Economics of Education Review 53 (2016) 311–328

Contents lists available at ScienceDirect

Economics of Education Review

journal homepage: www.elsevier.com/locate/econedurev

The Marginal effect of K-12 English language development programs: Evidence from Los Angeles Schools

Nolan G. Pope[#,*]

The University of Chicago, 1380 E Hyde Park Blvd. 309, Chicago, Illinois 60615, United States

ARTICLE INFO

Article history:
Received 14 April 2015
Revised 19 April 2016
Accepted 25 April 2016
Available online 7 May 2016

JEL classification codes:
I24
I28
J15
H75

Keywords:
School
Students
English
Los Angeles
Language
Spanish
Test
California

ABSTRACT

The growing number of K-12 non-native English speaking students increases the value of optimizing education policy to meet their academic needs. Using a regression discontinuity in test scores from the Los Angeles Unified School District, I analyze the optimal age and English proficiency level for students to enter and exit English language development (ELD) programs. I find marginal kindergarteners receive small academic gains from entering ELD classes. Marginal 2nd to 4th graders who are reclassified from ELD to mainstream English classes receive large benefits to their English test scores (0.25 SD) and GPA that persist over the next 7 years. Boys receive the majority of the benefits. I find no evidence that students reclassified in later grades receive any benefit. Achievement gains can be obtained by enrolling more students into ELD programs in kindergarten and choosing to transition them into mainstream English classes sooner.

Over the last several decades there have been large increases in the number of non-native English speaking students in US schools. Currently, 4.7 million K-12 students, 10% of all K-12 students, have limited English communication skills and are labeled as English Language Learners (ELLs). This is a 27% increase from the 3.7 million ELLs in 2001 (Aud et al., 2012). In addition, Thomas and Collier (2002) project that the ELL population will increase to 17% of K-12 students by 2030, with most of this change coming from increases in the proportion of students speaking Spanish. With this large and growing population of non-native English speaking students, closing the sizeable white-Hispanic educational achievement gap (Fryer & Levitt, 2004; Lee, 2002) could greatly benefit individuals, families, and society.

The growing number of ELLs has increased the attention on how to best meet ELLs' educational needs (Cheung & Slavin, 2005; Rossell & Baker, 1996). There has been particular policy interest in providing special instruction, coursework, and services to help support ELLs (August, 2002; Willig, 1985). Most schools provide separate classes for ELLs with specialized instruction until their English proficiency is deemed adequate. Much of the political debate has focused on how much instruction should be provided in English versus the students' native language. By determining the best policies for ELLs, administrators hope to reduce the Hispanic-White achievement gap.

* Corresponding author. Tel.: +18019959184.
E-mail address: nolanpope2@gmail.com, npope@uchicago.edu

I would like to thank Gary Becker, Steven Durlauf, Eric Eide, James Heckman, Robert Lalonde, Steven Levitt, Nathan Petek, Bryson Pope, Devin Pope, and Joseph Price for helpful comments and discussion.

http://dx.doi.org/10.1016/j.econedurev.2016.04.009

This paper uses a regression discontinuity design to answer two questions. First, at what level of English language proficiency should incoming kindergarteners be placed into mainstream English classes instead of ELD classes? To do this, I test whether entering an ELD program is beneficial to the marginal student with limited English proficiency. Second, after being placed in ELD classes, at what age and level of English language progression should ELLs be moved into mainstream English classes? In particular, I test whether ELLs are currently staying in specialized classes for too long or not long enough. These two questions address whether student achievement can be increased by changing the inflow and outflow of students in ELD programs. I answer these questions for both the short term (1–2 years after reclassification), and the medium term (3–7 years after reclassification) with several outcome measures: English and math state test scores, GPA, attendance, and grade retention.

The Los Angeles Unified School District (LAUSD) uses cutoff values in test scores to classify students as ELLs and reclassify them as mainstream English students. These cutoffs create discontinuities where students with very similar English language ability are assigned to ELD or mainstream classes based on small differences in test scores. Exploiting these discontinuities, I find that marginal kindergarteners placed in mainstream English classes perform slightly worse than marginal kindergarteners placed in ELD programs. I find no evidence of an effect on math test scores. For marginal kindergarteners, being placed in mainstream English classes decreases their 2nd to 8th grade English test scores by 0.014 to 0.104 standard deviations. I find no evidence of an effect on GPA, attendance, and grade retention.

In answering the question regarding the outflow of ELLs, I find that marginal students placed in mainstream classes in 2nd to 4th grade obtain large, persistent benefits in both their English test scores and English GPA. I find no evidence that marginal students in 5th to 10th grade receive such benefit. I also find no evidence that either group sees persistent benefits in math test scores or math GPA. One year after being reclassified, students in 2nd to 4th grade have a 0.163 standard deviation increase in their English test scores compared to students who were not reclassified and 7 years after being reclassified English test scores are 0.306 standard deviations higher. In addition, young boys receive a larger benefit from reclassification than young girls. The benefit for boys over the next 7 years of being reclassified in 2nd to 4th grade ranges from 0.247 to 0.443 standard deviations and for girls it ranges from 0.023 to 0.217 standard deviations. I find no evidence that being reclassified effects attendance or grade retention. These results indicate that reclassifying marginal ELLs sooner to mainstream English classes would lead to large academic benefits.

These results imply that academic gains could be obtained by slightly increasing the inflow of students into ELD programs and by increasing the early outflow of students from these programs. It appears that students with limited English communication skills benefit from ELD programs and they would benefit more by increasing the reclassification rate from ELD programs in earlier grades. In essence, the results imply that expanding ELD programs while shortening the length of time students spend in the program could increase ELLs academic achievement.

Much of the inequality in the United States is explained by the dispersion in human capital such as health and education (Eicher & Garcia-Penalosa, 2001; Mincer, 1958). The dispersion of human capital is caused by many sources, but a leading contributor to these differences is education, particularly at young ages. Restuccia and Urrutia (2004) find that about one-half of the intergenerational correlation in earnings is determined by parental investment into early education. Differences in classroom instruction of ELD and mainstream English classes for young children may have large effects on the differences in individuals' level of human capital and life outcomes.

Many school districts and education agencies have produced reports on the differences between kindergarteners place in ELD programs versus mainstream classes and ELLs versus reclassified ELLs. They find that students placed in mainstream classes perform substantially better than students placed in ELD programs. However, these reports are merely descriptive and do not take into account the clear selection problem that high English ability students are placed in mainstream English classes while low English ability students are not. Besides these descriptive reports, no previous research has looked at the effect of placing kindergarteners into ELD programs versus mainstream English classes. However, there has been a substantial amount of research looking at the effect of reclassification of students to mainstream English classes (Callahan, 2005; Slama, 2012; Umansky & Reardon, 2014; Valentino & Reardon, 2015). This research finds mixed results on the effect of reclassification on students' academic outcomes.

There are three papers that are particularly relevant to this paper. Robinson (2011) uses a regression discontinuity design to look at the effect of reclassification on test scores 1 year after reclassification for older students. His findings corroborate my findings that there is no evidence that older students receive academic gains from reclassification. Matsudaira (2005) also uses a regression discontinuity design for an undisclosed school district and also finds no evidence of an effect of reclassification on test scores for students in older grades 2 years after reclassification. In concurrent work, Robinson-Cimpian, Karen, and Thompson (2015) use both a regression discontinuity design and a policy change to look at the effect of reclassification. They find no evidence of an effect of reclassification on marginal younger students. However, they find that reclassifying a marginal student in 9th or 10th grade has a negative effect on the student's English test scores and likelihood of graduation.

This paper contributes to the literature in the following ways. First, this paper is the first to look at the effect of initial classification of kindergarteners into either mainstream English or ELD classes. Second, this paper contributes to the literature on the reclassification of ELL students by adding additional analysis to this subject. In contrast to past research that found no evidence of an effect for young students, this paper finds large positive academic benefits from the reclassification of young ELL students to mainstream English classes. Third, this paper is able to look at academic outcomes over the next 7 years after reclassification as compared to 1 or 2 years for most previous research. This allows the paper to find that these positive benefits not only persist, but tend to grow over the 7 years after reclassification. Fourth, it finds that young boys receive the majority of the benefit from reclassification. Lastly, this paper looks at additional outcomes variables such as students' GPA, grade retention, and attendance.

In this paper, Section I describes the LAUSD's ELD program and data. Section II describes the regression discontinuity design used. Section III describes the results for kindergarteners and students in 2nd to 10th grade. Section IV covers possible explanations for the results. Section V concludes.

1. Data and institutional background

I use administrative data from the Los Angeles Unified School District (LAUSD) of students who were ELLs in the 2002–2003 and 2003–2004 school years. These data track students' academic achievement through the 2011–2012 school year. A unique feature of the LAUSD is its high concentration of Hispanic students. In 2003 the school district's racial composition was 71.9% Hispanic, 12.1% black, 9.4% white, and 6.6% other[1]. In the LAUSD, 94% of ELLs are Hispanic. The LAUSD is also the second largest school district in the United States and contains the most ELLs of any school district in the nation. In 2003, the LAUSD had a total of 746,852 students. Of these students, 320,594 or 43.3%, were classified as ELLs, meaning they are considered to have limited English proficiency by the school district and as a result are placed in specialized classrooms to facilitate learning English. Also, 15.3% of students had been ELLs but had met all state and school district criteria and were placed in mainstream English classes. In addition to these two groups, for 7.5% of students, English was not their native language but met state and school district criteria to start school in mainstream English classes. Lastly, for 33.8% of students, English was their native language[2].

Similar to most school districts in the United States, the LAUSD uses a standardized procedure to classify incoming kindergarten students as ELLs. When a kindergarten student enrolls at a LAUSD school, the student's parent is given a home language survey. If the language first spoken by the student, the language most frequently used by the student, or the language most frequently spoken by the parent to the student is not English then it is determined that the student's home language is not English and they are a potential ELL. All potential ELLs are then administered the California English Language Development Test (CELDT). The CELDT was designed to determine and track English proficiency and is used by all schools in California. The CELDT has an overall score that ranges from 220 to 710 and is segmented into five proficiency levels with 1 denoting a beginning English proficiency level and 5 denoting an advanced English proficiency level[3]. From 2000 to 2006 the LAUSD students' CELDT scores and English proficiency has slowly risen. This rise is smooth over these years and has no sharp changes from year to year. Approximately, 32% of students score above the CELDT cutoff and have an English proficiency level of 4 or 5. Along with the overall CELDT score, students in 2nd to 11th grade also receive a score for three subtests in listening and speaking, reading, and writing which also have proficiency levels ranging from 1 to 5. With parents' consent, kindergarteners that receive a 4 or 5 overall proficiency level on the CELDT are placed in mainstream English classes. Also with parental consent, students who receive a 1–3 overall proficiency level are classified as an ELL. There is a clear cutoff value that splits each proficiency level.

The LAUSD also uses a standardized procedure to reclassify ELLs to mainstream English classes. This procedure has multiple criteria that must be met in order to be reclassified. At the beginning of each school year all students in 2nd to 11th grade who were an ELL the previous year are administered the CELDT (students in kindergarten and 1st grade are not considered for reclassification). Step 1: If a student receives a 4 or 5 overall proficiency level and 3–5 proficiency levels on each subtest then he moves to step 2; otherwise he remains an ELL. Step 2: The student's teacher or school language appraisal team look at classroom grades and performance and make a reclassification recommendation. If they recommend reclassification, then the student moves to step 3; otherwise he remains an ELL. Step 3: If the student receives a basic or better grade level score on the California Standards Test's English section, then the student is reclassified into mainstream English classes; otherwise he remains an ELL.

For kindergarteners the timing of the tests and placement into mainstream English or ELD classes is as follows. At the time of enrollment, potential ELLs are identified using the home language survey. Within 30 calendar days of the first day of school all kindergarteners are administered the CELDT. Kindergarteners' tests are hand scored at the schools. Kindergarteners are then placed into mainstream English or ELD classes based on their test scores.

For the reclassification of students in grades 2 to 11, the timing of tests and reclassification is more complicated. In the spring of the previous school year the California Standards Test is administered and submitted to the state for scoring. In the fall of the current school year, within 30 calendar days of the first day of school, the CELDT is administered and submitted to the state for scoring. By the spring of the current year the schools receive both the California Standards Test and CELDT scores and the school language appraisal team goes through the three steps of reclassification described above. Students who meet all of the requirements are reclassified and placed in mainstream English class in the fall of the next school year. Students who fail to meet the requirements remain an ELL and start the

[1] Statistics can be found at http://dq.cde.ca.gov/dataquest.

[2] Statistics can be found at http://dq.cde.ca.gov/dataquest.

[3] The correspondence between proficiency levels and CELDT overall scores is slightly different for each grade, however, in general students with CELDT overall scores between 220 and 440 have a proficiency level of 1. Students with CELDT scores between 441 and 487 have a proficiency level of 2. Students with CELDT scores between 488 and 531 have a proficiency level of 3. Students with CELDT scores between 532 and 566 have a proficiency level of 4. Students with CELDT scores between 567 and 710 have a proficiency level of 5.

same process over again. The decision made by the school language appraisal team is recorded in LAUSD's electronic record system and determines the student's ELL status in the fall regardless of whether the student transfers schools or not. In addition, the procedure used to classify and reclassify students did not change over the 2 years studied.

In the LAUSD, the ELLs receive different types of instruction. In 1998 California passed proposition 227 which required all public school instruction, including ELL instruction, to be conducted in English. Due to this California ballot initiative, ELLs were required to be taught in structured English immersion classrooms. As a result, 88.8% of ELLs in 2003 were taught in structured English immersion classrooms. However, parents could specifically request alternative instruction, and in 2003 7.5% of ELLs were taught in bilingual classrooms and 3% were taught in mainstream English classrooms with native language support. The data do not contain the type of ELD program a student is in and therefore all ELLs students are pooled together for the analysis regardless of the ELD program type. In structured English immersion classrooms at least 1 hour of English language development instruction is required daily. In addition, English is used to teach academic subjects with a curriculum and instruction designed for students with limited English proficiency.

The data contains over 400,000 ELL-year observations in 511 schools in the LAUSD for the 2002–2003 and 2003–2004 school years. For each student, the data contains information on gender, parents' education, English and math test scores, percent of days absent, and the school attended through 2012. The data also contains ELL status, CELDT scores, and GPA for 6th through 12th grades in each academic subject through 2009. The data excludes all students in special education. To be included in the sample a student must be a non-special education student, attended school in the LAUSD in the previous year, was an ELL in the previous year, and was administered the CELDT in the previous year. The data include all students in the LAUSD who meet these requirements.

The California Standards Test is a high stakes statewide multiple-choice test given to all California students in 2nd to 11th grade (both ELL and mainstream students). The English and math portions consist of 81 and 71 questions respectively and are each broken into two untimed parts. Also, students are given a grade of either A, B, C, D, or F in each class every semester. For the analysis, both English and math test scores are normalized and all effects are standard deviation changes. The GPA is measured on a 0 to 4 scale (e.g. A = 4.0 and B = 3.0). Table 1 contains a list of variables and summary statistics.

Table 1
Summary statistics.

Variables	Kindergarten		2nd–10th grade	
	Mean	Standard deviation	Mean	Standard deviation
Current year				
Reclassified to mainstream	0.25	0.43	0.09	0.29
Parents education in years	11.38	2.70	11.15	2.63
Male	0.50	0.50	0.52	0.50
CELDT overall score	444.98	85.40	510.29	55.75
CELDT Listening/Speaking Score	–	–	519.27	74.42
CELDT reading score	–	–	497.41	55.61
CELDT writing score	–	–	506.68	56.22
English test score	–	–	297.31	41.33
Math test score	–	–	308.95	60.84
Math GPA	2.67	0.64	2.52	0.66
Listening GPA	2.66	0.64	2.92	0.55
Reading GPA	2.43	0.68	2.72	0.61
Speaking GPA	2.59	0.66	2.91	0.54
Writing GPA	2.33	0.66	2.60	0.60
After 4 years				
English test score	339.71	50.52	309.12	46.48
Math test score	365.50	72.77	295.29	54.90
Fraction of days absent	0.02	0.02	0.08	0.11
English GPA	3.01	0.77	2.03	1.23
Math GPA	2.99	0.98	1.73	1.25
Enrolled in LAUSD (Attrition)	0.77	0.42	0.73	0.44
After 7 years				
English test score	335.87	54.26	323.78	51.92
Math test score	336.68	64.54	286.00	53.17
Fraction of days absent	0.03	0.05	0.07	0.11
Enrolled in LAUSD	0.65	0.48	0.62	0.49
Number of observations	55,273		396,077	

Notes: Summary statistics for kindergarteners include all incoming kindergarteners. Summary statistics for 2nd–10th graders includes only ELL students. Only students in 2nd grade and above have CELDT subtests and California Standards Test scores. For the 2nd–10th column the after 4 years panel only contains information for 2nd–7th grade and the after 7 years panel only contains information for 2nd and 5th grade.

2. Methodology

The classification and reclassification procedures of ELLs described in Section I lead to discontinuities in the classification and reclassification rates of ELLs at the CELDT cutoff score between a 3 and 4 overall proficiency level. These two discontinuities are shown in Fig. 1. Panel A shows the discontinuity in the classification rate for kindergarten and Panel B shows the discontinuity in the reclassification rate for 2nd to 10th grade students. For kindergarteners just below the cutoff and therefore with an overall proficiency level of 3, less than 10% were placed in mainstream English classes. Not all students below the cutoff are placed in ELD classes since parental consent is required and parents may choose to have their children placed in mainstream English classes. However, for kindergarteners just above the cutoff, and therefore with an overall proficiency level of 4, more than 99% were placed in mainstream English classes. This results in a 90-percentage point jump in the likelihood of being placed in mainstream English classes at the cutoff. Similarly, for students in 2nd to 10th grade there is an 11-percentage point jump in the reclassification rate from 1 to 12% for students just below and above the cutoff. Not all students below the cutoff stay in ELD classes since parental consent is required and parents may choose to change their student to a mainstream English class despite the school's decision. These two discontinuities are the variation in the classification and reclassification rates used to estimate the effect on the marginal student near the cutoff of being placed in mainstream English classes.

I exploit discontinuities in Fig. 1 to estimate these effects using a regression discontinuity design. The

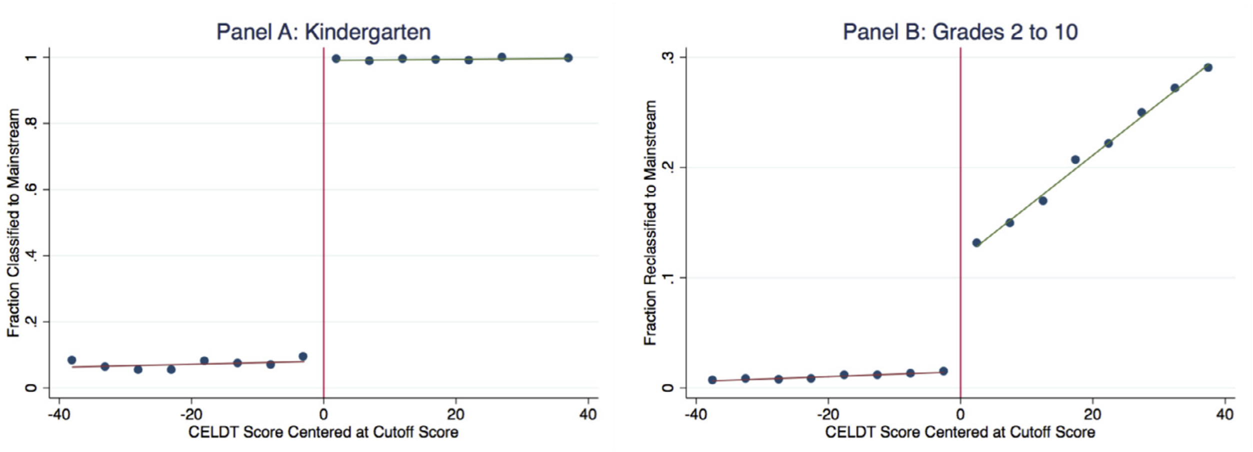

Fig. 1. Classification and reclassification rates at CELDT cutoff.
Notes: Panel A is for all kindergarteners and Panel B is for all students in 2nd–10th grade in the years 2003 and 2004. Each point represents the fraction of students in a given bin that are classified or reclassified. Each bin has a range of 5 overall CELDT points.

estimator of interest when estimating a regression discontinuity design is the Wald estimator:

$$\delta = \frac{\lim_{x\downarrow c}E[Y|X=x] - \lim_{x\uparrow c}E[Y|X=x]}{\lim_{x\downarrow c}E[T|X=x] - \lim_{x\uparrow c}E[T|X=x]} \quad (1)$$

where X is the overall CELDT score, c is the overall CELDT cutoff score, Y is the outcome of interest such as English test scores, GPA, or attendance, and T is a binary variable for a treatment of being placed into mainstream English classes. The estimator δ is interpreted as the effect of the treatment T on the outcome Y for individuals at the cutoff who are moved from one treatment to the other by crossing the cutoff. The numerator is the difference between the average outcomes for individuals right above and below the cutoff. The denominator is the difference in the probability of being classified or reclassified for individuals right above and below the cutoff.

Instrumental variable (IV) estimates of the parameter δ and robust standard errors are estimated using two-stage least squares restricted to individuals near the cut off (Hahn, Petra, & Van derKlaauw, 2001). Unless otherwise specified the two-stage least squares estimation includes individuals within 30 points (half a standard deviation) from the cutoff. Later I will allow this distance to vary to check the robustness of the results. The first stage is as follows:

$$T_i = \alpha_1 + \beta Z_i + \gamma_1(X_i - c) + \pi_1 D_i + \lambda_1 A_i + \theta_1 S_i + \eta_1 G_i + \nu_i \quad (2)$$

where Z_i is a binary variable equal to one if individual i is above the cutoff and zero if below, D_i is a vector of individual demographic information, A_i is a vector of individual academic achievement, S_i is a vector of school dummy variables, G_i is a vector of grade dummy variables, and ν_i is a random error term. The second stage is as follows:

$$Y_i = \alpha_2 + \delta T^*_i + \gamma_2(X_i - c) + \pi_2 D_i + \lambda_2 A_i + \theta_2 S_i + \eta_2 G_i + \varepsilon_i \quad (3)$$

where T^*_i is the predicted value of T_i from the first stage, and ε_i is a random error term. For both kindergarten and 2nd to 10th grade, D_i includes gender and parents' education. For 2nd to 10th grade, A_i includes the English test scores, math test scores, and CELDT subtest scores the year before reclassification. For kindergarteners A_i is not available.

The estimates of δ should be interpreted as a local average treatment effect. This local average treatment effect compares ELLs who are right above the CELDT cutoff score and have passed steps 2 and 3 in the reclassification process to ELLs who are right below the CLEDT cutoff and would have passed steps 2 and 3 in the reclassification process. Note that having passed steps 2 and 3 in the reclassification process is implicitly included since crossing the CELDT threshold would have no effect on reclassification unless ELLs pass steps 2 and 3. The population for which these estimates are applicable is ELLs who were near the CLEDT cutoff and met all other requirements for reclassification. In other words, the pertinent population for these estimates are ELLs who met the academic requirements to be recommended for reclassification by their school language committee, had sufficient state test scores, and had CELDT scores near the CELDT score cutoff. The population for which the estimates are valid can be thought of as the marginal student in the program who is at the CLEDT cutoff. This is a narrow and specific population, but it is also the policy relevant population. Estimates obtained will have policy implications, such as whether to contract or expand the size of the ELD program by making the criteria for reclassification more or less strict. It should also be noted that the results found will only be applicable to a small portion of all ELLs. In addition, the results for marginal older students are primarily for students who have been in ELD programs for an extended period of time and may have other educational problems.

There are two important underlying assumptions for regression discontinuity analysis that must be met in order for the estimates of δ to be unbiased (Imbens & Lemieux, 2008). First, individuals right above and below the cutoff should be similar in both their observables and unobservables. Although discontinuities at the cutoff in unobservable cannot be tested, discontinuities at the cutoff for the observable covariates can be tested. To test for discontinuity in the covariates, I regress each covariate

Table 2
Test for discontinuities in covariates at the CELDT cutoff.

	Parent's education			Test scores				GPA			
Variables	Male	Less than HS	HS	More than HS	English	Math	Listening	Reading	Speaking	Writing	Math
Kindergarten											
Over CELDT Cutoff	−0.028*	0.003	0.003	0.013	–	–	–	–	–	–	–
	[0.016]	[0.013]	[0.012]	[0.012]	–	–	–	–	–	–	–
2nd to 4th Grade											
Over CELDT Cutoff	−0.010	−0.002	0.007	0.003	−0.009	0.016	0.008	−0.003	0.006	−0.002	0.001
	[0.007]	[0.006]	[0.005]	[0.005]	[0.011]	[0.012]	[0.006]	[0.007]	[0.006]	[0.007]	[0.008]
5th to 7th Grade											
Over CELDT Cutoff	−0.009	0.001	0.002	0.001	0.012	0.006	–	–	–	–	–
	[0.007]	[0.007]	[0.005]	[0.005]	[0.010]	[0.010]	–	–	–	–	–
8th to 10th Grade											
Over CELDT Cutoff	−0.005	0.017**	−0.002	0.002	0.003	0.000	–	–	–	–	–
	[0.009]	[0.008]	[0.006]	[0.006]	[0.012]	[0.011]	–	–	–	–	–

Notes: Each cell in the table uses a specification that regresses the indicated outcome variable on a binary variable equal to one if the student's CELDT score is above the cutoff and the student's overall CELDT score using a bandwidth of 30 points. Each cell in the table reports the coefficient on the binary variable for if the student's CELDT score is above the cutoff along with its robust standard error. *** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level.

on a binary variable equal to one if the student's CELDT score is above the cutoff and the student's overall CELDT score. Table 2 reports the coefficient on if the student is above the cutoff for each covariate and each age group. Of the 26 estimated coefficients only two are statistically significant at the 10% level. This indicates that there do not appear to be discontinuities in the covariates at the CELDT cutoff. Although this does not rule out discontinuities in unobservables at the cutoff, this provides some evidence of individuals right above and below the cutoff being similar. In addition, Figs. A.3 and A.4 plots students' covariates by CELDT score. These figures also show that there do not appear to be discontinuities in the demographic or academic covariates at the CELDT score cutoff.

The second assumption that must be met is that students and parents are not able to manipulate students' CELDT scores. Since the difference between being above or below the cutoff can be the difference between answering one additional question correct and because students and parents do not ex-ante know what the cutoff will be, it is unlikely that manipulation of the CELDT score is occurring. Fig. 2 shows the density of the overall CELDT scores, over the range of 40 points (roughly two-thirds of a standard deviation) above and below the cutoff, for both kindergarten and students in 2nd to 10th grade. No clear discontinuities at the cutoff appear to exist for the overall CELDT scores density for both kindergarten and 2nd to 10th grade students. In addition, I test whether there is a discontinuity in the density of CELDT scores at the cutoff following McCrary (2008). However, this test produces an artificial density discontinuity due to the fact that some test values are unattainable in some years and in some grades (Robinson-Cimpian et al., 2015). These unattainable CELDT values cause a nonsmooth density that then causes the test to find artificial discontinuities in the density. These artificial discontinuities are less pronounced when each grade and year are tested separately, but they still exist due to some test values not being possible.

3. Results

The results for the classification of kindergarteners into mainstream English classes are presented first followed by the results for the reclassification of ELLs into mainstream English classes in 2nd to 10th grade. All the results use the two-stage least squares procedure described in the methodology section. Robust standard errors are reported throughout the results because they tend to give the largest standard errors and are therefore the most conservative. Appendix Tables A.15 through A.22 report the main results with OLS standard errors and standard errors clustered by CELDT score. To estimate the effect of reclassification on ELLs the students are split into three grade groups: 2nd to 4th grade, 5th to 7th grade, and 8th to 10th grade. First grade is excluded since students in 1st grade are not eligible for reclassification. Also 11th and 12th grade are excluded since students only take the California Standards Test in 2nd through 11th grade and therefore a future test score is unavailable for students in 11th and 12th grade. These three grades groups are chosen simply because they conveniently split the grades into three groups with three grades each. However, due to the arbitrary nature of this choice, the web appendix shows the results for groups with two grades each and for individual grades.

3.1. Kindergarten classification results

I start the analysis of how kindergarten classification affects students by plotting students' academic outcomes over students' overall CELDT score. Fig. 3 plots standardized English test scores from 2nd to 8th grade, ever retained a grade, and 6th grade GPA over students' overall

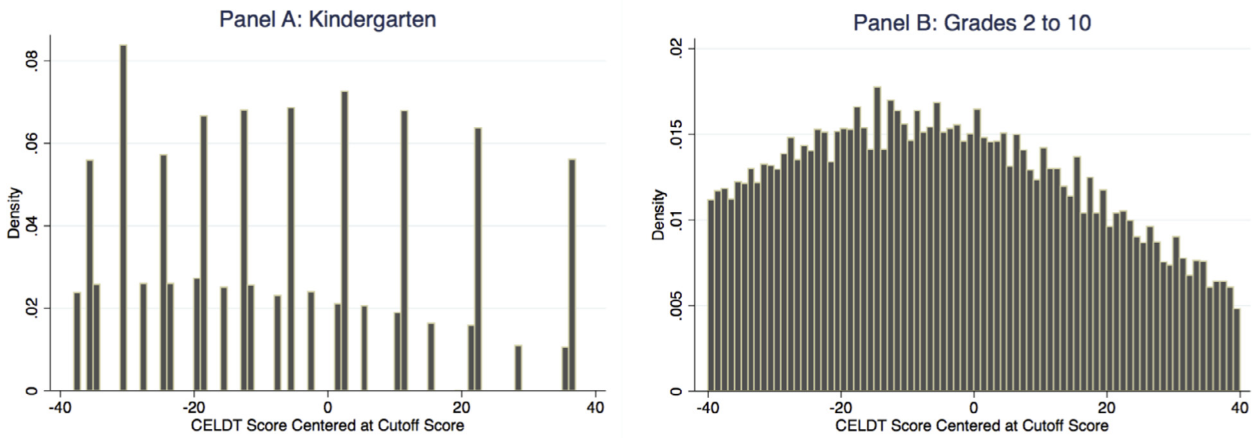

Fig. 2. Density of CEDLT scores.

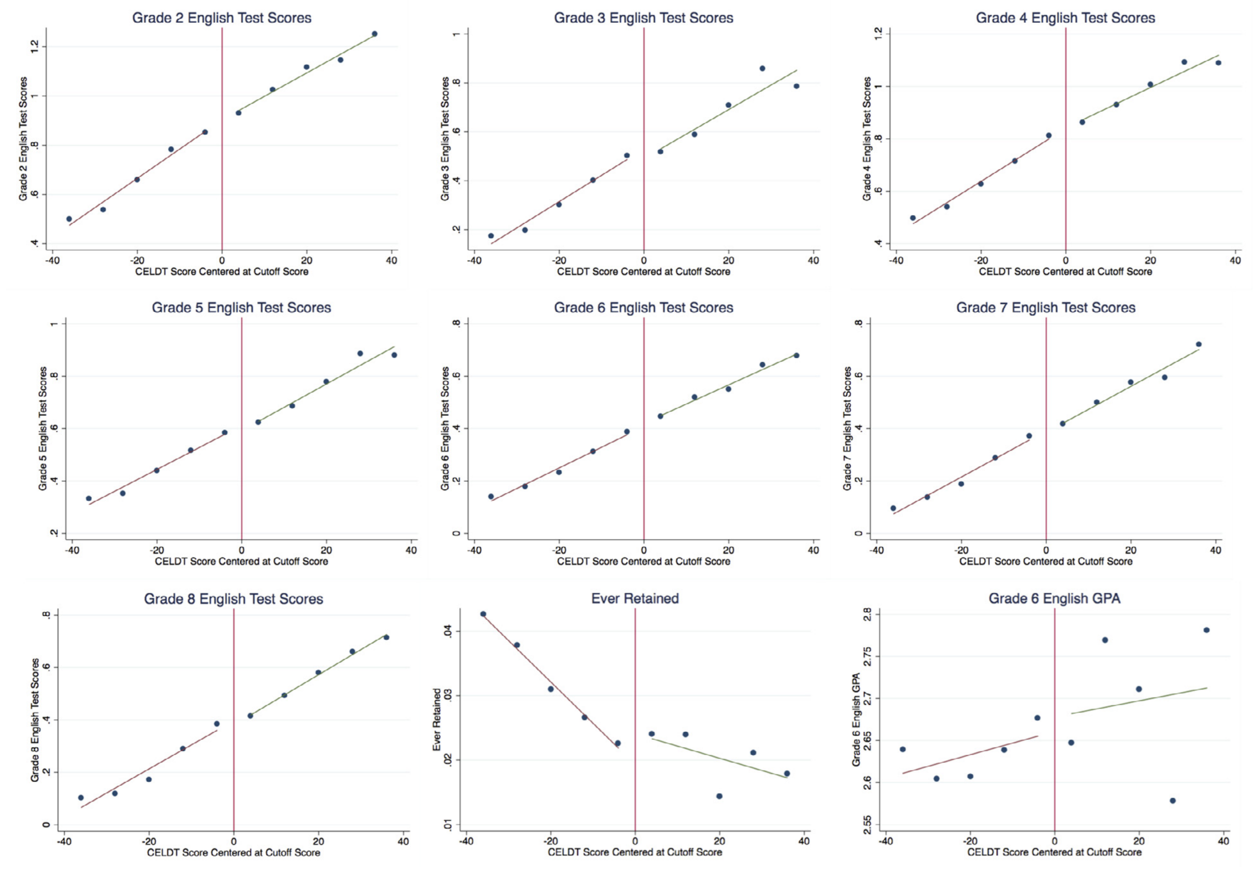

Fig. 3. Outcomes by CELDT score for Kindergarten.
Notes: Each point in the figure represents the mean of a bin. Each bin has a range of 5 CELDT points. The figures from first to last are 2nd–8th grade English test scores, the fraction of students ever retained, and English GPA in 6th grade. The English scores are standardized with a mean of zero, ever retained is a probability of being retained, and English GPA is on a scale from 0 to 4.

CELDT score. Each point represents the average of the outcome variable in a given bin of the overall CELDT scores. For each grade of English test scores there are small to no declines in scores at the cutoff CELDT value. For the ever retained a year in school graph there appears to be an increase in the probability of being retained at the cutoff, but the IV estimate of 0.008 is not statistically significant. The GPA in 6th grade (the only year with GPA data), however, has no jump at the cutoff.

All the additional analysis of the effect of kindergarten classification use the two-stage least squares procedure described in the methodology section. The results for

Table 3
IV Estimates for 8th grade English and Math test scores.

Variables	English test scores in 8th grade			Math test scores in 8th grade		
	Kindergarten			Kindergarten		
	No covariates	Demographic controls	School Fixed effects	No covariates	Demographic controls	School Fixed effects
Mainstream	−0.063	−0.085**	−0.082**	−0.023	−0.031	−0.033
	[0.042]	[0.042]	[0.042]	[0.036]	[0.036]	[0.035]
CELDT score	0.744***	0.718***	0.721***	0.473***	0.436***	0.405***
	[0.087]	[0.087]	[0.086]	[0.074]	[0.074]	[0.072]
Male		−0.189***	−0.180***		−0.049***	−0.045***
		[0.020]	[0.020]		[0.017]	[0.017]
Less than High school		−0.015	0.015		−0.011	0.015
		[0.026]	[0.031]		[0.022]	[0.026]
High school		0.092***	0.088***		0.080***	0.076***
		[0.028]	[0.032]		[0.024]	[0.027]
Some college		0.181***	0.148***		0.198***	0.162***
		[0.037]	[0.039]		[0.032]	[0.034]
College graduate		0.593***	0.449***		0.498***	0.354***
		[0.044]	[0.048]		[0.039]	[0.042]
Post graduate school		0.498***	0.381***		0.454***	0.331***
		[0.069]	[0.066]		[0.062]	[0.059]
School fixed effects			X			X
Observations	9854	9506	9505	9894	9542	9541
Number mainstream	4246	4081	4081	4253	4086	4086
R-squared	0.021	0.058	0.138	0.014	0.044	0.144
F-statistic	28,296	30,418	29,803	28,306	30,397	29,875

Notes: All estimates are for kindergarteners near the cutoff. For columns 1–3 the outcome variable is English test scores in 8th grade and for columns 4–6 the outcome variable is math test scores in 8th grade. The first and fourth columns are the IV results for kindergarteners using Eqs. (2) and (3) with no demographic controls. The second and fifth columns include gender and parents' education dummies as demographic controls. The omitted parent education level is the not reported category. The third and sixth columns add school fixed effects. Robust standard errors are reported in brackets. *** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level.

kindergarteners show how those just above the cutoff, and therefore placed in mainstream English classes, perform compared to kindergarteners just below the cutoff and therefore classified as an ELL and placed in ELD classes.

Table 3 reports the IV estimates of δ for the outcome variable 8th grade standardized English and math test scores. The 8th grade standardized English and math test scores are used for two reasons. First, it is the last grade in the data in which students are administered the California Standards Test. Second, the estimates of δ with this outcome can be thought of as the cumulative effect of being placed in mainstream English classes in kindergarten from kindergarten to 8th grade on English and math academic achievement. The first column for both English and math test scores is the two-stage least squares specification with no covariates. The second column is the baseline specification, which adds the demographic controls. This baseline specification is used since it includes all available demographic controls and it allows for between school variation since the classification policy is at the school district level. Table A.12 and Fig. 3 show the results for various specification choices. Since kindergarten is the first year of formal schooling no prior achievement controls are available. The third column adds school fixed effects. Using the baseline specification, the effect on the marginal kindergartener of being placed in mainstream English classes decreases 8th grade English test scores by 0.085 standard deviations and is statistically significant. The effect on 8th grade math test scores is statistically indistinguishable from zero. These results indicate that when marginal kindergarteners are placed in ELD programs they perform slightly better on future English tests than when they are placed in mainstream English classes.

Table 4 and Fig. A.6 show the dynamic effect of being placed in mainstream English classes as a kindergartener from 2nd to 8th grade. In Table 4 each cell is an IV coefficient on the binary variable of being placed in mainstream English classes as a kindergartener with its robust standard error. Column 1 shows that the effect of being placed in mainstream English classes on English test scores is negative for each of the 7 grades and is statistically significant for 3 of the 7 grades. These effect sizes range from −0.014 to −0.104 standard deviations. Column 2 shows that the effect of being placed in mainstream English classes on math test scores is always close to zero and never statistically significant. In addition to test scores, columns 3 and 4 show there are no significant effects on the probability of being retained a year in school and the fraction of school days absent.

To test the robustness of the kindergarten estimates, Fig. 4 plots the IV estimates of the effect of being placed in mainstream English classes on English test scores for the marginal kindergartener over the bandwidth choice. The bandwidth of 5 is omitted due to the limited amount of data and the large confidence intervals produced. The estimates are relatively constant over the bandwidth choice and the confidence intervals narrow as the bandwidth increases. Table A.10 reports the kindergarten estimates when estimated separately for boys and girls. There is some suggestive evidence that the negative effects of

Table 4
IV estimates of English and Math test scores by grade.

Variables	Kindergarten			
	English	Math	Retained	Fraction of days absent
1st grade	–	–	−0.001	–
	–	–	[0.004]	–
2nd grade	−0.060	−0.058	0.007*	–
	[0.039]	[0.038]	[0.004]	–
3rd grade	−0.104***	−0.007	0.002	–
	[0.034]	[0.038]	[0.002]	–
4th grade	−0.049	−0.024	0.001	–
	[0.035]	[0.036]	[0.002]	–
5th grade	−0.071**	−0.030	–	–
	[0.032]	[0.042]	–	–
6th grade	−0.014	0.024	–	0.000
	[0.034]	[0.035]	–	[0.002]
7th grade	−0.054	−0.042	–	−0.001
	[0.039]	[0.033]	–	[0.002]
8th grade	−0.085**	−0.031	–	0.000
	[0.042]	[0.036]	–	[0.002]

Notes: Each cell in the table is an IV coefficient on the variable placed in mainstream English classes along with its robust standard error. Each cell uses the baseline specification from columns 2 and 5 from Table 8. This baseline specification includes gender and parents' education controls and a bandwidth of 30 points. The columns denote the test subject, retained, and fraction of school days absent and the rows represent the grade in which the outcome was observed. For example the cell in the English column and 6th grade row is the baseline specification with English test scores in grade 6 as the outcome variable for students near the cutoff in kindergarten. *** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level.

being placed in mainstream English classes are larger for girls than boys, particularly in the long run. However, these difference are not statistically significant. In addition, Table A.12 is a robustness table showing the kindergarten results for different specifications.

3.2. Reclassification results

Fig. 5 plots standardized 11th grade English test scores over the overall CELDT score. The standardized 11th grade English test score is used since it is the last grade in which students are administered the California Standards Test and the effect on this outcome can be thought of as the cumulative effect of reclassification on K-11 English academic achievement. The 2nd to 4th grade figure shows a small to no jump in the average English test scores at the cutoff. The 5th to 7th grade figure shows no jump in the average English test scores at the cutoff. The lack of clear graphical evidence of effects at the cutoff for the 2nd to 4th group should lead to a cautious interpretation of the following IV estimates. One possible explanation for the lack of clear graphical evidence is that since moving from below to above the cutoff only increases the likelihood of being reclassified from 1% to 12%, the average jump at the cutoff for the entire population is being muted by the 88% of students above the cutoff who where not reclassified. The 11-percentage point jump in the likelihood of being reclassified is a strong instrument for the IV estimates; however, due to the discontinuity only being 11-percentage points, it does not allow the results to be viewed graphically very easily.

Table 5A reports the IV estimates of δ from each grade group for the outcome variable standardized 11th grade English test score. The first column for each grade group is the two-stage least squares specification with no covariates. This column uses Eqs. (2) and (3) without demographic or academic achievement controls. For marginal students at the cutoff in 2nd to 4th grade being reclassified has a statistically significant increase on their 11th grade English test scores of 0.331 standard deviations. However, for the 5th to 7th and 8th to 10th grade groups I find no evidence of an effect of reclassification on 11th grade English test scores. The second column of each grade group contains controls for gender, parents' education, grade, English and math test scores, CELDT subtest scores, and English GPA. Similar to the no covariates specification, this specification finds large and statistically significant benefits to 11th grade English test scores for marginal students in 2nd to 4th grade, but statistically insignificant effect for students in either 5th to 7th or 8th to 10th grade. The third column of each grade group adds a fixed effect for the student's school. The school fixed effect has a minimal effect on the reclassified coefficient. These results

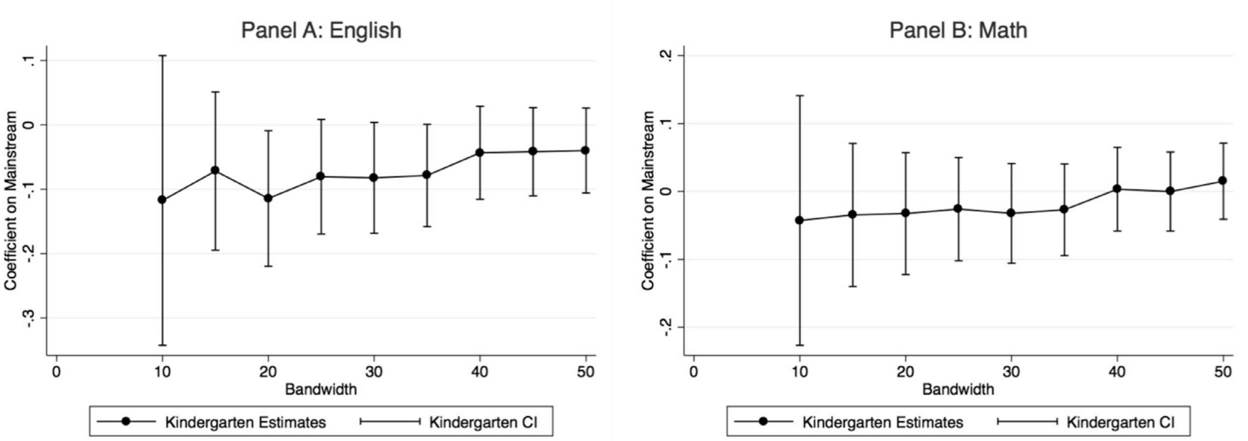

Fig. 4. IV Estimate of mainstream classes on 8th grade test scores over bandwidth.
Notes: These two figures show the robustness of the kindergarten estimates to the choice of bandwidth. The baseline specification in the demographic controls column in Table 9 is used with varying bandwidths. Bandwidths vary from 10 to 50 points.

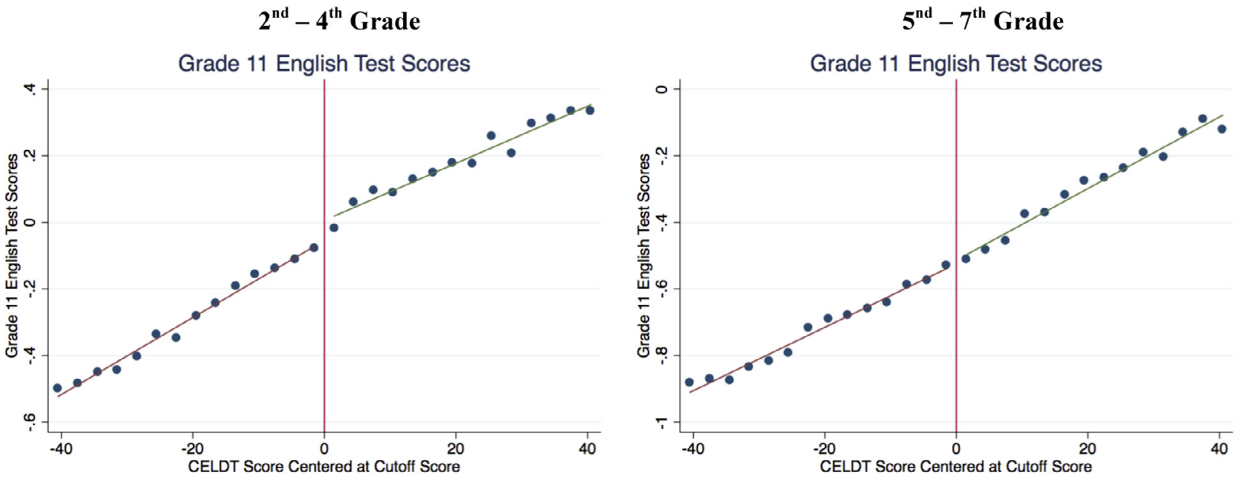

Fig. 5. 11th Grade English Test scores by CELDT score.
Notes: The first graph is for 2nd to 4th grade. The second graph is for 5th to 7th grade. Each point in the figure represents the mean test score of a bin. Each bin has a range of 5 CELDT points.

Table 5A
IV estimates for 11th grade English test scores.

Variables	English test scores in 11th grade								
	2nd–4th grade			5th–7th grade			8th–10th grade		
	No covariates	Demographic/ Performance controls	School fixed effects	No covariates	Demographic/ Performance controls	School fixed effects	No covariates	Demographic/ performance controls	School fixed effects
Reclassified	0.331***	0.297**	0.272**	−0.059	−0.056	−0.051	0.014	−0.181	−0.133
	[0.127]	[0.119]	[0.119]	[0.102]	[0.090]	[0.091]	[0.263]	[0.221]	[0.224]
CELDT Overall	0.684***	−0.071	−0.068	0.747***	−0.468	−0.514	0.664***	−0.794	−0.590
	[0.047]	[0.924]	[0.919]	[0.051]	[0.693]	[0.693]	[0.074]	[1.007]	[1.007]
Male		−0.076***	−0.078***		−0.077***	−0.079***		−0.064***	−0.060***
		[0.008]	[0.008]		[0.007]	[0.007]		[0.009]	[0.009]
English test score		0.314***	0.313***		0.377***	0.370***		0.430***	0.423***
		[0.008]	[0.008]		[0.011]	[0.011]		[0.018]	[0.018]
Math test score		0.120***	0.122***		0.194***	0.202***		0.232***	0.224***
		[0.006]	[0.006]		[0.007]	[0.007]		[0.009]	[0.010]
CELDT Subtests		X	X		X	X		X	X
English GPA		X	X						
Parent's education FE		X	X		X	X		X	X
Grade FE		X	X		X	X		X	X
School FE			X			X			X
Observations	40,542	37,060	36,435	41,499	40,069	39,087	28,379	24,908	24,143
Number reclassified	3047	2693	2590	5253	5208	4981	2326	2051	1942
R-squared	0.061	0.286	0.315	0.041	0.274	0.293	0.043	0.302	0.318
F-statistic	750	682	697	647	684	664	134	149	144

Notes: For all specification the outcome variable of interest is 11th grade English test scores. The grade groups of 2nd–4th grade, 5th–7th grade, and 8th–10th grade each have three columns with a different specification in each of the three columns. For all specification the baseline bandwidth of 30 points is used. The first, fourth, and seventh columns are the IV results for each grade group using the two-stage least squares procedure from Eqs. (2) and (3) with no demographic or prior academic performance controls. The second, fifth, and eighth columns are the IV results for each grade group using Eqs. (2) and (3) including demographic and prior academic performance controls. The third, sixth, and ninth columns add school fixed effects. Robust standard errors are reported in brackets. *** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level.

indicate there are large English language benefits to marginal younger students when reclassified to mainstream English classes. Similar to Robinson's (2011) results I find no evidence of an English language benefit from being reclassified for marginal older students.

Table 5B reports the corresponding IV estimates of δ from each grade group for the outcome variable of standardized 11th grade math test scores. For all grade groups and specifications, there is no evidence of an effect on math scores from being reclassified for marginal students. The baseline specification used in columns 2, 5, and 7 in Table 5A and Table 5B are used for all the subsequent results unless otherwise specified. This baseline specification is used since it includes all available demographic and prior academic achievement controls and it allows for between school variation since the reclassification policy is at the school district level. Table 10 and Fig. 7 show the results for various specification choices.

Table 5B
IV Estimates for 11th grade Math test scores.

Variables	Math test scores in 11th grade								
	2nd–4th grade			5th–7th grade			8th–10th grade		
	No covariates	Demographic/ performance controls	School fixed effects	No covariates	Demographic/ performance controls	School fixed effects	No covariates	Demographic/ performance controls	School fixed effects
Reclassified	0.093	−0.009	−0.026	0.054	0.052	0.078	0.107	0.025	−0.040
	[0.098]	[0.096]	[0.094]	[0.070]	[0.066]	[0.065]	[0.178]	[0.159]	[0.158]
CELDT Overall	0.302***	0.273	0.230	0.210***	0.733	0.664	0.094*	0.143	0.597
	[0.037]	[0.765]	[0.749]	[0.037]	[0.509]	[0.495]	[0.054]	[0.768]	[0.748]
Male		0.056***	0.056***		0.051***	0.047***		0.022***	0.022***
		[0.007]	[0.006]		[0.005]	[0.005]		[0.007]	[0.006]
English test score		0.041***	0.037***		0.048***	0.035***		0.067***	0.070***
		[0.006]	[0.006]		[0.008]	[0.008]		[0.013]	[0.013]
Math test score		0.228***	0.226***		0.240***	0.237***		0.345***	0.302***
		[0.005]	[0.005]		[0.006]	[0.006]		[0.009]	[0.009]
CELDT subtests		X	X		X	X		X	X
English GPA		X	X						
Parent's Education FE		X	X		X	X		X	X
Grade FE		X	X		X	X		X	X
School FE			X			X			X
Observations	35,470	32,454	31,875	34,626	33,532	32,659	22,489	20,005	19,374
Number reclassified	2736	2414	2315	4584	4554	4345	1929	1749	1644
R-squared	0.020	0.189	0.249	0.019	0.152	0.211	0.009	0.229	0.281
F-statistic	680	619	632	546	582	560	123	127	121

Notes: For all specification the outcome variable of interest is 11th grade math test score. The grade groups of 2nd–4th grade, 5th–7th grade, and 8th–10th grade each have three columns with a different specification in each of the three columns. For all specification the baseline bandwidth of 30 points is used. The first, fourth, and seventh columns are the IV results for each grade group using the two-stage least squares procedure from Eqs. (2) and (3) with no demographic or prior academic performance controls. The second, fifth, and eighth columns are the IV results for each grade group using Eqs. (2) and (3) including demographic and prior academic performance controls. The third, sixth, and ninth columns add school fixed effects. Robust standard errors are reported in brackets. *** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level.

Table 6
IV estimates of English and Math test scores by years after reclassification.

Variables	2nd–4th grade		5th–7th grade		8th–10th grade	
	ELA	Math	ELA	Math	ELA	Math
1 year	0.163**	0.122	0.071	0.037	−0.150	0.076
	[0.076]	[0.090]	[0.060]	[0.052]	[0.184]	[0.154]
2 years	0.179**	0.063	0.106	0.061	0.205	−0.092
	[0.078]	[0.090]	[0.068]	[0.056]	[0.224]	[0.162]
3 years	0.223***	0.331***	−0.012	0.066	–	–
	[0.085]	[0.092]	[0.076]	[0.061]	–	–
4 years	0.270***	0.181**	-0.044	0.073	–	–
	[0.093]	[0.084]	[0.082]	[0.061]	–	–
5 years	0.136	0.070	–	–	–	–
	[0.103]	[0.092]	–	–	–	–
6 years	0.250**	0.063	–	–	–	–
	[0.112]	[0.097]	–	–	–	–
7 years	0.306***	0.085	–	–	–	–
	[0.119]	[0.098]	–	–	–	–
Max observations	77,984	77,967	60,261	60,030	31,057	28,061
Min observations	49,835	46,882	44,047	40,735	21,507	18,262

Notes: Each cell in the table is an IV coefficient on reclassified along with its robust standard error. Each cell uses the baseline specification from columns 2, 5, and 8 from Table 2A and Table 2B. This baseline specification includes demographic and prior academic performance controls and a bandwidth of 30 points. The columns represent grade groups and test subject, and the rows represent the number of years after reclassification in which the test is taken. For example the cell in the 2nd–4th grade English column and 1 year row is the baseline specification with English test scores one year after reclassification as the outcome variable for students near the cutoff in 2nd–4th grade. *** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level.

Table 6 and Fig. A.5 show the estimates for the dynamic effect of reclassification on students' test scores each year after reclassification. In Table 6, each cell contains the IV coefficient on the reclassified variable along with its robust standard error. For marginal 2nd to 4th grade students, being reclassified increases their English test scores by 0.163 standard deviations the year after being reclassified. This effect persists over the next six years and is largest 7 years after reclassification with an effect size of 0.306 standard deviations. The IV coefficients are statistically

Table 7
IV estimates of English test scores by grade group for boys and girls.

Variables	CST ELA score					
	2nd–4th grade		5th–7th grade		8th–10th grade	
	Boys	Girls	Boys	Girls	Boys	Girls
1 Year	0.308***	0.040	0.080	0.061	−0.395	0.037
	[0.116]	[0.100]	[0.103]	[0.071]	[0.330]	[0.208]
2 Years	0.364***	0.023	0.131	0.093	0.017	0.331
	[0.121]	[0.101]	[0.119]	[0.079]	[0.396]	[0.260]
3 Years	0.345***	0.119	0.028	−0.038	–	–
	[0.131]	[0.110]	[0.132]	[0.088]	–	–
4 Years	0.340**	0.217*	−0.019	−0.063	–	–
	[0.145]	[0.120]	[0.143]	[0.096]	–	–
5 Years	0.247	0.045	–	–	–	–
	[0.159]	[0.132]	–	–	–	–
6 Years	0.417**	0.111	–	–	–	–
	[0.175]	[0.144]	–	–	–	–
7 Years	0.433**	0.195	–	–	–	–
	[0.185]	[0.153]	–	–	–	–
Max observations	38,915	39,069	30,717	29,544	16,478	14,579
Min observations	24,815	25,020	22,351	21,696	11,365	10,142

Notes: Each cell in the table is an IV coefficient on reclassified along with its robust standard error. Each cell uses the baseline specification from columns 2, 5, and 8 from Table 2A. This baseline specification includes demographic and prior academic performance controls and a bandwidth of 30 points. For all cells the outcome variable of interest is English test scores. However, the number of years after reclassification, grade group, and gender of each cell differ. The columns represent grade groups and gender, and the rows represent the number years after reclassification in which the English test is taken. For example the cell in the 2nd to 4th grade boys' column and 1 year row is the baseline specification with boys' English test scores 1 year after reclassification as the outcome variable for students near the cutoff in 2nd–4th grade. *** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level.

significant for 6 of the 7 years. The effect of reclassification on math test scores for 2nd to 4th grade students near the cutoff are also positive for all 7 years. However, only 2 of the 7 estimates are statistically significant. For students in both 5th to 7th grade and 8th to 10th grade there are no statistically significant effects on either English or math test scores. The estimates from Table 6 suggest that there are large and persistent benefits to English test scores of being reclassified for younger students near the cutoff. However, there is no evidence of an effect on English test scores for marginal older students or an effect on math test scores for all marginal students.

In addition to estimating the effect of reclassification for the marginal student for different grades, I also estimate the effect by gender. I find that boys receive substantial benefits from reclassification whereas girls receive relatively little benefit. The results by gender are in Table 7 for English and Table A.6 for math. The effect of reclassification on English test scores for 2nd to 4th grade ranges from 0.247 to 0.433 standard deviations for boys and from 0.023 to 0.217 standard deviations for girls. For boys, 6 of the 7 years after reclassification estimates are statistically significant at the 5% level, whereas none of the estimates for girls are statistically significant. For the older grade groups all estimates are near zero and are not statistically significant.

Although it is unclear why young boys receive larger benefits from reclassification than young girls, there are a few possible explanations. First, reclassified marginal boys may have higher academic ability than reclassified marginal girls and therefore benefit more from reclassification. In these data and as has been documented in the education literature, on average girls tend to perform better on English tests while boys tend to perform better on math tests (Dee, 2007; Roland & Steven, 2010). Hence, conditional on students' CELDT scores and covariates, young boys who are reclassified score 0.231 standard deviations higher on math test scores than do young girls. However, since English test scores are used in the reclassification process young boys who are reclassified only score 0.040 standard deviations lower on English tests scores. Therefore, reclassified young boys have higher academic ability as measured by test scores than do young reclassified girls. Additional evidence that reclassified marginal boys may have higher ability is that young boys near the cutoff are 18% less likely to be reclassified than young girls, implying that boys who are reclassified may have higher ability. Second, there is some evidence that girls tend to have higher willingness to communicate in a second language (MacIntyre, Baker, Clément, & Donovan, 2002) and have a more positive attitude toward language learning (Wright, 1999; Zafar & Meenakshi, 2012). These possible gender differences may make the reclassification to mainstream English classes more helpful for boys and therefore more impactful.

Besides the effect of reclassification on English and math test scores, other school outcomes are examined. Table 8 reports the effect of reclassification for marginal students on 6th through 12th grade English and math GPA (GPA is only available for these grades in the data). The results for GPA are similar to the results for test scores. For students near the cutoff in 2nd to 4th grade, being reclassified increases students' 6th grade English GPA by 0.334 points, or on average from a B to a B+. It also has a positive but not statistically significant effect on math GPA. Reclassification has a statistically significant positive effect on

Table 8
IV estimates of GPA.

Variables	GPA					
	2nd–4th grade		5th–7th grade		8th–10th grade	
	English	Math	English	Math	English	Math
6th grade	0.334**	0.187	–	–	–	–
	[0.139]	[0.149]	–	–	–	–
7th grade	0.510***	0.269	−0.195	−0.291**	–	–
	[0.161]	[0.165]	[0.126]	[0.133]	–	–
8th grade	0.063	0.122	−0.058	−0.115	–	–
	[0.154]	[0.157]	[0.127]	[0.130]	–	–
9th grade	0.323	−0.042	−0.278*	−0.118	–	–
	[0.228]	[0.232]	[0.154]	[0.145]	–	–
10th grade	–	–	−0.314**	−0.036	0.195	0.597
	–	–	[0.155]	[0.145]	[0.422]	[0.376]
11th grade	–	–	0.020	0.244	−0.614	−0.423
	–	–	[0.169]	[0.165]	[0.390]	[0.362]
12th grade	–	–	−0.115	−0.369	0.208	0.189
	–	–	[0.246]	[0.341]	[0.373]	[0.485]
Max observations	58,398	63,383	52,682	57,558	27,587	25,838
Min observations	26,077	26,128	17,818	11,617	22,175	15,164

Notes: Each cell in the table is an IV coefficient on reclassified along with its robust standard error. Each cell uses the baseline specification from columns 2, 5, and 8 from Table 2A. This baseline specification includes demographic and prior academic performance controls and a bandwidth of 30 points. For all cells the outcome variable of interest is GPA. However, the GPA grade, grade group, and subject (English or math) of each cell differ. The columns represent grade groups and subject, and the rows represent the grade in which the GPA is received. For example the cell in the 2nd–4th grade English column and 6th grade row is the baseline specification with English GPA in grade 6 as the outcome variable for students near the cutoff in 2nd–4th grade. Both English and Math GPA are measured on a scale from 0 to 4. *** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level.

two of the four future years of English GPA. However, for 5th to 7th grade and 8th to 10th grade the effect tends to be negative. These results seem to be in line with the test score results that younger students have large gains in English and minor to no gains in math. Similarly, there is no evidence that reclassification affects marginal older students' English or math GPA.

The transition from ELD classes to mainstream classes can potentially be difficult due to changes in language, culture, friends, and other features of the school environment. Besides just affecting academic achievement, the effect of reclassification on the social aspect of school may also lead to changes in school attendance and being retained a year in school. The first three columns in Table 9 shows estimates of the baseline specification for whether a student is retained a year in school by each grade group. Columns 4 to 6 estimate the effect of reclassification on the fraction of school days absent by grade group. For each grade group and in each future grade, I find no evidence of an effect from reclassification on retention. Similarly, I find no evidence of an effect from being reclassified on the fraction of days absent from school. These results indicate that there is no evidence that any difficulties that might arise from reclassification for marginal students affects students' decision to attend school or their likelihood of being held back a year in school.

One possible problem and effect of reclassification is reclassification's effect on attrition from the sample or in other words its effect on students leaving the LAUSD. I test to see if attrition rates differ for ELLs who were reclassified versus not reclassified. Fig. 6 shows attrition rates for ELLs who were reclassified and not reclassified up to 7 years after the reclassification decision was made. In order for the data to identify whether an ELL was reclassified or not he must enroll in the LAUSD the next year; therefore, the fraction of reclassified students and not reclassified students enrolled in the LAUSD the next year is one. The attrition rate for reclassified and not reclassified students is similar. To verify this similarity, the baseline specification was estimated with the binary outcome variable equal to one if the student was enrolled in the LAUSD a given number of years after the reclassification decision was made. For all future years after the first one and for all three grade groups the effect of reclassification on attrition from the LAUSD is close to zero and not statistically significant (see Table A.14).

To check the robustness of the results, I use ten different specifications and report the results in Table 10. Each row of the table indicates a different specification and each column indicates a grade group and test subject. The first, second and fourth rows correspond with the second, first and third columns from Table 5A and 5B. The third row includes demographic controls, but excludes any prior achievement controls. The fifth row is the same as the baseline specification, but includes a fourth degree polynomial for the running variable, the CELDT score. The sixth row also uses the baseline specification, but uses a fourth degree polynomial for English and math test scores (all grade groups) and English GPA (2nd to 4th grade group). The seventh and ninth rows use the baseline specification with a bandwidth of 40 and 20 points respectively. The eighth and tenth rows are the same as the sixth and eighth rows respectively, except they also use a fourth degree polynomial for prior achievement. For all specifications the reclassified coefficient for the 2nd to 4th grade English test scores is similar and statistically significant, whereas the coefficient for the 2nd to 4th grade math test scores are all similarly close to zero with no estimates being statistically significant. For all specifications the reclassified coefficients for the 5th to 7th and 8th to 10th grade

Table 9
IV estimates of retained and fraction of days absent.

Variables	Retained			Fraction of days absent		
	2nd–4th grade	5th–7th grade	8th–10th grade	2nd–4th grade	5th–7th grade	8th–10th grade
4th grade	−0.004	–	–	–	–	–
	[0.009]	–	–	–	–	–
5th grade	0.002	–	–	–	–	–
	[0.003]	–	–	–	–	–
6th grade	0.000	0.004	–	0.002	0.001	–
	[0.003]	[0.006]	–	[0.008]	[0.015]	–
7th grade	0.002	0.000	–	0.003	−0.009	–
	[0.003]	[0.002]	–	[0.009]	[0.008]	–
8th grade	0.002	0.001	–	0.002	−0.001	–
	[0.004]	[0.004]	–	[0.010]	[0.010]	–
9th grade	–	−0.054	–	−0.008	0.011	–
	–	[0.040]	–	[0.015]	[0.014]	–
10th grade	–	0.035	−0.020	−0.009	0.004	−0.016
	–	[0.028]	[0.084]	[0.016]	[0.016]	[0.041]
11th grade	–	0.024	0.057	−0.010	0.033*	0.055
	–	[0.025]	[0.048]	[0.016]	[0.017]	[0.040]
12th grade	–	–	–	0.040*	0.011	0.018
	–	–	–	[0.022]	[0.017]	[0.043]

Notes: Each cell in the table is an IV coefficient on reclassified along with its robust standard error. Each cell uses the baseline specification from columns 2, 5, and 8 from Table 2A. This baseline specification includes demographic and prior academic performance controls and a bandwidth of 30 points. The first three columns are IV estimates of the effect of reclassification on the binary variable of being retained a year in school for different grade groups. Columns 4 through 6 are IV estimates of the effect of reclassification on the fraction of school days absent for different grade groups. The columns represent grade groups and either retained or fraction of day absent, and the rows represent the grade in which the outcome variable is observed. For example the cell in the 2nd–4th grade retained column and 6th grade row is the baseline specification with retained in grade 6 as the outcome variable for students near the cutoff in 2nd–4th grade. *** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level.

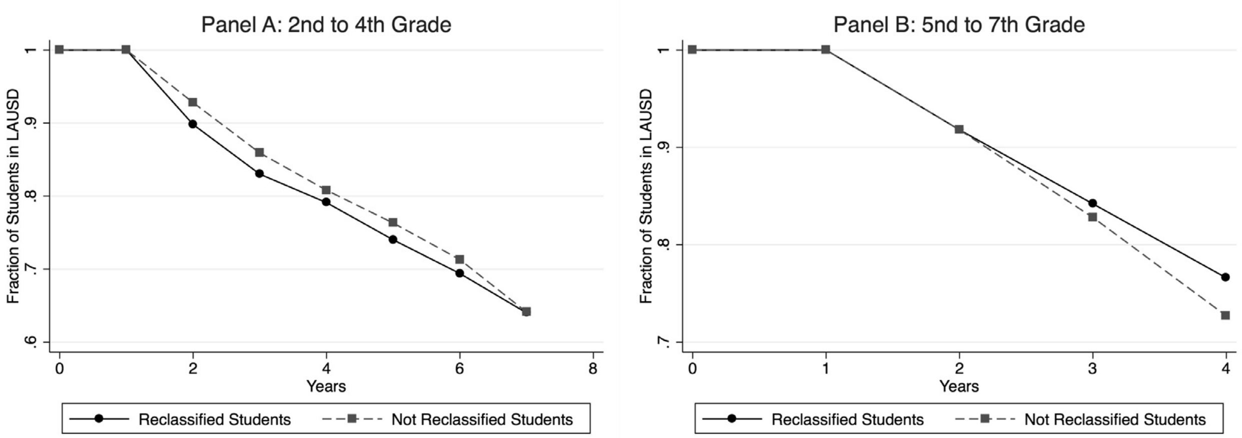

Fig. 6. Attrition rates over times by grade group.
Notes: Panel A is for students in 2nd–4th grade and Panel B is for students in 5th–7th grade in the years 2003 and 2004. The *y*-axis represents the fraction of students in the sample still enrolled in the LAUSD. The *x*-axis represents the number years after the year the student was either reclassified or not reclassified. The black line is for students who were reclassified in the years 2003 and 2004. The red line is for students who remained in ELD programs in the years 2003 and 2004. In order to be reclassified a student must attend a LAUSD school the following year, therefore the faction of reclassified students in the LAUSD in the first year is one. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

English and math test scores are statistically indistinguishable from zero.

For an additional robustness check, Fig. 7 graphs the reclassified coefficient from the baseline specification for bandwidths that vary from 5 to 50 points (0.08 to 0.83 standard deviations) for both the 2nd to 4th and 5th to 7th grade groups and for both English and math test scores. For each grade group and test subject the reclassified coefficient is relatively constant over the bandwidth choice. Also, the confidence intervals narrow as the bandwidth choice increases and more individuals on each side of the cutoff are included.

3.3. Policy implications

The kindergarten results indicate that ELD programs are effective for the students near the cutoff. These results fit with the conclusion that ELD programs are in general

Table 10
Robustness check.

Specification	Test scores in 11th grade					
	2nd–4th grade		5th–7th grade		8th–10th grade	
	English	Math	English	Math	English	Math
Baseline	0.297**	−0.009	−0.056	0.052	−0.181	0.025
	[0.119]	[0.096]	[0.090]	[0.066]	[0.221]	[0.159]
No covariates	0.331***	0.093	−0.059	0.054	0.014	0.107
	[0.127]	[0.098]	[0.102]	[0.070]	[0.263]	[0.178]
Demographic covariates	0.337***	0.102	−0.060	0.061	−0.080	0.038
	[0.125]	[0.097]	[0.101]	[0.069]	[0.262]	[0.174]
School fixed effects	0.272**	−0.026	−0.051	0.078	−0.133	−0.04
	[0.119]	[0.094]	[0.091]	[0.065]	[0.224]	[0.158]
4th degree polynomial for CELDT score	0.296*	−0.043	−0.100	0.085	−0.206	−0.041
	[0.178]	[0.141]	[0.123]	[0.093]	[0.275]	[0.192]
4th degree polynomial for achievement	0.244**	−0.045	−0.074	0.056	−0.204	0.012
	[0.124]	[0.100]	[0.092]	[0.068]	[0.223]	[0.155]
Bandwidth of 40	0.259**	−0.008	−0.060	0.045	−0.189	0.026
	[0.112]	[0.091]	[0.081]	[0.059]	[0.211]	[0.145]
Bandwidth of 40 and 4th degree polynomial for prior performance	0.254**	−0.021	−0.058	0.043	−0.180	0.022
	[0.111]	[0.090]	[0.080]	[0.059]	[0.206]	[0.144]
Bandwidth of 20	0.301**	−0.048	−0.119	0.052	−0.211	−0.034
	[0.153]	[0.122]	[0.113]	[0.084]	[0.257]	[0.183]
Bandwidth of 20 and 4th Degree polynomial for prior performance	0.289*	−0.050	−0.130	0.048	−0.284	−0.071
	[0.154]	[0.122]	[0.114]	[0.085]	[0.263]	[0.183]

Notes: Each cell in the table is an IV coefficient on reclassified along with its robust standard error. Each row represents a different specification and each column represents a different outcome variable of interest. The first row is the baseline specification from columns 2, 5, and 8 from Tables 2A and 2B. The second row is the no covariate specification from columns 1, 4, and 7 from Tables 2A and 2B. The third row just includes the demographic controls. The fourth row is the baseline specification plus school fixed effects. The fifth row uses a 4th degree polynomial for the running variable CELDT score. The sixth row uses a 4th degree polynomial to allow for more flexibility in prior achievement. The seventh row is the baseline specification with a bandwidth of 40 points. The eighth row is a combination of the fifth and sixth row. The ninth row is the baseline specification with a bandwidth of 20 points. The tenth row is a combination of the fifth and eighth row. *** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level.

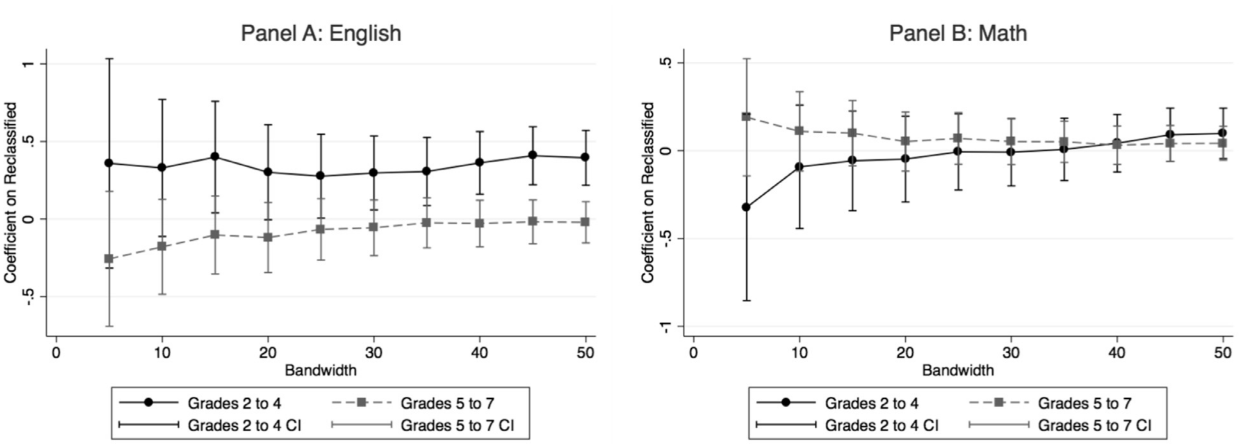

Fig. 7. IV Estimate of reclassification on 11th grade test scores over bandwidth.
Notes: These two figures show the robustness of the estimates to the choice of bandwidth. The baseline specification in the demographic and achievement controls column in Tables 2A and 2B is used with varying bandwidths. Bandwidths vary from 5 to 50 points. Confidence intervals are created using robust standard errors. The black line represents 2nd–4th grade and the red represents 5th–7th grade. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

effective in increasing the English achievement of non-native English students. In fact, the results suggest academic gains could be obtained by slightly expanding ELD programs to include more students. The results for reclassifying 2nd to 10th grade students indicate that there are large gains to be had by reclassifying the marginal student in earlier grades and allowing more students to exit ELD programs sooner. These results suggest that a slight increase in the inflow of students into ELD programs and a quicker outflow of younger students from ELD programs could create gains for marginal students.

The regression discontinuity design only produces estimates for the marginal student, and does not estimate the overall effect of ELL policies at the LAUSD. However, considering the costs and benefits of increasing the inflow and outflow of students in ELD programs can give a better understanding of the overall policy effect. For this cost benefit comparison there are two main benefits, and one possible cost.

From the results of this paper, it is shown that the marginal students' English academic achievement can be improved by increasing the inflow and outflow of students in ELD programs. The second benefit of increasing program inflow and outflow is its reduction on education expenditures. In California, per student expenditures are roughly 6–9% higher for ELLs than mainstream English students (Jepsen & Alth, 2005). However, both of these benefits are only applicable to the marginal ELL, which is at most 5% of the school district's students.

Although there are potential benefits to increasing the inflow and outflow of students in ELD programs, there is also a possible cost through changes in classroom composition and peer effects. Much of the literature finds little to no impact of additional non-native English speakers in a classroom on native speakers' academic achievement. Geay, McNally, and Telhaj (2013) find no causal impact of changes in the percentage of non-native speakers on educational outcomes. Ohinata and van Ours (2013) also find little evidence of a negative spill-over effect on academic performance from the increased presence of immigrant children. However, some research has shown negative spill-over effects on native speakers' academic achievement (Ahn & Jepsen, 2015; Diette & Oyelere, 2012). When a marginal ELL is reclassified or when a marginal kindergartener is placed in mainstream English class, this removes a high performing ELL from ELD classes and adds a relatively low performing student to mainstream English classes. ELLs near the cutoff have 0.104 and 0.088 standard deviations higher math and English test scores, respectively, than do average ELLs. The average mainstream student has 0.0979 and 0.4973 standard deviations higher math and English test scores respectively than do ELLs near the cutoff. This change lowers the average student achievement in both ELD classes and mainstream English classes. With class sizes of 20, reclassifying an ELL near the cutoff decreases the average math and English test scores of the ELD class he left by 0.0052 and 0.0044 standard deviations respectively. This reclassification would also decrease the average math and English test scores of the mainstream class he enters by 0.0049 and 0.0238 standard deviations respectively.

This decrease in the average student performance in both classes could have a negative effect on all students in ELL and mainstream English classes through peer effects, although the existence and size of these peer effects are debated (Sacerdote, 2011). Using estimates from the peer effect literature, the large benefit obtained by reclassifying a young marginal ELL is as much as 5–10 times larger than the negative effect on peers (Carrell, Sacerdote, & West, 2013; Burke and Sass, 2013; Lefgren, 2004). Under a tracking or boutique model the reclassification of an ELL near the cutoff could have no effect or even a positive effect on other ELLs and mainstream students (Dobbelsteen, Levin, & Oosterbeek, 2002).

4. Explanations

There are four potential explanations for why large benefits exist when young marginal ELLs are reclassified. First, the benefits could be directly due to more exposure to English during class when reclassified. Because teachers and students speak more and higher quality English in mainstream English classes than in ELD classes, this exposure could be the mechanism through which the academic gains are obtained. This mechanism fits with the fact that young marginal ELLs receive large benefits to English achievement and small to no benefit to math achievement when reclassified.

Similarly, the benefits from being reclassified could come from developing more and closer friendships with native English speakers when placed in mainstream English classes. When students are reclassified to mainstream classes they spend a much higher percentage of their class time with native English speaking students. This additional time spent with native English speakers is likely to lead to more and closer friendships with native English speakers (Suárez-Orozco, Suárez-Orozco, & Todorova, 2009). These friendships with native English speakers could play a key role in reclassified ELL students' English development (Carhill, Suárez-Orozco, & Páez, 2008; Zafar, 2001). Also, since children at younger ages are more likely to create friendships outside of their demographic clique (Killen, Lee-Kim, McGlothlin, Stangor, & Helwig, 2002, Aboud, 1988; Parker, Rubin, Erath, Wojslawowicz, & Buskirk, 2006), this could possibly explain why younger children receive larger benefits from reclassification.

Alternatively, the benefits from being reclassified could come from changes in teacher quality. If teaching quality was higher in mainstream classes than ELD classes due to a limited supply of qualified ELD teachers, then reclassification could increase reclassified ELLs' achievement. For this to be a plausible explanation, achievement gains should also appear in math as well as English, but they do not. Additionally, the teacher quality explanation would also imply that marginal kindergarteners should perform better when placed in mainstream classes, which is inconsistent with my results.

Lastly, peer effects could explain the benefits. As discussed above, the average achievement of students in mainstream English classes is higher than that of students in ELD classes. Therefore, moving from ELL to mainstream classes could positively affect marginal ELLs' achievement. However, similar to the teacher quality explanation, achievement gains should also occur in math, but do not. Also, the peer effects story cannot explain both the negative kindergarten effect and positive 2nd to 4th grade effect.

Although these explanations may explain why there might be gains from reclassification, they do not necessarily explain why young ELLs benefit from reclassification while there is no evidence older ELLs benefit. An additional consideration is why marginal students in younger grades experience large gains to reclassification whereas there is no evidence that students in older grades experience gains. One possible explanation is that younger children may have more language acquisition capability than older children. Another possible explanation is that reclassification rates are arbitrarily lower for younger grades than older grades and students with higher English proficiency benefit more from reclassification. These explanations are not mutually exclusive and may each play a role in explaining the difference in the effect of reclassification for younger and older grades.

There is a literature that finds language acquisition capability decreases with age (Bialystok, Hakuta, & Wiley, 2003; Hyltenstam & Abrahamsson, 2003). Meisel (2008) indicates that bilingual acquisition capability peaks around the age of two, with gradual declines after the age of five. However, others argued that older children are more efficient second language learners because they have a larger linguistic base on which to build (Bialystok & Hakuta, 1999; Marinova-Todd, Bradford Marshall, & Snow, 2000). More broadly, Cunha and Heckman (2009) find that education in later grades plays only a minor role in creating or reducing academic achievement gaps.

Reclassification moves students to classrooms with more native speaking English students. The assumption of a steady decline in bilingual acquisition capability as age increases, along with the increased exposure to native English speakers leads to the hypothesis that there should be steady decline in the effect of reclassification on the marginal student near the cutoff. The results show a steady and steep decline in the effect of reclassification from 2nd to 4th grade (ages 8–10), but then the effect of reclassification levels out with an effect size near zero (see Table A.2 and Table A.4). If the decline in language acquisition capability was much steeper for ages 8–10 than for ages 11–16 then the results could be consistent with the explanation that differing language acquisition capabilities at different ages is the cause for differing effects of reclassification for younger and older grades. However, if the decline in language acquisition capability is constant over the grades, there should be just as large of an effect size difference between 6th and 9th grade students as there is between 2nd and 5th grade students. Differing language acquisition capability is a plausible explanation for at least some of the difference, but requires steeper declines at younger ages than older ages to be the entire explanation.

The other explanation discussed is that reclassification rates are arbitrarily lower for younger grades than older grades and students with higher levels of English proficiency benefit more from reclassification. Fig. 8 shows the fraction of students reclassified for each grade for all ELLs, ELLs above the cutoff, and ELLs 20 points or more above the cutoff. Reclassification rates start low for younger grades, increase until they peak around 6th grade, decrease, and then plateau through 8th to 10th grade. This pattern is also observed for the reclassification rates for ELLs above the cutoff and ELLs 20 points or more above the cutoff.

The fact that all kindergarteners above the cutoff are placed in mainstream English classes and therefore all ELLs start school below the reclassification cutoff likely explains why the reclassification rate of all ELLs is lower for the younger grades than older grades. However, the reclassification rates conditional on being above the cutoff are also lower for younger than older grades through 6th grade. With similar reclassification policies for each grade it would be expected that the reclassification rate conditional on being above the cutoff would be the same for all grades. School language appraisal teams may use their discretion to arbitrarily reclassify younger ELLs above the cutoff less often than older ELLs above the cutoff. This arbitrarily lower reclassification rate of younger ELLs above the cutoff could cause the marginal younger ELL to have a higher English proficiency than the marginal older ELL and therefore reclassification could benefit the younger ELL more. Since these explanations are not mutually exclusive, the differing effect of reclassification for younger and older grades is likely not attributed to a single cause.

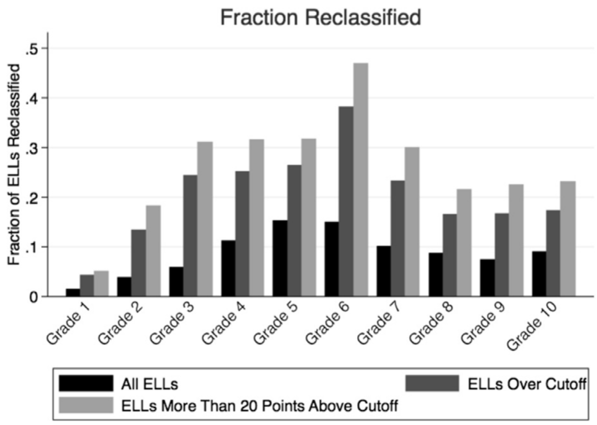

Fig. 8. Fraction of ELLS reclassified by grade.
Notes: The *y*-axis is the fraction of ELLs who are reclassified into mainstream English classes. The *x*-axis is each grade. The blue bars include all ELLs regardless of the distance of their overall CELDT score from the cutoff. The red bars only include ELLs who have overall CELDT scores above the cutoff and the green bar are for ELLs with an overall CELDT score 20 points or more above the cutoff. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

5. Conclusion

The growing number of students in which English is a second language makes it increasingly important to find ways to better meet their academic needs. Using a regression discontinuity design to test the effect of the classification and reclassification of ELLs, I find that marginal incoming kindergarteners benefit slightly from being placed into ELD programs. This suggests that small gains could be achieved by increasing the inflow of students into these programs. I also find that moving marginal 2nd to 4th grade students into mainstream classes causes large gains in their English test scores and English GPA. These gains are persistent over the 7 years following reclassification and are much larger for boys. I find no evidence that such gains exist for math test scores, math GPA, retention, or school attendance. I also find no evidence that gains exist for older students in 5th to 10th grade. It is likely that achievement gains could be obtained by increasing the inflow and early outflow in ELD programs.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.econedurev.2016.04.009.

References

Aboud, F. E. (1988). *Children and prejudice. The development of ethnic awareness and identity*. Oxford UK: Basil Blackwell.

Ahn, T., & Jepsen, C. (2015). The effect of sharing a mother tongue with peers: Evidence from North Carolina middle schools. *IZA Journal of Migration, 4*(1), 1–21.

Aud, S. (2012).et al. "The condition of education 2012." National Center for Education Statistics. Available at: http://nces.ed.gov/pubs2012/2012045.pdf

August, D. (2002). *English as a second language instruction: Best practices to support the development of literacy for English language learners*. Baltimore: Johns Hopkins.

Bialystok, E., & Hakuta, K. (1999). Confounded age: Linguistic and cognitive factors in age differences for second language acquisition. *Second language acquisition and the critical period hypothesis*, Edited by David Birdsong: 161–181.

Burke, M., & Sass, T. (2013). Classroom peer effects and student achievement. *Journal of Labor Economics, 31*(1), 51–82.

Callahan, R. M. (2005). Tracking and high school English learners: Limiting opportunity to learn. *American Educational Research Journal, 42*(2), 305–328.

Carhill, A., Suárez-Orozco, C., & Páez, M. (2008). Explaining English language proficiency among adolescent immigrant students. *American Educational Research Journal, 45*(4), 1155–1179.

Carrell, S., Sacerdote, B., & West, J. (2013). From natural variation to optimal policy? The Importance of endogenous peer group formation. *Econometrica, 81*(3), 855–882.

Cheung, A., & Slavin, R. (2005). A synthesis of research on language of reading instruction for English language learners. *Review of Education Research, 75*, 247–284.

Cunha, F., & Heckman, J. (2009). The economics and psychology of inequality and human development. *Journal of the European Economic Association, 7*, 320–364.

Dee, T. (2007). Teachers and the gender gaps in student achievement. *Journal of Human Resources, 42*(3), 528–554.

Diette, T., & Oyelere, R. (2012). "Do significant immigrant inflows create negative education impacts? Lessons from the North Carolina Public School System" IZA discussion paper 6561.

Dobbelsteen, S., Levin, J., & Oosterbeek, H. (2002). The causal effect of class size on scholastic achievement: distinguishing the pure class size effect from the effect of changes in class composition. *Oxford Bulletin of Economics and Statistics, 64*(1), 17–38.

Eicher, T. S., & Garcia-Penalosa, C. (2001). Inequality and growth: The dual role of human capital in development. *Journal of Development Economics, 66*(1), 173–197.

Fryer, R., & Levitt, S. (2004). Understanding the black-white test score gap in the first two years of school. *Review of Economics and Statistics, 86*(2), 447–464.

Geay, C., McNally, S., & Telhaj, S. (2013). Non-native speakers of English in the classroom: What are the effects on pupil performance? *Economic Journal, 123*(570), F281–F307.

Hahn, J., Petra, T., & Van derKlaauw, W. (2001). Identification and estimation of treatment effects with a regression discontinuity design. *Econometrica, 69*, 201–209.

Hakuta, K., Bialystok, E., & Wiley, E. (2003). Critical evidence: A test of the critical period hypothesis for second language acquisition. *Psychological Science, 14*, 31–38.

Hyltenstam, K., & Abrahamsson, N. (2003). Maturational constraints in SLA. In C. J. Doughty, & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 539–588). Malden, MA: Blackwell.

Imbens, G., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics, 142*, 615–635.

Jepsen, C., & De Alth, S. (2005). *English learners in California schools*. Public Policy Institute of California.

Killen, M., Lee-Kim, J., McGlothlin, H., Stangor, C., & Helwig, C. C. (2002). How children and adolescents evaluate gender and racial exclusion. *Monographs of the society for research in child development, 67*(4), 1–129.

Lee, J. (2002). Racial and ethnic achievement gap trends: Reversing the progress toward equity? *Educational Researcher, 31*(1), 3–12.

Lefgren, L. (2004). Educational peer effects and the Chicago public schools. *Journal of Urban Economics, 56*, 169–191.

MacIntyre, P. D., Baker, S. C., Clément, R., & Donovan, L. A. (2002). Sex and age effects on willingness to communicate, anxiety, perceived competence, and L2 motivation among junior high school French immersion students. *Language Learning, 52*(3), 537–564.

Marinova-Todd, S., Bradford M.D., & Snow, C. (2000). Three misconceptions about age and L2 learning. *TESOL quarterly*, 9–34.

Matsudaira, J. (2005). Sinking or swimming? Evaluating the impact of English immersion versus bilingual education. In *Paper presented at the annual conference of the Association for Public Policy Analysis and Management*, Washington, DC.

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

Meisel, J. M. (2008). The bilingual child. In T. K. Bhatia, & W. C. Ritchie (Eds.), *The handbook of bilingualism* (pp. 91–113) (Eds.). Oxford: Blackwell.

Mincer, J. (1958). Investment in human capital and personal income distribution. *Journal of Political Economy, 66*, 281–302.

Ohinata, A., & van Ours, J. C. (2013). How immigrant children affect the academic achievement of native Dutch children. *Economic Journal, 123*(570), F308–F331.

Parker, J., Rubin, K., Erath, S., Wojslawowicz, J., & Buskirk, A. (2006). Peer relationships, child development, and adjustment: A developmental psychopathology perspective. *Developmental psychopathology, theory and method*, Edited by Dante Cicchetti and Donald Cohen: 419–477.

Restuccia, D., & Urrutia, C. (2004). Intergenerational persistence of earnings: The role of early and college education. *American Economic Review, 94*, 1354–1378.

Rossell, C. H., & Baker, K. (1996). The educational effectiveness of bilingual education. *Research in the Teaching of English, 30*(1), 7–69.

Robinson, J. P. (2011). Evaluating criteria for English learner reclassification: A causal-effects approach using a binding-score regression discontinuity design with instrumental variables. *Educational Evaluation and Policy Analysis, 33*(3), 267–292.

Robinson-Cimpian, J. P., & Thompson, K. D. (2015). The effects of changing test-based policies for reclassifying English learners. *Journal of Policy Analysis and Management, 35*(2), 279–305.

Sacerdote, B. (2011). Peer effects in education: How might they work, how big are they and how much do we know thus far? *Handbook of the Economics of Education, 3*, 249–277.

Slama, R. B. (2012). A longitudinal analysis of academic English proficiency outcomes for adolescent English language learners in the United States. *Journal of Educational Psychology, 104*(2), 265.

Suárez-Orozco, C., Suárez-Orozco, M., & Todorova, I. (2009). *Learning a new land: Immigrant students in American society*. Harvard University Press.

Thomas, W., & Collier, V. (2002). *A national study of school effectiveness for language minority students' long-term academic achievement*. Santa Cruz, CA: University of California at Santa Cruz, Center for Research on Education, Diversity, and Excellence.

Umansky, I., & Reardon, S. (2014). Reclassification patterns among Latino English learner students in bilingual, dual immersion, and English immersion classrooms. *American Educational Research Journal, 51*(5), 879–912.

Valentino, R., & Reardon, S. (2015). Effectiveness of four instructional programs designed to serve English learners variation by ethnicity and initial English proficiency. *Educational Evaluation and Policy Analysis*, 0162373715573310.

Willig, A. (1985). A meta-analysis of selected studies on the effectiveness of bilingual education. *Review of Educational Research, 55*(3), 269–317.

Wright, M. (1999). Influences on learner attitudes towards foreign language and culture. *Educational Research, 41*(2), 197–208.

Zafar, S., & Meenakshi, K. (2012). individual learner differences and second language acquisition: A review. *Journal of Language Teaching and Research, 3*(4), 639–646.