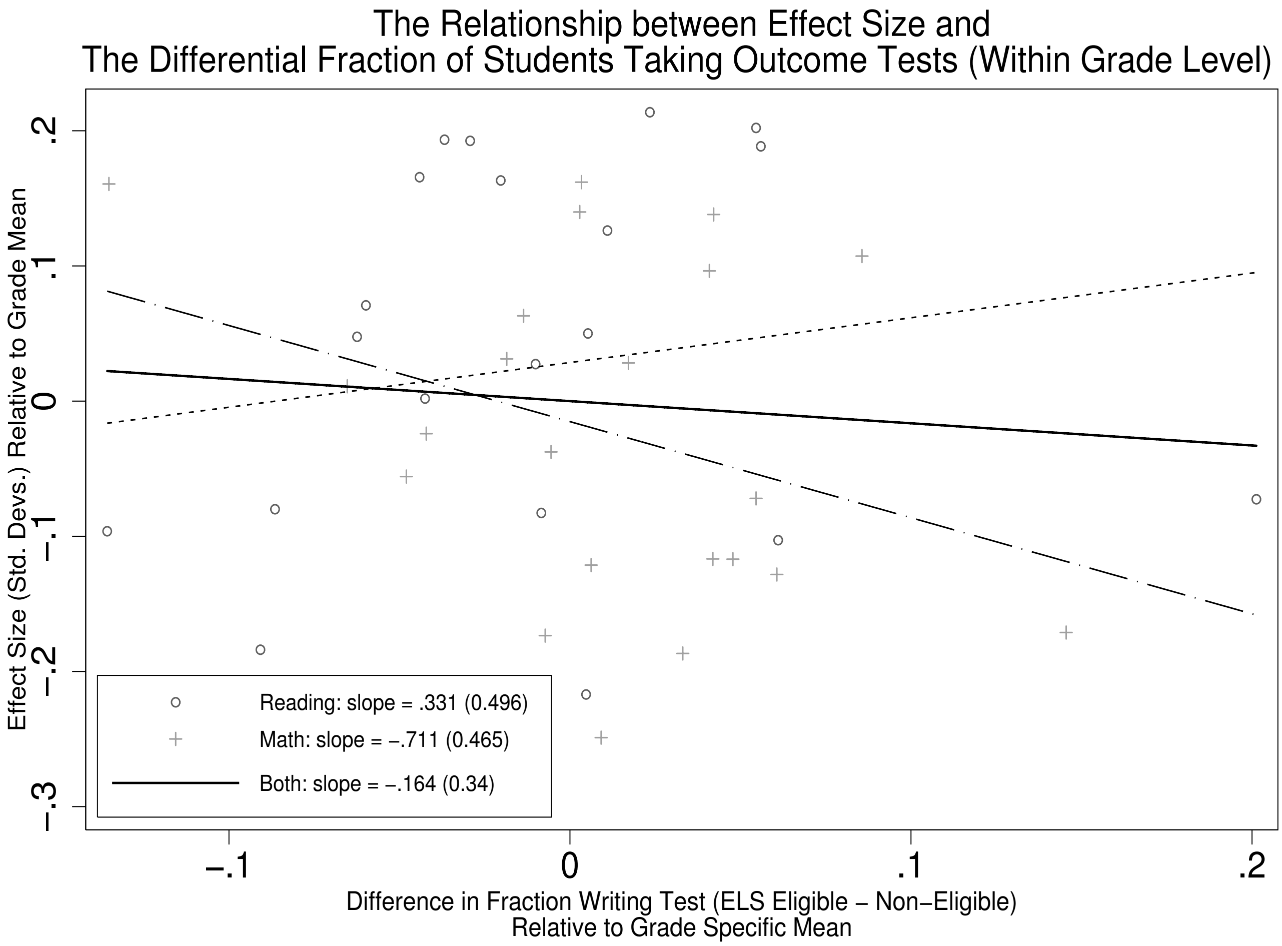


Math Achievement Outcomes for Grade 4
Three years after LAT (All Years)
by LAT Score Relative to 40th Percentile
Math Z−Score
Fraction Tested
1
0
−1
−2
1
.5
0
−20
0
20
40
60
LAT Score: 0 = 40th Percentile
Math: −.025 (0.063)
%Tested: −.002 (0.004)
Discontinuities estimated using polynomial of degree 3 on 3−year sample.



Math Achievement Outcomes for Grade 4
Four years after LAT (All Years)
by LAT Score Relative to 40th Percentile
Math Z−Score
Fraction Tested
1
0
−1
−2
1
.5
0
−20
0
20
40
60
LAT Score: 0 = 40th Percentile
Math: .156 (0.119)
%Tested: −.005 (0.015)
Discontinuities estimated using polynomial of degree 3 on 4−year sample.

Figure 8:



Note: Effect and attrition sizes are estimated from models using polynomial of degree 3.

Table 1:
Sample Means of Selected Student Characteristics
by Program Participation One Year after LAT

Variable Name	Full Sample	All Students: ELS	All Students: No ELS	All Students: Diff.	Students Receiving ELS: Bilingual	Students Receiving ELS: ESL	Students Receiving ELS: Diff.
Following year outcomes							
Reading z-score	-.627	-.872	-.237	-.635	-.949	-.815	-.134
	(.003)	(.003)	(.004)	(.005)	(.005)	(.004)	(.006)
Math z-score	-.432	-.611	-.07	-.541	-.602	-.622	.02
	(.002)	(.003)	(.004)	(.005)	(.004)	(.004)	(.006)
Promoted	.914	.901	.946	-.045	.893	.908	-.015
	(.001)	(.001)	(.001)	(.002)	(.002)	(.001)	(.002)
Baseline characteristics							
Female	.478	.477	.482	-.004	.499	.461	.038
	(.001)	(.001)	(.002)	(.003)	(.002)	(.002)	(.003)
Age	11.922	12.093	11.471	.622	12.034	12.137	-.103
	(.005)	(.006)	(.01)	(.011)	(.009)	(.008)	(.012)
Asian	.154	.144	.181	-.037	.048	.216	-.168
	(.001)	(.001)	(.002)	(.002)	(.001)	(.001)	(.002)
Hispanic	.706	.731	.639	.093	.912	.595	.318
	(.001)	(.001)	(.002)	(.002)	(.001)	(.002)	(.002)
Black	.056	.053	.064	-.011	.028	.072	-.044
	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)
White	.083	.071	.115	-.045	.01	.116	-.106
	(.001)	(.001)	(.001)	(.002)	(0)	(.001)	(.001)
Free/Red. Lunch	.966	.969	.958	.011	.976	.963	.012
	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)
Foreign born	.475	.493	.43	.063	.492	.494	-.002
	(.001)	(.001)	(.002)	(.003)	(.002)	(.002)	(.003)
Home Lang = Spanish	.698	.731	.612	.119	.916	.592	.325
	(.001)	(.001)	(.002)	(.002)	(.001)	(.002)	(.002)
West Region	.182	.186	.172	.014	.255	.135	.12
	(.001)	(.001)	(.002)	(.002)	(.002)	(.001)	(.002)
North	.299	.318	.249	.069	.428	.236	.192
	(.001)	(.001)	(.002)	(.002)	(.002)	(.002)	(.003)
Southeast	.262	.259	.271	-.011	.205	.3	-.095
	(.001)	(.001)	(.002)	(.002)	(.002)	(.002)	(.002)
Northeast	.242	.225	.289	-.065	.111	.31	-.198
	(.001)	(.001)	(.002)	(.002)	(.001)	(.002)	(.002)
South	.014	.012	.019	-.007	.001	.02	-.02
	(0)	(0)	(.001)	(.001)	(0)	(.001)	(.001)
Grade 3	.206	.16	.326	-.166	.186	.14	.046
	(.001)	(.001)	(.002)	(.002)	(.002)	(.001)	(.002)
Grade 4	.185	.203	.137	.066	.213	.195	.019
	(.001)	(.001)	(.002)	(.002)	(.002)	(.001)	(.002)
Grade 5	.179	.186	.159	.027	.183	.189	-.005
	(.001)	(.001)	(.002)	(.002)	(.002)	(.001)	(.002)
Grade 6	.155	.149	.171	-.022	.142	.155	-.012
	(.001)	(.001)	(.002)	(.002)	(.001)	(.001)	(.002)
Grade 7	.137	.148	.107	.041	.134	.159	-.026
	(.001)	(.001)	(.001)	(.002)	(.001)	(.001)	(.002)
Grade 8	.139	.153	.1	.054	.141	.163	-.022
	(.001)	(.001)	(.001)	(.002)	(.001)	(.001)	(.002)

Notes: The means above are for the baseline sample, consisting of all students taking the LAT exam who are recorded to remain in the school system for either the math or reading achievement exam one year later.

Table 2:

Sensitivity Analysis of the Effects of Sample Composition:
Estimates of Effect of ELS Eligibility on Reading and Math Achievement
by Sample Definition and Number of Years From LAT Test to Achievement Test

	Years Since LAT Test			
	1 Year	2 Years	3 Years	4 Years
	Reading Results without Covariates			
1 Year Sample	.115	n.a.	n.a.	n.a.
	(.026)	(—)	(—)	(—)
2 Year Sample	.105	.145	n.a.	n.a.
	(.039)	(.053)	(—)	(—)
3 Year Sample	.108	.122	.100	n.a.
	(.058)	(.057)	(.089)	(—)
4 Year Sample	-.106	.077	.104	n.a.
	(.110)	(.056)	(.107)	(—)
	Models for Math without covariates			
1 Year Sample	.016	n.a.	n.a.	n.a.
	(.049)	(—)	(—)	(—)
2 Year Sample	.038	-.001	n.a.	n.a.
	(.080)	(.061)	(—)	(—)
3 Year Sample	.091	.009	-.025	n.a.
	(.090)	(.068)	(.063)	(—)
4 Year Sample	.026	.009	-.003	.156
	(.108)	(.080)	(.092)	(.119)

Notes: All discontinuities are estimated with a 3^{rd} degree polynomial.
Standard errors in parentheses are computed following Lee and Card (2004).

Table 3:

The Effect of Eligibility on Participation English Language Services
by Grade, Year, and Program Type - One year after LAT Test

	1998			1999			2000			2001			All Year Average		
Grade	ELS	Bil.	ESL	ELS	Bil.	ESL	ELS	Bil.	ESL	ELS	Bil.	ESL	ELS	Bil.	ESL
Models without covariates															
Grade 3	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	.904	.302	.602	.926	.344	.582	.926	.337	.584
	(—)	(—)	(—)	(—)	(—)	(—)	(.023)	(.067)	(.079)	(.002)	(.031)	(.03)	(.002)	(.028)	(.028)
Grade 4	.92	.358	.562	.818	.184	.634	.907	.194	.714	.935	.312	.624	.908	.279	.629
	(.014)	(.06)	(.055)	(.025)	(.038)	(.04)	(.02)	(.057)	(.053)	(.018)	(.022)	(.021)	(.009)	(.017)	(.017)
Grade 5	.918	.531	.388	.873	.207	.665	.854	.265	.59	1.007	.212	.794	.935	.289	.644
	(.034)	(.088)	(.088)	(.059)	(.166)	(.131)	(.064)	(.121)	(.062)	(.037)	(.054)	(.057)	(.022)	(.041)	(.036)
Grade 6	.944	.286	.658	.846	.241	.605	.931	.313	.618	.951	.224	.728	.940	.227	.719
	(.033)	(.047)	(.029)	(.024)	(.128)	(.145)	(.002)	(.02)	(.019)	(.002)	(.004)	(.005)	(.001)	(.004)	(.005)
Grade 7	.917	.266	.651	.849	.249	.599	.664	.15	.514	.867	.123	.744	.86	.22	.643
	(.053)	(.036)	(.054)	(.043)	(.045)	(.063)	(.1)	(.052)	(.1)	(.038)	(.063)	(.069)	(.024)	(.023)	(.033)
Grade 8	.869	.243	.626	.85	.183	.667	.801	.225	.575	.934	.228	.706	.867	.2	.658
	(.018)	(.028)	(.017)	(.013)	(.011)	(.009)	(.026)	(.02)	(.038)	(.019)	(.028)	(.031)	(.009)	(.008)	(.008)
All Grades	.906	.279	.625	.845	.187	.664	.929	.259	.614	.938	.228	.719	.934	.227	.694
	(.01)	(.019)	(.014)	(.01)	(.01)	(.009)	(.002)	(.013)	(.015)	(.001)	(.004)	(.005)	(.001)	(.003)	(.004)
Models with covariates															
Grade 3	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	.895	.262	.633	.919	.306	.613	.918	.298	.615
	(—)	(—)	(—)	(—)	(—)	(—)	(.022)	(.063)	(.075)	(.003)	(.029)	(.026)	(.003)	(.026)	(.025)
Grade 4	.915	.317	.598	.820	.15	.67	.898	.147	.751	.934	.273	.662	.904	.245	.669
	(.014)	(.058)	(.054)	(.025)	(.038)	(.037)	(.02)	(.047)	(.045)	(.018)	(.018)	(.021)	(.009)	(.015)	(.016)
Grade 5	.912	.47	.443	.871	.162	.709	.855	.236	.618	1.004	.186	.818	.933	.248	.671
	(.033)	(.076)	(.083)	(.059)	(.16)	(.128)	(.064)	(.123)	(.067)	(.036)	(.043)	(.059)	(.021)	(.035)	(.037)
Grade 6	.940	.26	.681	.849	.202	.646	.926	.277	.649	.950	.205	.744	.936	.211	.733
	(.028)	(.033)	(.03)	(.023)	(.135)	(.152)	(.001)	(.016)	(.016)	(.002)	(.004)	(.005)	(.001)	(.004)	(.005)
Grade 7	.923	.277	.646	.858	.224	.634	.667	.101	.567	.85	.051	.799	.855	.179	.665
	(.055)	(.049)	(.045)	(.039)	(.042)	(.067)	(.095)	(.052)	(.1)	(.037)	(.058)	(.071)	(.024)	(.025)	(.031)
Grade 8	.871	.154	.717	.852	.164	.689	.803	.204	.599	.929	.205	.724	.871	.177	.700
	(.02)	(.028)	(.021)	(.013)	(.021)	(.023)	(.026)	(.04)	(.048)	(.017)	(.026)	(.029)	(.009)	(.013)	(.013)
All Grades	.907	.237	.681	.847	.171	.679	.925	.246	.651	.943	.21	.735	.933	.212	.719
	(.01)	(.018)	(.015)	(.01)	(.017)	(.019)	(.001)	(.013)	(.014)	(.001)	(.004)	(.005)	(.001)	(.004)	(.004)

Notes: All discontinuities are estimated with a 3^{rd} degree polynomial on the subset of students scoring between the 15th and the 98th percentile on the LAT. Standard errors in parentheses are computed following Lee and Card (04).

Table 4:

The Effect of ELS Eligibility on
Reading Score Outcomes and the Probability of Being Tested
One year after LAT Test

Grade	1998 z-score	1998 tested	1999 z-score	1999 tested	2000 z-score	2000 tested	2001 z-score	2001 tested	All Year Average z-score	All Year Average tested
Models without covariates										
Grade 3	n.a.	n.a.	n.a.	n.a.	.365	-.341	-.013	-.395	.268	-.366
	(—)	(—)	(—)	(—)	(.074)	(.102)	(.126)	(.112)	(.064)	(.076)
Grade 4	-.023	-.215	.259	-.074	.107	-.123	.085	-.139	.183	-.119
	(.076)	(.043)	(.037)	(.027)	(.087)	(.045)	(.098)	(.047)	(.03)	(.019)
Grade 5	-.257	-.232	-.234	.105	.005	-.141	-.114	-.159	-.084	-.156
	(.118)	(.058)	(.261)	(.108)	(.055)	(.057)	(.045)	(.027)	(.033)	(.022)
Grade 6	.247	-.103	-.049	-.013	-.163	-.069	-.029	-.083	-.091	-.079
	(.084)	(.04)	(.076)	(.07)	(.023)	(.01)	(.027)	(.006)	(.017)	(.005)
Grade 7	.126	-.109	.057	-.091	.243	.007	n.a.	n.a.	.114	-.03
	(.223)	(.084)	(.072)	(.067)	(.11)	(.039)	(—)	(—)	(.058)	(.031)
Grade 8	.079	-.081	.116	-.112	-.34	-.138	.167	-.069	.121	-.101
	(.23)	(.034)	(.101)	(.022)	(.299)	(.034)	(.084)	(.03)	(.061)	(.014)
All Grades	.039	-.135	.165	-.088	-.08	-.077	-.029	-.087	-.009	-.087
	(.049)	(.02)	(.029)	(.016)	(.019)	(.009)	(.022)	(.006)	(.013)	(.005)
Models with covariates										
Grade 3	n.a.	n.a.	n.a.	n.a.	.443	-.342	.007	-.384	.364	-.361
	(—)	(—)	(—)	(—)	(.063)	(.094)	(.134)	(.1)	(.057)	(.068)
Grade 4	.022	-.236	.281	-.091	.133	-.141	.097	-.152	.197	-.131
	(.067)	(.045)	(.036)	(.025)	(.07)	(.057)	(.091)	(.05)	(.028)	(.019)
Grade 5	-.158	-.306	-.144	.021	.018	-.139	-.037	-.199	-.031	-.196
	(.131)	(.052)	(.259)	(.092)	(.063)	(.051)	(.047)	(.025)	(.036)	(.02)
Grade 6	.252	-.083	.046	-.054	-.182	-.068	.039	-.111	-.092	-.101
	(.085)	(.042)	(.045)	(.047)	(.018)	(.007)	(.029)	(.004)	(.014)	(.003)
Grade 7	.129	-.14	.115	-.119	.231	-.084	n.a.	n.a.	.156	-.105
	(.123)	(.071)	(.074)	(.068)	(.089)	(.046)	(—)	(—)	(.052)	(.034)
Grade 8	.106	-.116	.167	-.132	-.213	-.184	.165	-.076	.126	-.114
	(.186)	(.031)	(.101)	(.028)	(.203)	(.048)	(.09)	(.026)	(.061)	(.015)
All Grades	.082	-.159	.175	-.099	-.098	-.075	.031	-.113	0	-.106
	(.044)	(.019)	(.025)	(.017)	(.016)	(.007)	(.023)	(.004)	(.011)	(.003)

Notes: All discontinuities are estimated with a 3^{rd} degree polynomial on the subset of students scoring between the 15th and the 98th percentile on the LAT. Standard errors in parentheses are computed following Lee and Card (04).

Table 5:

The Effect of ELS Eligibility on
Math Score Outcomes and the Probability of Being Tested
One year after LAT Test

Grade	1998 z-score	1998 tested	1999 z-score	1999 tested	2000 z-score	2000 tested	2001 z-score	2001 tested	All Year Average z-score	All Year Average tested
Models without covariates										
Grade 3	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
	(—)	(—)	(—)	(—)	(—)	(—)	(—)	(—)	(—)	(—)
Grade 4	.018	-.133	.107	-.143	-.012	-.069	-.015	-.115	.02	-.118
	(.109)	(.04)	(.077)	(.029)	(.091)	(.036)	(.057)	(.04)	(.038)	(.018)
Grade 5	-.157	-.166	-.33	.043	-.032	-.094	-.183	-.136	-.119	-.115
	(.1)	(.087)	(.092)	(.113)	(.049)	(.049)	(.094)	(.039)	(.037)	(.028)
Grade 6	.213	-.208	-.061	-.025	.194	-.033	.081	-.057	.092	-.055
	(.115)	(.055)	(.101)	(.056)	(.064)	(.013)	(.024)	(.004)	(.022)	(.004)
Grade 7	-.048	-.104	-.124	-.037	-.07	.002	.069	-.02	-.074	-.026
	(.109)	(.064)	(.089)	(.07)	(.158)	(.039)	(.191)	(.04)	(.06)	(.024)
Grade 8	-.321	-.05	.046	-.064	-.085	-.109	.081	-.082	0	-.072
	(.104)	(.016)	(.08)	(.019)	(.159)	(.025)	(.056)	(.014)	(.041)	(.009)
All Grades	-.073	-.075	-.053	-.079	.03	-.049	.057	-.06	.023	-.06
	(.048)	(.014)	(.039)	(.015)	(.034)	(.01)	(.02)	(.004)	(.015)	(.003)
Models with covariates										
Grade 3	n.a.	n.a.	n.a.	n.a.	.35	-.241	.064	-.264	.177	-.249
	(—)	(—)	(—)	(—)	(.058)	(.085)	(.047)	(.118)	(.036)	(.069)
Grade 4	.07	-.16	.154	-.167	.019	-.101	-.039	-.143	.039	-.148
	(.086)	(.038)	(.085)	(.024)	(.107)	(.036)	(.065)	(.046)	(.041)	(.017)
Grade 5	.022	-.23	-.213	-.024	-.023	-.096	-.084	-.177	-.08	-.148
	(.107)	(.073)	(.081)	(.1)	(.068)	(.045)	(.06)	(.035)	(.037)	(.025)
Grade 6	.172	-.2	.118	-.067	.175	-.036	.191	-.085	.186	-.08
	(.093)	(.054)	(.146)	(.036)	(.05)	(.009)	(.022)	(.003)	(.019)	(.003)
Grade 7	.032	-.115	-.052	-.08	.008	-.071	.266	-.049	.028	-.08
	(.118)	(.049)	(.096)	(.066)	(.116)	(.046)	(.155)	(.055)	(.058)	(.026)
Grade 8	-.106	-.114	.066	-.124	-.158	-.125	.135	-.097	.043	-.108
	(.167)	(.036)	(.114)	(.029)	(.154)	(.029)	(.087)	(.016)	(.059)	(.012)
All Grades	.067	-.148	-.011	-.126	.147	-.05	.131	-.087	.115	-.085
	(.047)	(.02)	(.044)	(.016)	(.03)	(.008)	(.018)	(.003)	(.014)	(.003)

Notes: All discontinuities are estimated with a 3^{rd} degree polynomial on the subset of students scoring between the 15th and the 98th percentile on the LAT. Standard errors in parentheses are computed following Lee and Card (04).

Table 6:

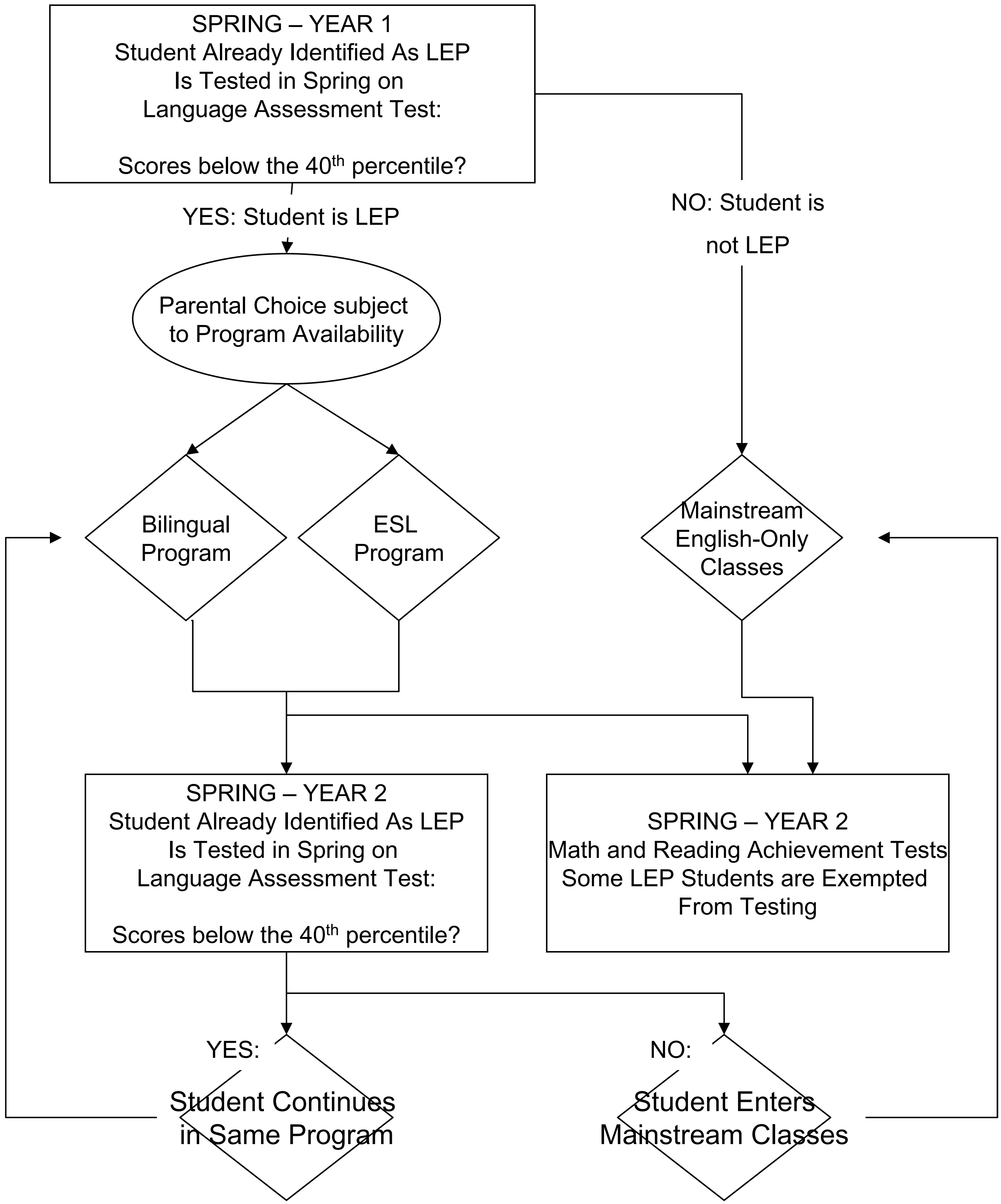
The Effect of ELS Eligibility on
Reading and Math Score Outcomes and the Probability of Being Tested
Two years after LAT Test

	1998		1999		2000		All Year Average	
Grade	z-score	tested	z-score	tested	z-score	tested	z-score	tested
Models for Reading with covariates								
Grade 3	n.a.	n.a.	n.a.	n.a.	.091	.011	.091	.011
	(—)	(—)	(—)	(—)	(.103)	(.027)	(.103)	(.027)
Grade 4	.035	-.007	.206	-.021	.254	-.034	.123	-.016
	(.048)	(.011)	(.07)	(.012)	(.078)	(.02)	(.035)	(.008)
Grade 5	-.153	.057	.074	.084	.168	-.023	.035	-.008
	(.133)	(.037)	(.101)	(.055)	(.14)	(.014)	(.07)	(.013)
Grade 6	.166	-.013	-.21	-.011	n.a.	n.a.	-.07	-.012
	(.161)	(.024)	(.124)	(.04)	(—)	(—)	(.098)	(.021)
Grade 7	.203	-.009	.343	-.037	-.082	.013	.06	-.01
	(.149)	(.036)	(.083)	(.031)	(.054)	(.029)	(.043)	(.018)
All Grades	.038	-.004	.167	-.018	.048	-.017	.08	-.012
	(.042)	(.009)	(.044)	(.011)	(.039)	(.01)	(.024)	(.006)
Models for Math with covariates								
Grade 3	n.a.	n.a.	n.a.	n.a.	.27	-.008	.27	-.008
	(—)	(—)	(—)	(—)	(.098)	(.014)	(.098)	(.014)
Grade 4	-.03	-.017	.027	-.023	.067	-.049	.022	-.024
	(.072)	(.009)	(.086)	(.014)	(.07)	(.016)	(.043)	(.007)
Grade 5	-.162	.015	-.069	.072	-.092	-.031	-.098	-.023
	(.138)	(.028)	(.113)	(.028)	(.07)	(.007)	(.055)	(.007)
Grade 6	-.057	-.074	-.053	.011	.024	.009	.01	.003
	(.059)	(.02)	(.186)	(.038)	(.026)	(.005)	(.024)	(.005)
Grade 7	.077	-.045	.172	.018	.225	-.021	.178	-.015
	(.126)	(.035)	(.073)	(.031)	(.079)	(.022)	(.05)	(.016)
All Grades	-.043	-.024	.067	-.001	.044	-.009	.03	-.011
	(.041)	(.008)	(.048)	(.011)	(.022)	(.004)	(.018)	(.003)

Notes: Different from Tables 3-5, results in this table come from estimating models only for students present in the school system for two years following the LAT. All discontinuities are estimated with a 3^{rd} degree polynomial on the subset of students scoring between the 15th and the 98th percentile on the LAT. Standard errors in parentheses are computed following Lee and Card (04).

Appendix Figure 1:

Assignment to English Language Services In LUSDiNE



Appendix Table 1:

Description of Grade & Year LAT Test "Experiments" and Contributions to Sample Sizes

Grade at LAT Test	Year of LAT Test: 1998 Outcome Years (Grades) Sample Sizes	1999 Outcome Years (Grades) Sample Sizes	2000 Outcome Years (Grades) Sample Sizes	2001 Outcome Years (Grades) Sample Sizes	All Years Outcome Years (Grades) Sample Sizes
Grade 3	**1999-2002 (3-6)**	**2000-2002 (3-5)**	**2001-2002 (3-4)**	**2002 (3)**	**1999-2002 (3-6)**
N 1yr sample	**10,811**	**9,473**	**8,861**	**8,518**	**37,663**
N 2yr sample	**10,064**	**8,845**	**8,265**	**0**	**27,174**
N 3yr sample	**9,321**	**8,319**	**0**	**0**	**17,640**
N 4yr sample	**8,657**	**0**	**0**	**0**	**8,657**
Grade 4	**1999-2002 (4-7)**	**2000-2002 (4-6)**	**2001-2002 (4-5)**	**2002 (4)**	**1999-2002 (4-7)**
N 1yr sample	**9,432**	**8,796**	**8,159**	**7,440**	**33,827**
N 2yr sample	**8,663**	**8,094**	**7,583**	**0**	**24,340**
N 3yr sample	**7,940**	**7,452**	**0**	**0**	**15,392**
N 4yr sample	**7527***	**0**	**0**	**0**	**0**
Grade 5	**1999-2002 (5-8)**	**2000-2002 (5-7)**	**2001-2002 (5-6)**	**2002 (5)**	**1999-2002 (5-8)**
N 1yr sample	**8,866**	**8,159**	**7,855**	**7,916**	**32,796**
N 2yr sample	**8,106**	**7,415**	**7,168**	**0**	**22,689**
N 3yr sample	**7,519**	**6997***	**0**	**0**	**7,519**
N 4yr sample	**5,898**	**0**	**0**	**0**	**5,898**
Grade 6	**1999-2001 (6-8)**	**2000-2002 (6-8)**	**2001-2002 (6-7)**	**2002 (6)**	**1999-2002 (6-8)**
N 1yr sample	**8,024**	**6,958**	**6,756**	**6,724**	**28,462**
N 2yr sample	**7,388**	**6,421**	**6263***	**0**	**13,809**
N 3yr sample	**6,886**	**5,244**	**0**	**0**	**12,130**
N 4yr sample	**0**	**0**	**0**	**0**	**0**
Grade 7	**1999-2000 (7-8)**	**2000-2001 (7-8)**	**2001-2002 (7-8)**	**2002 (7)**	**1999-2002 (7-8)**
N 1yr sample	**7,029**	**6,077**	**6,002**	**5984***	**19,108**
N 2yr sample	**6,501**	**5,601**	**5,145**	**0**	**17,247**
N 3yr sample	**0**	**0**	**0**	**0**	**0**
N 4yr sample	**0**	**0**	**0**	**0**	**0**
Grade 8	**1999 (8)**	**2000 (8)**	**2001 (8)**	**2002 (8)**	**1999-2002 (8)**
N 1yr sample	**6,813**	**6,676**	**5,771**	**6,154**	**25,414**
N 2yr sample	**0**	**0**	**0**	**0**	**0**
N 3yr sample	**0**	**0**	**0**	**0**	**0**
N 4yr sample	**0**	**0**	**0**	**0**	**0**
All Grades	**1999-2002 (3-8)**	**2000-2002 (3-8)**	**2001-2002 (3-8)**	**2002 (3-8)**	**1999-2002 (3-8)**
N 1yr sample	**50,975**	**46,139**	**43,404**	**42,736**	**183,254**
N 2yr sample	**41,609**	**37,196**	**35,542**	**0**	**114,347**
N 3yr sample	**38,552**	**19,693**	**0**	**0**	**58,245**
N 4yr sample	**22,082**	**0**	**0**	**0**	**22,082**

KEY: Contributes to 1-year sample: (at least this dark) Contributes to 2-year sample: (at least this dark) Contributes to 3-year sample: (at least this dark) Contributes to 4-year sample: (at least this dark)

Note: The table describes each of 24 separate "experiments" arising from the Language Assessment Test being administered in 4 different years to students in 6 different grade levels (3-8). line of each cell describes the years and grades for which outcome information (reading and math achievement test scores) are available for each experiment. The following 4 lines cell records the number of observations available for analyzing outcomes each of 4 years after the LAT experiment. Zeros occur because outcome data is available only for student grades 3 - 8, and only between 1999 and 2002. Cells marked with an '*' have no reading outcomes available due to the lack of reading score information for 7th graders in 2002.

Appendix Table 2: Density of LAT Score Density by Grade Level for Baseline Sample

	LAT Score	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	All Grades	F(labscr)
Eligible for English Language Services	-38	834	2,590	1,896	2,864	1,915	2,435	12,534	0.068
	-37	1,249	1,881	2,139	2,217	1,907	1,869	11,262	0.130
	-36	1,620	2,377	1,991	1,214	1,210	1,313	9,725	0.183
	-35	1,300	1,993	1,924	939	1,351	995	8,502	0.229
	-34	818	1,390	1,299	1,081	1,090	1,210	6,888	0.267
	-33	886	1,357	1,787	1,254	867	1,096	7,247	0.306
	-32	0	1,359	1,072	707	865	1,164	5,167	0.335
	-31	993	1,597	1,236	734	952	1,011	6,523	0.370
	-30	0	1,288	679	882	488	1,027	4,364	0.394
	-29	1,135	1,302	650	0	508	848	4,443	0.418
	-28	0	1,171	742	886	1,028	827	4,654	0.444
	-27	1,274	924	773	0	740	616	4,327	0.467
	-26	0	616	866	1,005	817	608	3,912	0.489
	-25	1,332	675	895	0	530	575	4,007	0.511
	-24	0	640	890	1,122	510	641	3,803	0.531
	-23	1,545	641	0	0	572	624	3,382	0.550
	-22	0	618	997	1,269	558	0	3,442	0.569
	-21	1,532	648	0	0	556	587	3,323	0.587
	-20	0	0	999	0	580	0	1,579	0.595
	-19	0	615	0	1,225	257	654	2,751	0.610
	-18	0	0	1,094	0	534	0	1,628	0.619
	-17	1,682	642	0	0	231	602	3,157	0.636
	-16	0	0	1,088	620	529	0	2,237	0.649
	-15	0	591	0	702	0	619	1,912	0.659
	-14	0	0	0	0	510	0	510	0.662
	-13	1,840	654	1,098	0	0	591	4,183	0.685
	-12	0	0	0	0	469	0	469	0.687
	-11	0	639	0	595	0	0	1,234	0.694
	-10	0	0	0	0	430	576	1,006	0.699
	-9	919	604	1,036	635	0	0	3,194	0.717
	-8	1,088	0	0	0	426	0	1,514	0.725
	-7	0	570	0	0	0	509	1,079	0.731
	-6	0	0	0	0	344	0	344	0.733
	-5	0	513	422	0	0	0	935	0.738
	-4	0	0	512	509	0	445	1,466	0.746
	-3	932	0	0	0	344	0	1,276	0.753
	-2	1,069	457	0	0	0	0	1,526	0.761
	-1	0	0	0	627	0	0	627	0.765
	0	0	0	0	0	0	404	404	0.767
Ineligible for English Language Services (ELS)	1	0	862	782	0	588	0	2,232	0.779
	2	0	0	0	0	0	0	0	0.779
	3	0	0	1,024	1,216	0	0	2,240	0.791
	4	0	859	0	0	547	0	1,406	0.799
	5	0	0	0	0	0	893	893	0.804
	6	1,694	0	0	0	0	0	1,694	0.813
	7	0	677	814	0	493	0	1,984	0.824
	8	2,410	0	0	1,422	0	0	3,832	0.845
	9	0	0	0	0	0	0	0	0.845
	10	0	292	0	1,025	393	724	2,434	0.858
	11	0	313	769	0	0	0	1,082	0.864
	12	0	0	0	0	0	0	0	0.864
	13	0	0	0	0	0	0	0	0.864
	14	0	253	596	0	311	0	1,160	0.870
	15	0	0	0	0	0	476	476	0.873
	16	1,795	257	0	0	0	0	2,052	0.884
	17	0	0	0	947	281	0	1,228	0.891
	18	0	0	0	884	0	0	884	0.896
	19	0	245	0	0	0	0	245	0.897
	20	2,495	0	600	0	270	386	3,751	0.917
	21	0	0	490	0	0	0	490	0.920
	22	0	244	0	0	0	0	244	0.921
	23	0	0	0	0	229	0	229	0.923
	24	0	0	0	0	0	0	0	0.923
	25	0	237	0	0	0	329	566	0.926
	26	0	0	0	0	0	0	0	0.926
	27	0	0	0	557	202	0	759	0.930
	28	0	0	0	645	0	0	645	0.933
	29	0	0	789	0	0	0	789	0.938
	30	2,226	186	0	0	0	243	2,655	0.952
	31	0	198	0	0	149	0	347	0.954
	32	2,694	0	0	0	0	0	2,694	0.969
	33	0	0	0	0	0	0	0	0.969
	34	0	0	0	0	0	0	0	0.969
	35	0	0	0	0	0	0	0	0.969
	36	0	0	0	0	107	207	314	0.971
	37	0	0	0	0	0	0	0	0.971
	38	0	319	0	0	0	0	319	0.972
	39	2,301	0	266	0	0	0	2,567	0.986
	40	0	0	284	0	102	0	386	0.988
	41	0	0	0	0	0	0	0	0.988
	42	0	0	0	0	0	111	111	0.989
	43	0	0	0	0	0	0	0	0.989
	44	0	239	0	0	0	0	239	0.990
	45	0	0	0	0	100	0	100	0.991
	46	0	0	0	0	0	38	38	0.991
	47	0	0	0	0	0	46	46	0.991
	48	0	0	0	679	0	0	679	0.995
	49	0	173	0	0	71	0	244	0.996
	50	0	0	0	0	0	0	0	0.996
	51	0	0	297	0	0	0	297	0.998
	52	0	0	0	0	0	67	67	0.998
	53	0	0	0	0	60	0	60	0.999
	54	0	121	0	0	0	0	121	0.999
	55	0	0	0	0	0	0	0	0.999
	56	0	0	0	0	47	0	47	1.000
	57	0	0	0	0	0	48	48	1.000
	58	0	0	0	0	24	0	24	1.000
	Total	37,663	33,827	32,796	28,462	25,092	25,414	183,254	