

EDUCATION POLICY ANALYSIS ARCHIVES

A peer-reviewed scholarly journal
Editor: Gene V Glass
College of Education
Arizona State University

Copyright is retained by the first or sole author, who grants right of first publication to the **Education Policy Analysis Archives**. EPAA is a project of the Education Policy Studies Laboratory. Articles are indexed in the Directory of Open Access Journals (www.doaj.org).

Volume 12 Number 36

July 30, 2004

ISSN 1068-2341

Reclassification of English Learners

James B. Grissom

California Department of Education

Citation: Grissom, J. B. (2004, July 30). Reclassification of English learners, *Education Policy Analysis Archives*, 12(36). Retrieved [date] from <http://epaa.asu.edu/epaa/v12n36/>.

Abstract

Ron Unz, originator of Proposition 227, claimed, prior to the passage of Prop. 227, that the five percent annual reclassification rate of English learners to fluent English proficient indicated bilingual education was a failure. Critics of Prop. 227 have countered that the annual reclassification rate has changed little since the passage of Prop. 227, indicating the new legislation had no effect on reclassification rates. Unfortunately, the annual reclassification rate does not provide a clear indicator of how long it takes students to be reclassified after entering the school system. To better estimate reclassification rates for English learners in California, cohorts were created to track the same groups of students over time. Ron Unz also claimed that test scores for immigrant students improved dramatically after the passage of Prop. 227. To evaluate his claim, average test scores were calculated by language fluency. Based on statewide data from three different cohorts tracked across four years, Prop. 227 has had no effect on reclassification rates or test scores.

Introduction

Ron Unz, originator of Proposition 227, stated, prior to the passage of Prop. 227, that a five percent annual reclassification rate of English learners (EL) to fluent English proficient (R-FEP) in California implied to him a failure rate for bilingual education of 95 percent (Unz, 1997). Critics of Prop. 227 (Crawford, 2003; Hakuta, 2002; Mora, 2000) have challenged Unz's statements about EL reclassification two ways. First, they present evidence that reclassification rates, available from the California Department of Education (CDE) were closer to seven than five percent and were rising prior to the passage of Prop. 227 (CDE, 2004). Second, the annual reclassification rate, since the passage of Prop. 227, has stabilized around eight percent, indicating that Prop. 227 has had little or no effect on reclassification rates. In addition, critics of Prop. 227 have emphasized that less than 30 percent of EL students were enrolled in bilingual programs prior to the passage of Prop. 227 (Gandara, 2000). As such, annual reclassification rates could not be interpreted as evidence that bilingual education programs were failing since more than 70 percent of EL students were not in bilingual programs. Although Unz has claimed Prop. 227 a success, he has been quiet about its effect on reclassification rates.

It is the contention of this study that the reclassification rates cited by Unz and his critics are misleading in two ways. First, the data upon which these reclassification rates are based do not account for students moving into and out of the California school system. The EL student population is not stable. It is increasing each year (CDE, 2004). When there are more EL students entering the school system than leaving, the denominator is inflated and the proportion of students who have been reclassified (i.e., the number of reclassified students divided by the number of EL students) is underestimated. Second, the reported reclassification rates are simply the proportion of EL students who have been reclassified in a particular year. The rates do not provide an indicator of how long it takes students to be reclassified after they have enrolled in the California school system.

According to Unz, most EL students can learn English in just a few months (*Ron Unz Exposes*, 2001) and so, the EL designation should not last much longer than a year. The language of Prop. 227, now part of California Law's Education Code (EC), reflects this philosophy.

Children who are English learners shall be educated through sheltered English immersion during a temporary transition period not normally intended to exceed one year... Once English learners have acquired a good working knowledge of English, they shall be transferred to English language mainstream classrooms (EC, Section 305).

The notion that EL students can learn English in just a few months has been called into question by researchers in language development. Hakuta, Butler, & Witt (2000) reported that English oral proficiency takes 3 to 5 years to develop and academic proficiency takes 4 to 7 years. They considered academic proficiency to be academic success in an English speaking classroom. This seems to be a tautology because the number of years to achieve academic proficiency was based on the length of time it took to reclassify students.

Reclassifying students from EL to R-FEP status is a process that uses multiple criteria (EC, Sec. 313), which include:

- 1) Assessment of English language proficiency
- 2) Teacher evaluation
- 3) Parent opinion and consultation
- 4) Comparison of performance in basic skills

The intent of using multiple criteria is to protect EL students from being reclassified before they are ready. It is thought if students are reclassified before they have achieved academic language skills or content-area knowledge and abilities they are at risk of academic failure.

The first reclassification criterion, language proficiency, is determined by an English language proficiency test. English language proficiency tests are designed to measure students' communication, reading, and writing skills in English. In May 2001 all Local Education Agencies (LEAs) were mandated by law (EC, Sec. 313) to use the California English Language Development Test (CELDT) to evaluate the English language proficiency of students whose home language is other than English. Prior to this date, LEAs were free to select from a list of CDE-approved English language development tests.

The other crucial reclassification criterion is the assessment of basic skills. Scores on a standardized achievement test are used to evaluate basic skills. In September 2002, all LEAs were advised to use the California Standards Test (CST) to evaluate the proficiency of EL students in basic skills. Prior to this date, LEAs had discretion in determining academic proficiency. It was common for districts to require EL students to score at or above the 36th percentile on one or more portions of the statewide norm-referenced test (NRT), the Stanford Achievement Test version 9 (SAT/9), form T, to be reclassified. However, proficiency could be defined as higher or lower than the 36th percentile.

Academic proficiency as defined by Hakuta, Butler, & Witt (2000) is a tautology because the length of time to achieve academic proficiency was based on the length of time it took students to be reclassified, and reclassification depends on academic performance. School districts report the biggest barrier to reclassification was not English proficiency but academic proficiency (Parrish, Linquanti, Merickel, Quick, Laird, & Esra, 2002). That is, students might be English fluent, based on results from an English language proficiency test, but would not be reclassified R-FEP because they could not meet the threshold (e.g., the 36th percentile) on a standardized achievement test. As a result, it could not be known if students would be able to demonstrate academic proficiency in the classroom if they only had to demonstrate proficiency in English to be reclassified.

Whatever length of time it takes EL students to be academically proficient, Hakuta, Butler, & Witt (2000) argued that

linguistic competence is complex, and even the most privileged second language learners take a significant amount of time to attain mastery, especially for the level of language required for school success.

Given that reclassification rates have been used by proponents of Prop. 227 to support its passage and opponents to criticize its effectiveness, there should be interest in how long it takes students to be reclassified. Toward that end, the purpose of this study is to track three different cohorts of EL students over a span of time in order to calculate the proportion of EL students reclassified R-FEP during this span.

Although Unz has been quiet about the effect of 227 on reclassification rates, Unz claims that test scores for EL students have improved dramatically since the passage of Prop. 227. Unz's claims are based on an initial CDE achievement test report in which EL and R-FEP scores were mistakenly combined and reported as EL. When R-FEP scores were disaggregated from EL scores, the dramatic EL improvement disappeared. However, even after being informed of the error, Unz refused to modify his statements (Weintraub & Chey, 1999).

Test scores of over one million immigrant students in California have risen by more than 50% since 1998, with those districts most rigorously embracing Prop. 227 having actually doubled their academic performance (Unz, 2001).

A second purpose of this study is to evaluate Unz's claim that EL test scores have improved dramatically since the passage of Prop. 227.

Method

Each spring California public schools administer a series of standardized achievement tests: the Standardized Testing and Reporting (STAR) program. These tests are administered to all public school students enrolled in grades two through eleven. As part of the testing program, demographic information, including language fluency, is collected. Students are classified into one of four language fluency categories: (1) English Only (EO), (2) Fluent English Proficient (FEP), (3) Reclassified Fluent English Proficient (R-FEP), or (4) English Learner (EL).

The STAR tests were first administered in the spring of 1998. Through 2002, the standardized NRT, SAT/9, form T, was administered as part of the STAR program. In 2003, the NRT was changed to the California Achievement Tests, Sixth Edition Survey (CAT/6). This study uses data from tests administered from the spring of 1998 through 2003.

STAR data were used to create three matched cohort files. For the first cohort file second-grade students tested in 1998 were matched with third-grade students tested in 1999, fourth-grade students tested in 2000, and fifth-grade students tested in 2001. Students were matched on the county/district/school (CDS) code, birth date, and gender. Each public school in California has a unique CDS code. The matched-cohort file contained information about the same group of students in the same school for four years at four different grade levels. Students who left the school, entered the school after grade two, or were held back were not part of the matched cohort.

There could be errors in the matching process but there was no reason to believe matching errors biased the results. The second cohort file was created by matching second-grade students in 1999 with third-grade students in 2000, fourth-grade students in 2001 and fifth-grade students in 2002. The third cohort was created by matching second-grade students in 2000 with third-grade students in 2001 fourth-grade students in 2002 and fifth-grade students in 2003. Again, these files contained information about the same group of students in the same school for four years at four different grade levels. The first cohort file (i.e., 1998-2001) included 192,023 students and the 1999-2002 and 2000-2003 cohort files had 224,425 and 277,373 students, respectively.

Matching students on home language generated a sub-sample of students, since the data field for home language could be missing or contain inconsistencies. If home language for a student was missing or inconsistent, a match could not be made and the student was dropped from the sample and the sample size was reduced. After the matching process, home language was constrained to two categories: Spanish (i.e., EL students whose home language was Spanish) and other language (i.e., EL students whose home language was something other than Spanish). The sub-sample for the 1998-2001 cohort had 57,348 students and was created to compare reclassification of Spanish EL students with other EL students.

Three different types of analyses were conducted. In one set of analyses the probability that EL students would be reclassified as R-FEP between second and fifth-grade was estimated. Toward that end the number and percent of students reclassified as R-FEP between second and fifth-grade were calculated. These analyses also calculated the number and percent of students not reclassified (i.e., the students who remained EL) between second and fifth-grade. These percents can be interpreted as probabilities. Analyses were also conducted for subgroups: gender (i.e., females compared to males), the national school lunch program (NSLP) participation (i.e., students receiving free and reduced lunch compared to those who do not), and home language (i.e., students whose home language is Spanish compared to students whose home language is neither English or Spanish).

A second series of analyses used logistic regression to test for subgroup differences in reclassification rates after accounting for differences in achievement. Reclassification, defined as

whether a student had been reclassified or not by the end of fifth-grade, was regressed on gender, NSLP, and home language and NRT scores.

A third series of analyses evaluated academic performance by language fluency. Average Stanford 9 total reading NCE scores were calculated for EO, FEP, R-FEP, and EL students across four years. For the 2000-2003 cohort the CAT/6 was administered in fifth-grade. The fifth-grade CAT/6 average reading NCE scores were converted to SAT/9 average reading NCE scores through equipercentile equating. These analyses were conducted to evaluate the effect of Prop. 227 on the test scores of EL and R-FEP students.

Results

Students in the matched cohorts have higher test scores on average than the state as a whole. It is assumed that scores are higher because students have remained in the same school for at least four years. Figure 1 compares the SAT/9 mean total reading scale scores for the whole state and for the 1998-2001 cohort sample.

Figure 1. Average SAT/9 total reading scale score for all students compared to the 1998-2001 matched cohort sample

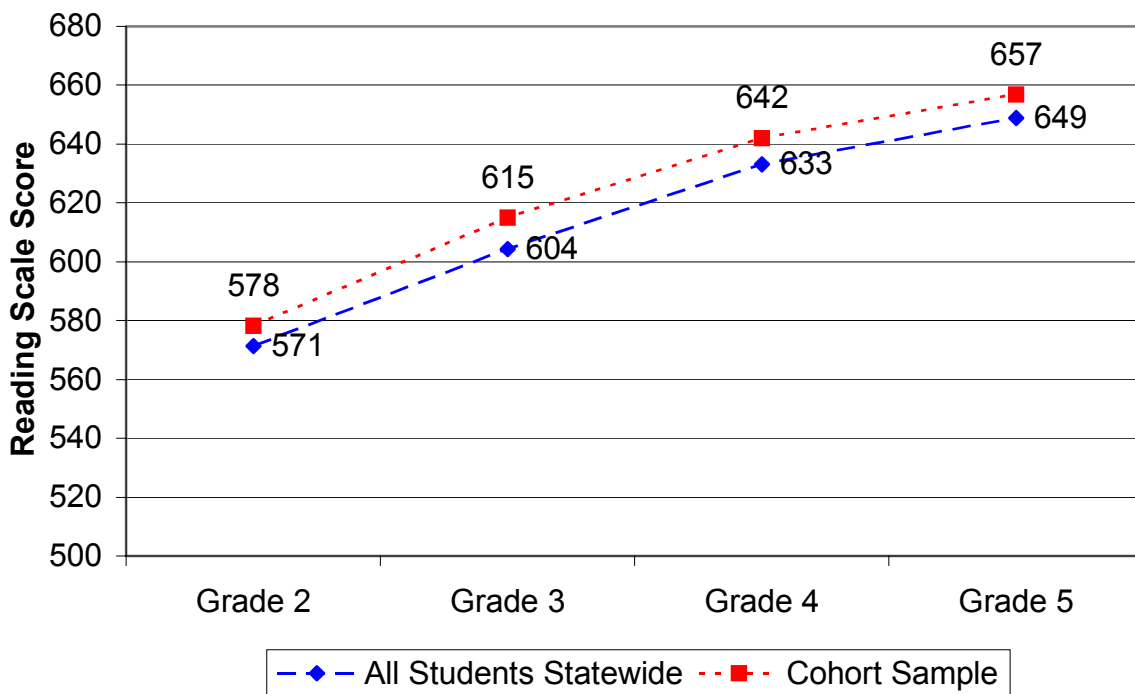
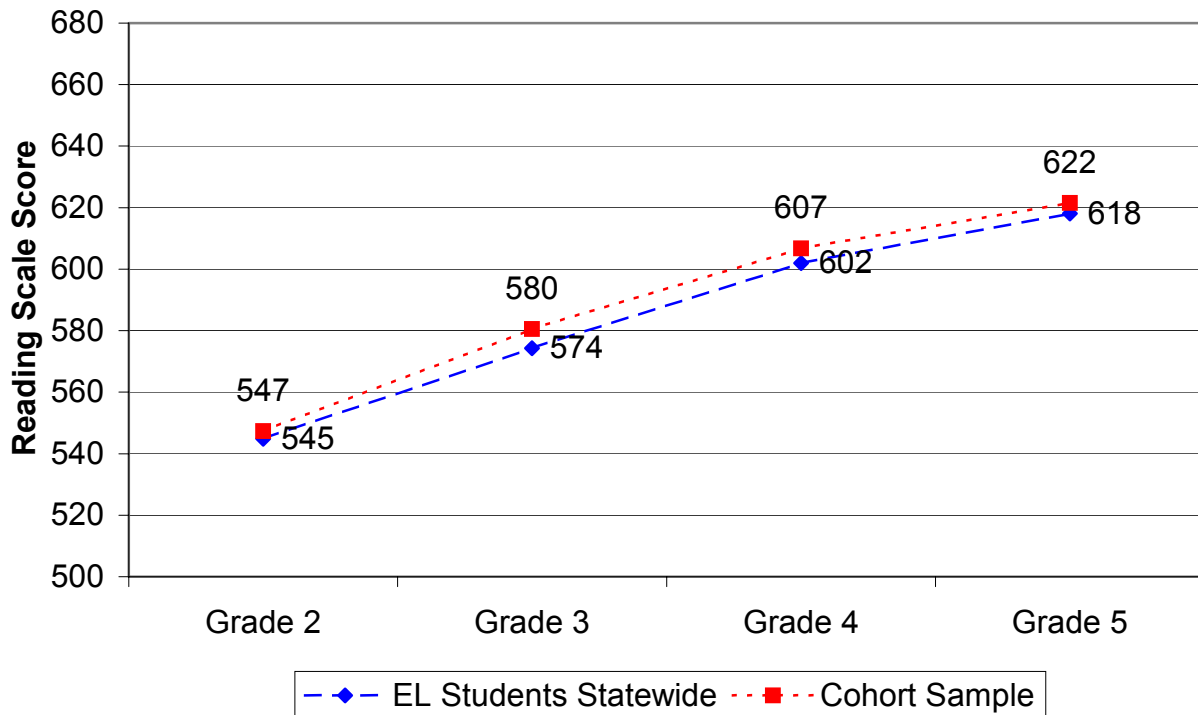


Figure 1 shows that the 1998-2001 cohort sample on average had higher SAT/9 mean total reading scale scores than the state as a whole. Figure 2 shows these same data for EL students.

Figure 2. Average SAT/9 total reading NCE score for EL students statewide compared to the 1998-2001 matched cohort sample

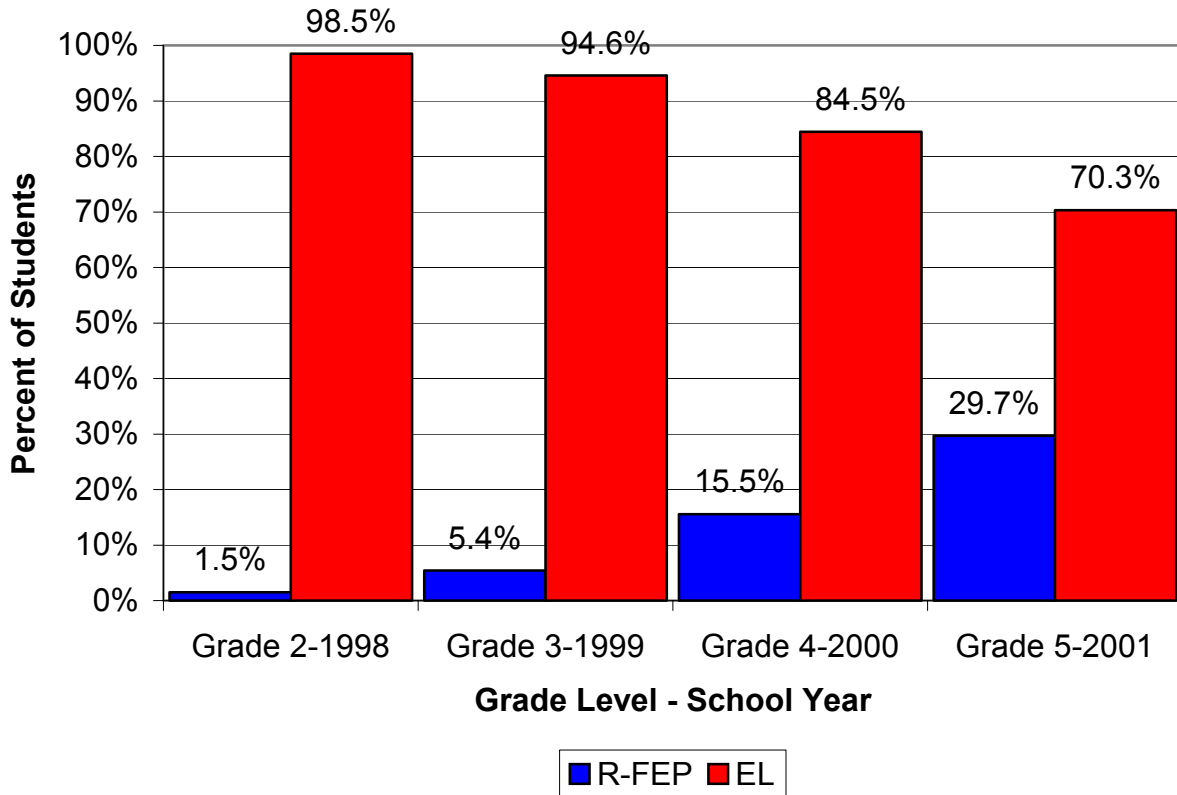


EL students in the 1998-2001 cohort sample on average scored slightly higher in total reading than EL students for the state as a whole. Results were consistent across other cohorts and indicate that subsequent analyses are based on groups of students that have higher test scores than the state as a whole. Results and conclusions need to be interpreted with these results in mind.

Reclassification Rates

Figure 3 shows the reclassification rate for the 1998-2001 matched cohort. It is a truer indicator of the reclassification process than annual reclassification rates because students did not move in or out of the group and a single group of students was tracked for four years.

Figure 3. Proportion of EL students reclassified R-FEP for the 1998-2001 cohort, n = 58,775

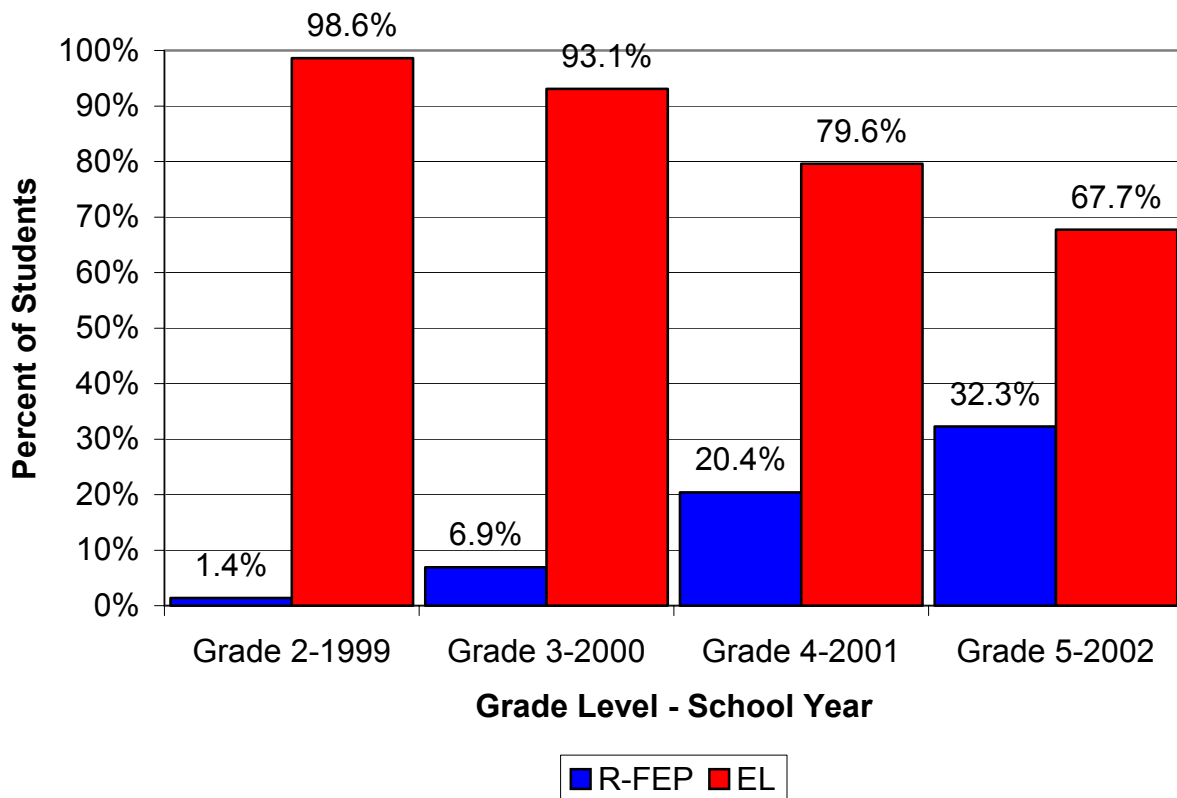


Results indicate the length of time to be reclassified is different than what might be imagined from the annual reclassification rates reported by Unz and his critics. That is, the percent of students reclassified each year is neither 5 nor 8 percent but varies from year to year. Within a year or two of being classified EL, few students are reclassified as R-FEP. It is not unreasonable to think that most of the second-grade EL students in this cohort were also first-grade EL students. In any case, less than 2 percent of EL students who started second, and possibly first grade, as EL were reclassified R-FEP by the end of the school year. By the end of third-grade, only an additional 4 percent of these same students had been reclassified. However, after two or three years of EL designation the reclassification rate began to increase. By the end of fourth-grade, an additional 10 percent were reclassified and by the end of fifth-grade, 14 percent more were reclassified. The pattern indicates that few students were reclassified within one to three years of entering the school system. State law asserts that the EL designation should not normally exceed one year, but after four or five years of schooling, only 30 percent of EL students had been reclassified.

These results can be interpreted as probabilities. That is, after four or five years of schooling (i.e., by the end of fifth-grade) EL students had a 30 percent probability of being reclassified as R-FEP and a 70 percent probability of remaining EL. Since reclassification is based in part on achievement data and the 1998-2001 cohort is higher achieving than the state as a whole, the true rate may be something less than 30 percent.

Figure 4 shows the same results for the 1999-2002 cohort. The 1999-2002 cohort is the class that is one year behind the 1998-2001 cohort.

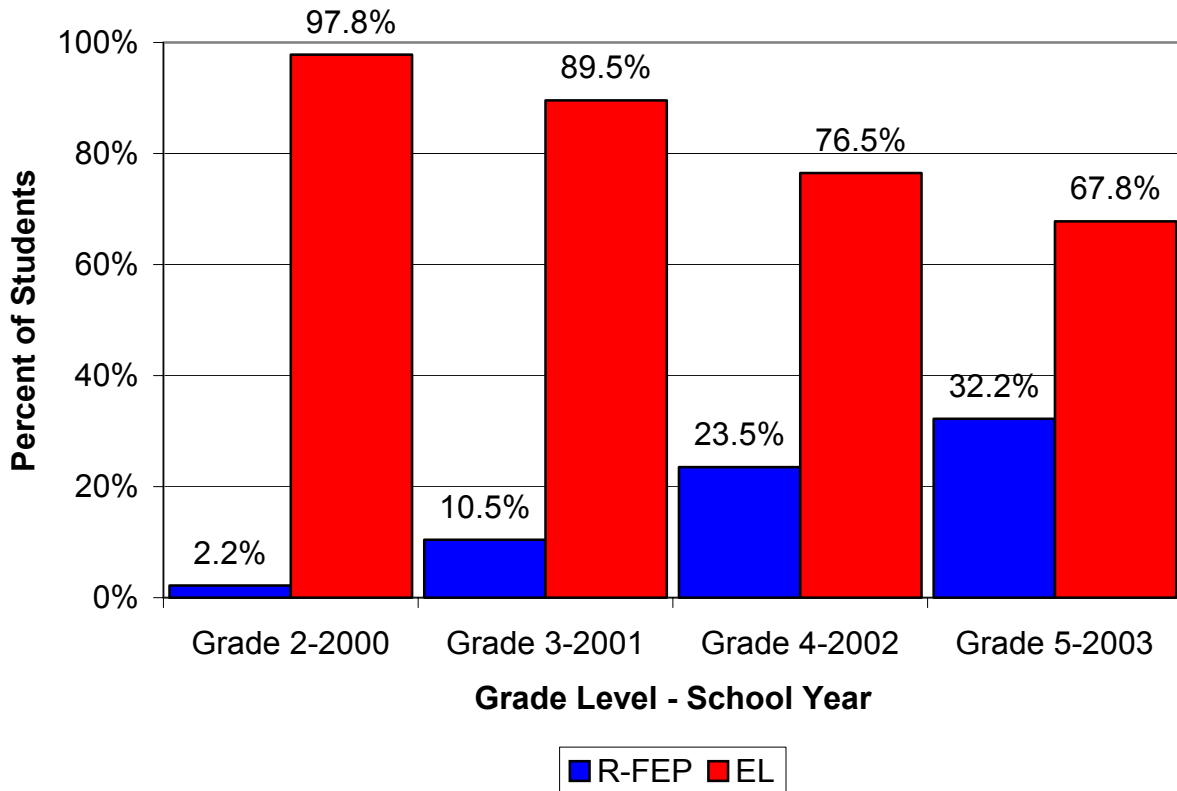
Figure 4. Proportion of EL students reclassified R-FEP for the 1999-2002 cohort, n = 72,806



The pattern is the same but the probability of being reclassified by the end of fifth-grade improved slightly. EL students now had a 32 percent probability of being reclassified and a 68 percent probability of remaining EL.

Figure 5 shows results for the 2000-2003 cohort.

Figure 5. Proportion of EL students reclassified R-FEP for the 2000-2003 cohort, n = 78,729



The pattern is comparable to both the 1998-2001 and 1999-2002 cohorts. EL students had a 32 percent probability of being reclassified and a 68 percent probability of remaining EL.

Data were available to estimate the reclassification rates through sixth-grade. Therefore, 1998-2002 and 1999-2003 cohorts could have been created. However, creating a matched file with the additional year/grade reduced the number of students in the cohort samples considerably. For example, the number of students in the cohort drops from 192,023 to 71,429 when the 1998-2001 cohort becomes the 1998-2002 cohort. In addition, the pattern of reclassification changes rather dramatically when sixth-grade is added. For the 1998-2002 cohort, there is only a 24 percent probability of an EL student being reclassified by the end of fifth-grade as opposed to a 30 percent probability for the 1998-2001 cohort. Given the larger sample and the consistency of

reclassification rates across cohorts, results are not reported for the 1998-2002 and 1999-2003 cohorts.

Data across three different cohorts indicates the probability of remaining an EL student after four (or five) years of school is approximately 70 percent. The passage of Prop. 227 has not produced a one or even two year transition process described in law.

Reclassification Rates by Subgroups

Next, reclassification rates were calculated for three subgroups: gender (i.e., females compared to males), NSLP participation (i.e., students who receive free or reduced lunch compared to those who do not), and home language (i.e., students whose home language is Spanish compared to students whose home language is neither English nor Spanish). Figure 6 shows reclassification rates by gender for the 1998-2001 cohort.

Figure 6. Proportion of EL students reclassified R-FEP for the 1998-2001 cohort by gender, n = 58,775

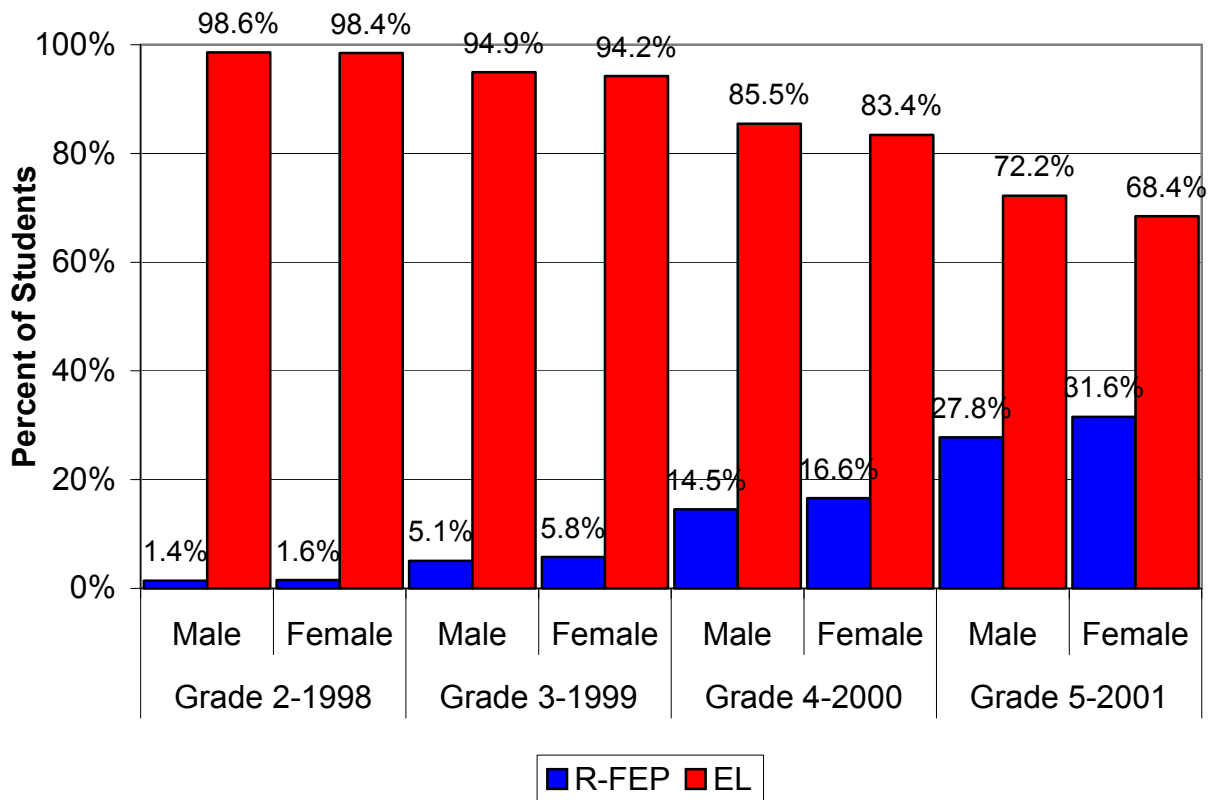


Figure 6 shows female EL students were more likely to be reclassified than males. By the end of fifth-grade the probability of females being reclassified R-FEP was 32 percent and for males the probability was 28 percent.

Figure 7 shows reclassification rates for NSLP students and non-NSLP students.

Figure 7. Proportion of EL students reclassified R-FEP for the 1998-2001 cohort by NSLP, n = 58,775

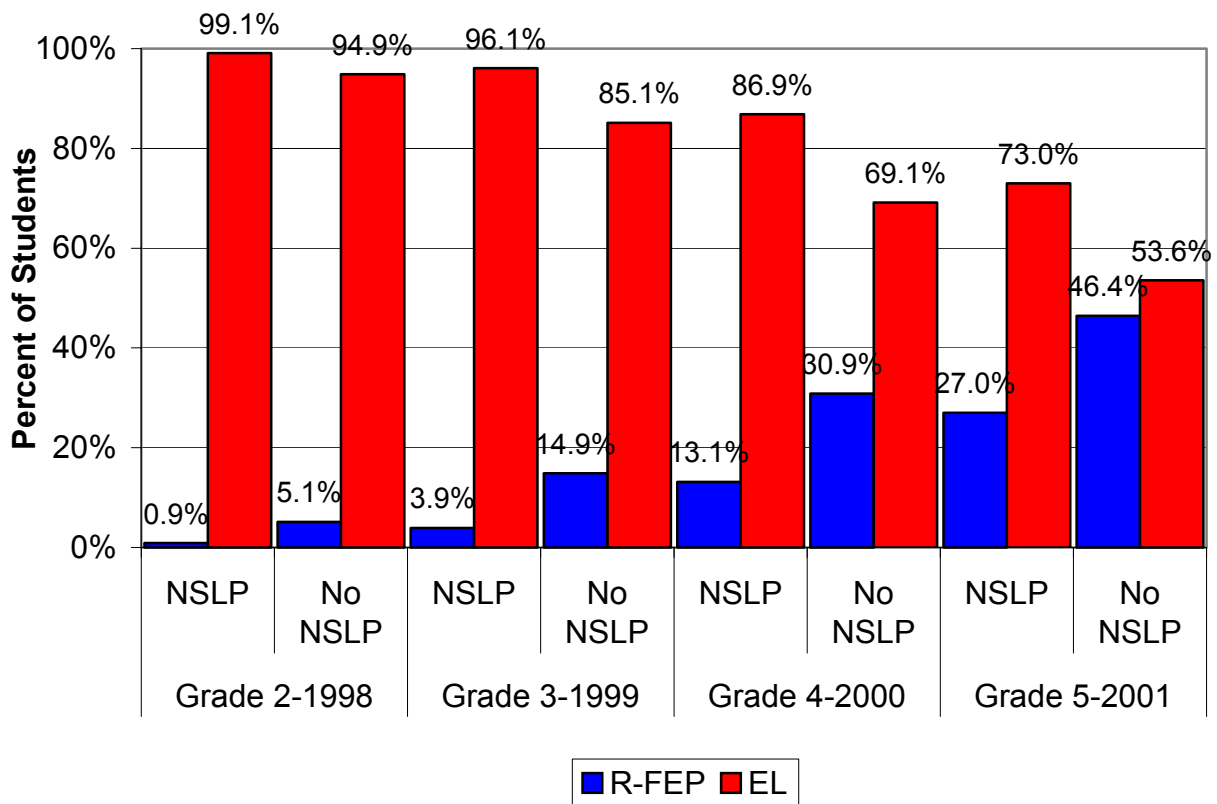


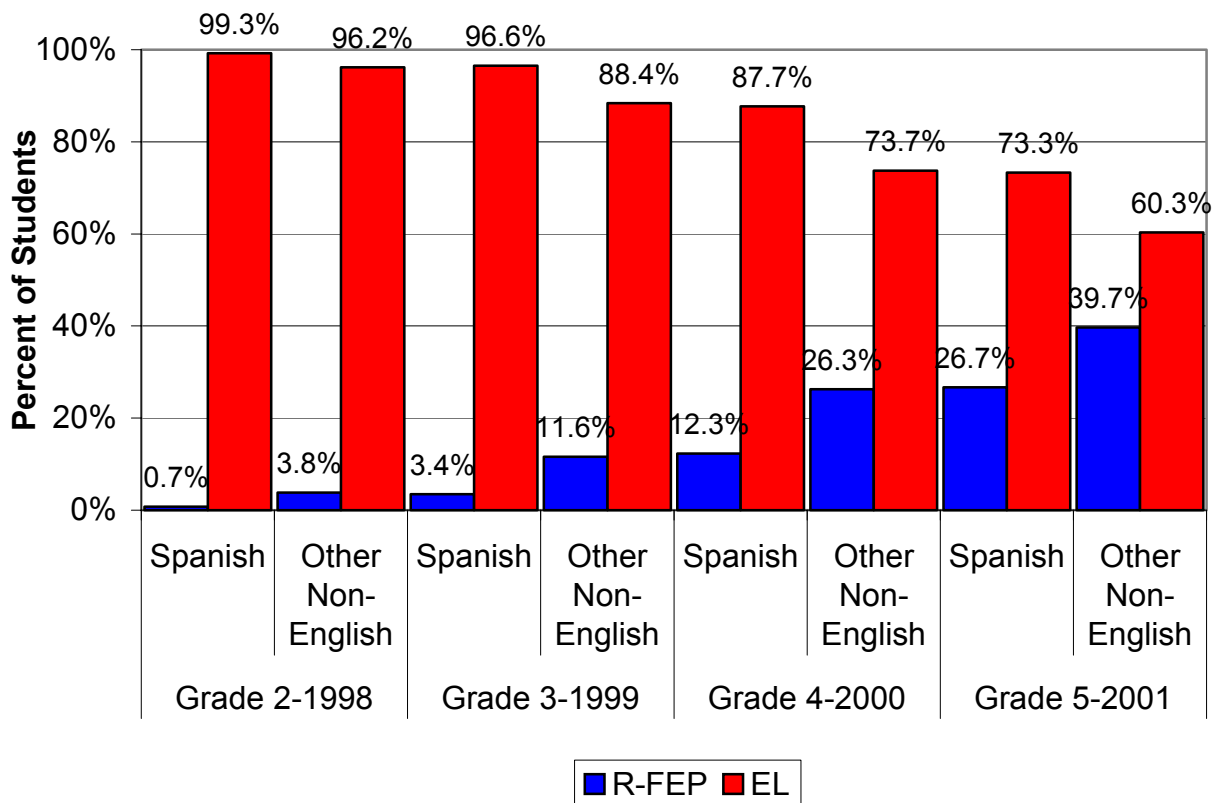
Figure 7 indicates that EL NSLP students had a 27 percent chance of being reclassified R-FEP by the end of fifth-grade and EL no NSLP students had a 46 percent chance.

NSLP serves as a proxy for socio-economic status (SES). Participation in NSLP is an indicator of lower SES. No NSLP is an indicator of higher SES. Parent education was another available proxy

for SES. However, analyses of parent education were consistent with NSLP and results are not displayed.

Figure 8 shows the reclassification rates for Spanish EL students compared to other language EL students.

Figure 8. Proportion of EL students reclassified R-FEP for the 1998-2001 cohort by home language, n = 57,348



In Figure 8 Spanish EL students had a 27 percent chance to be reclassified R-FEP by the end of fifth-grade and other language EL students had a 40 percent.

Results were consistent across cohorts. The data suggest that male, NSLP, and Spanish EL students have a lower probability of being reclassified R-FEP than female, non-NSLP, and other language EL students. However, the reclassification process relies on multiple criteria and a crucial aspect of the reclassification process was the assessment of basic skills. Scores on a standardized

achievement test were used to evaluate basic skills. To account for the relationship between academic achievement and reclassification, logistic regression was used to test for group differences after holding achievement constant. Reclassification, defined as to whether a student had been reclassified or not by the end of fifth grade, was regressed on gender, NSLP, home language, and NRT total reading normal curve equivalent (NCE) scores for four years. Table 1 shows these results for the 1998-2001 cohort.

Table 1
Reclassification regressed on gender, NSLP, language, and reading scores
for the 1998-2001 cohort

Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Chi-Square	Pr > ChiSq
Intercept	1	-5.3359	0.0579	8502.53	<.0001
Gender (female)	1	0.0511	0.0192	7.09	0.0077
NSLP (NSLP)	1	0.0227	0.0195	1.36	0.2441
Language (Spanish)	1	0.1009	0.0198	25.99	<.0001
Gender*NSLP	1	-0.0418	0.0192	4.74	0.0296
Gender*Language	1	-0.0448	0.0192	5.44	0.0197
NSLP*Language	1	0.1373	0.0192	51.24	<.0001
Gender*NSLP*Language	1	0.0173	0.0192	0.81	0.3683
Reading_NCE98	1	0.0252	0.0012	444.14	<.0001
Reading_NCE99	1	0.0251	0.0015	300.59	<.0001
Reading_NCE00	1	0.0286	0.0015	389.21	<.0001
Reading_NCE01	1	0.0366	0.0014	708.65	<.0001

The intercept (i.e., -5.3359) represents the probability of being reclassified. This logit value represents approximately .005%. Parameter estimates with positive values move this percent closer to 1 (i.e., increase the likelihood of being reclassified) and negative values move the value away from 1 (i.e., decrease the likelihood of being reclassified). For example, holding achievement and other variables constant, females were significantly, at the .008 level of significance, more likely than males to be reclassified.

Even though NSLP students were more likely to be reclassified after holding achievement and other variables constant, the difference between reclassification rates for NSLP and non-NSLP EL students was not significant, at a .01 level of significance. Non-NSLP students scored higher than NSLP EL students on the SAT/9 reading test and were thus more likely to be reclassified. However, when achievement was held constant the difference in reclassification rates disappeared.

Home language was significant at the .0001 level of significance. After controlling for the effects of achievement and other variables, Spanish EL students were *more* likely to be reclassified R-FEP than other language EL students. If both home language groups were being treated in the same way, controlling for test score differences should have the same effect as NSLP. That is, the differences between the groups would have no longer been significant. However, the direction of the parameter estimate raises the suspicion that a large number of other language EL students, who were eligible for reclassification, given their NRT test scores, were not reclassified.

To test this suspicion, test scores for EL students from the two different home language groups were compared. Table 2 shows these results.

Table 2
Mean Reading NCE Score

Test	Other Non-English	Spanish
Reading_NCE98	48.4	31.5
Reading_NCE99	47.7	32.7
Reading_NCE00	51.5	35.6
Reading_NCE01	51.1	37.0

Other language EL students on average had higher SAT/9 reading scores than Spanish EL students and were thus more likely to be reclassified.

The next analysis attempted to determine if other language EL students were under-represented in the R-FEP language category. If so, that would explain the regression results. For each year in the 1998-2001 the EL students who had not been reclassified and who had scored at or above the 36th

percentile on the SAT/9 were identified. The 36th percentile was selected because it has been a traditional score to determine student reclassification. Figure 9 shows these results.

Figure 9. Percent of EL students scoring at or above the 36th percentile by home language for the 1998-2001 cohort, n = 26,970

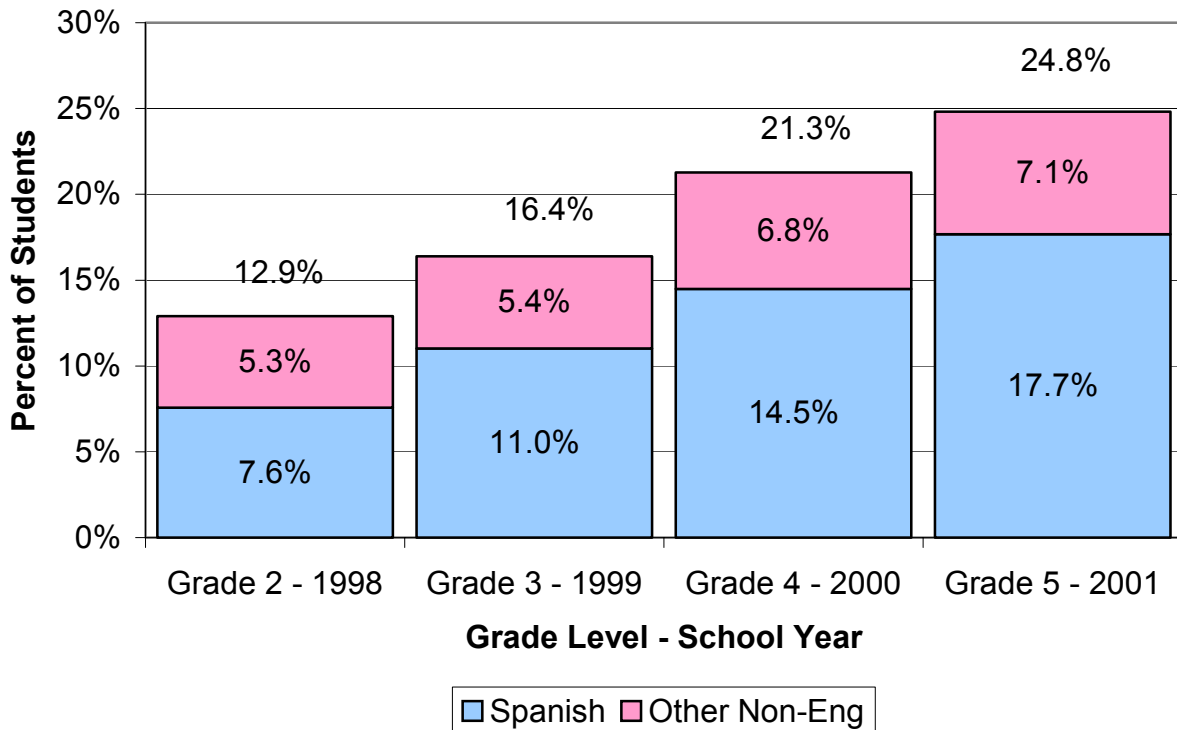


Figure 9 shows a couple of different things. First, it shows the percentage of students who were possible candidates for reclassification based on scoring at or above the 36th percentile on the SAT/9 reading test. In 1998, 12.9 percent (i.e., 5.3% + 7.6%) met the reclassification threshold of the 36th percentile but were not reclassified. In 1999, 16.4 percent met the threshold value, and in 2000 and 2001 there were 21.3 and 24.8 percent, respectively, that met the threshold value. Each year there were a certain percentage of students who were strong candidates for reclassification but were not reclassified and each year this percentage increased. By grade five, 25 percent of EL students who were strong candidates for reclassification had not been reclassified.

Second, Figure 9 shows the percent of other language and Spanish EL students who met the reclassification threshold of the 36th percentile but were not reclassified. In 1998 for example, the

percent of other language EL students was 5.3 percent. For Spanish EL students it was 7.6 percent. Continuing to use 1998 as an example, other language EL students represented 41.2 percent of the 12.9 percent total and Spanish EL students represented 58.8 percent. However, for the full 1998-2001 cohort, other language EL students represented 26 percent of the total and Spanish EL students represented 74 percent. The other language EL students represented a larger percentage of EL students that met the NRT threshold for reclassification but were not reclassified than they did of EL students overall. That is why the regression analysis indicated that Spanish EL students were more likely to be reclassified than other language EL students when achievement was controlled.

Figure 3 shows the number of EL students after grade 5 for the 1998-2001 cohort as 41,143. Figure 9 shows this same value as 26,970 students. The number of students in Figure 9 represents those EL students who had reading test scores for grades 2 through 5 and non-missing home language information. The requirement to have non-missing data for the four different reading tests and home language reduced the sample size.

Back to Table 1, there is also a significant interaction effect for NSLP and home language. The interpretation is that even though Spanish EL students were more likely to be reclassified than other language EL students, the Spanish / NSLP students were even more likely than the Spanish / non-NSLP students to be reclassified after holding achievement and other variables constant.

The regression analysis indicates that the strongest predictors of whether students would be reclassified were reading test scores. As test scores went up, the probability of being reclassified increased. In addition, when achievement was held constant Spanish students were more likely, rather than less likely, than other language EL students to be reclassified. Table 3 shows the regression analysis for the 1999-2002 cohort.

Results for the 1999-2002 cohort were consistent with the 1998-2001 cohort. Reading test scores and language (i.e., Spanish) were again the variables most strongly related to reclassification. And again, female EL students were more likely to be reclassified than males after controlling for the effects of achievement and other variables.

Table 3
Reclassification regressed on gender, NSLP, language, and reading scores
for the 1999-2002 cohort

Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Chi-Square	Pr > ChiSq
Intercept	1	-5.8563	0.0520	12661.58	<.0001
Gender (female)	1	0.0464	0.0167	7.75	0.0054
NSLP (NSLP)	1	0.0491	0.0169	8.49	0.0036
Language (Spanish)	1	0.1496	0.0171	76.48	<.0001
Gender*NSLP	1	-0.0108	0.0166	0.42	0.5177
Gender*Language	1	-0.0376	0.0166	5.09	0.0240
NSLP*Language	1	0.0818	0.0166	24.12	<.0001
Gender*NSLP*Language	1	0.1330	0.0166	0.64	0.5177
Reading_NCE99	1	0.0259	0.0010	674.71	<.0001
Reading_NCE00	1	0.0377	0.0013	914.17	<.0001
Reading_NCE01	1	0.0265	0.0012	463.52	<.0001
Reading_NCE02	1	0.0355	0.0012	918.04	<.0001

However, for the 1999-2002 cohort the difference in reclassification rates for NSLP and non-NSLP students was statistically significant at the .01 level. Students receiving free and reduced lunch were more likely to be reclassified R-FEP after controlling for achievement and other variables.

Table 4 shows the regression analysis for the 2000-2003 cohort. Results are consistent with the other cohorts except female EL students were neither more nor less likely to be reclassified than male EL students holding achievement and other variables constant. EL students receiving NSLP were neither more or less likely to be reclassified than non-NSLP EL students and Spanish

speaking EL students were neither more or less likely to be reclassified than non-Spanish EL students. Again, there was a significant interaction effect for NSLP and home language but the direction was reversed from the other cohorts. The interpretation is that, even though there was no relationship between being reclassified, NSLP and home language, the NSLP / non-Spanish students were more likely than the NSLP / Spanish students to be reclassified after holding achievement and other variables constant. As with the other cohorts, the strongest predictors of reclassification were reading test scores.

Table 4
Reclassification regressed on gender, NSLP, language, and reading scores
for the 2000-2003 cohort

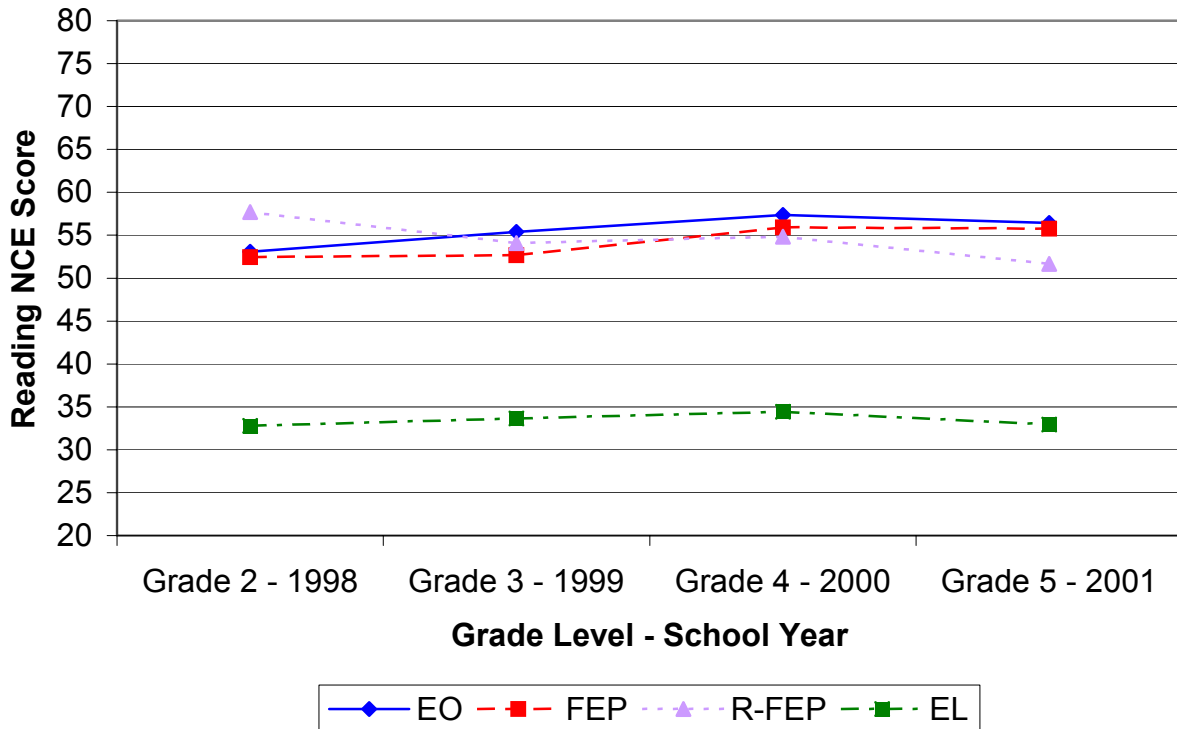
Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Chi-Square	Pr > ChiSq
Intercept	1	8.9398	0.1219	5382.05	<.0001
Gender (female)	1	-0.0170	0.0163	1.09	0.2986
NSLP (NSLP)	1	0.0232	0.0163	2.02	0.1555
Language (Spanish)	1	-0.0325	0.0165	3.89	0.0486
Gender*NSLP	1	0.0033	0.0162	0.04	0.8376
Gender*Language	1	0.0170	0.0162	1.09	0.2956
NSLP*Language	1	-0.0502	0.0162	9.57	0.002
Gender*NSLP*Language	1	0.0195	0.0162	1.45	0.228
Reading_NCE00	1	-0.0440	0.0009	218931.00	<.0001
Reading_NCE01	1	-0.0366	0.0012	887.70	<.0001
Reading_NCE02	1	-0.0272	0.0012	529.96	<.0001
Rreading_CST_SS03	1	-0.0115	0.0005	640.87	<.0001

Academic Performance by Language Fluency

Academic performance was estimated by calculating average SAT/9 total reading NCE scores by language fluency (i.e., EO, FEP, R-FEP, & EL) across four years. Figure 10 shows these results for the 1998-2001 cohort.

Figure 10. Average SAT/9 total reading NCE score by language fluency for the 1998-2001 cohort, n = 145,873



The average reading NCE scores for EO, FEP, and R-FEP students were comparable, but for EL students the average reading scores were much lower. For EO and FEP students there was a slight upward trend in the average reading score but for R-FEP students there was a slight downward trend. For EL students the average reading scores remained fairly constant over time. For EO and FEP students, the test scores were computed for the same students each year. For R-FEP and EL students test scores were computed for different students each year. Each year the number of R-FEP students increased and the number of EL students decreased because each year more students were reclassified.

Students reclassified as R-FEP in 1998 were the most academically precocious EL students by virtue of the fact that they were the first to meet both language and academic reclassification requirements. Students reclassified as R-FEP in 2001 were the least academically proficient of those reclassified by virtue of the fact that it took them the longest time to meet the reclassification requirements.

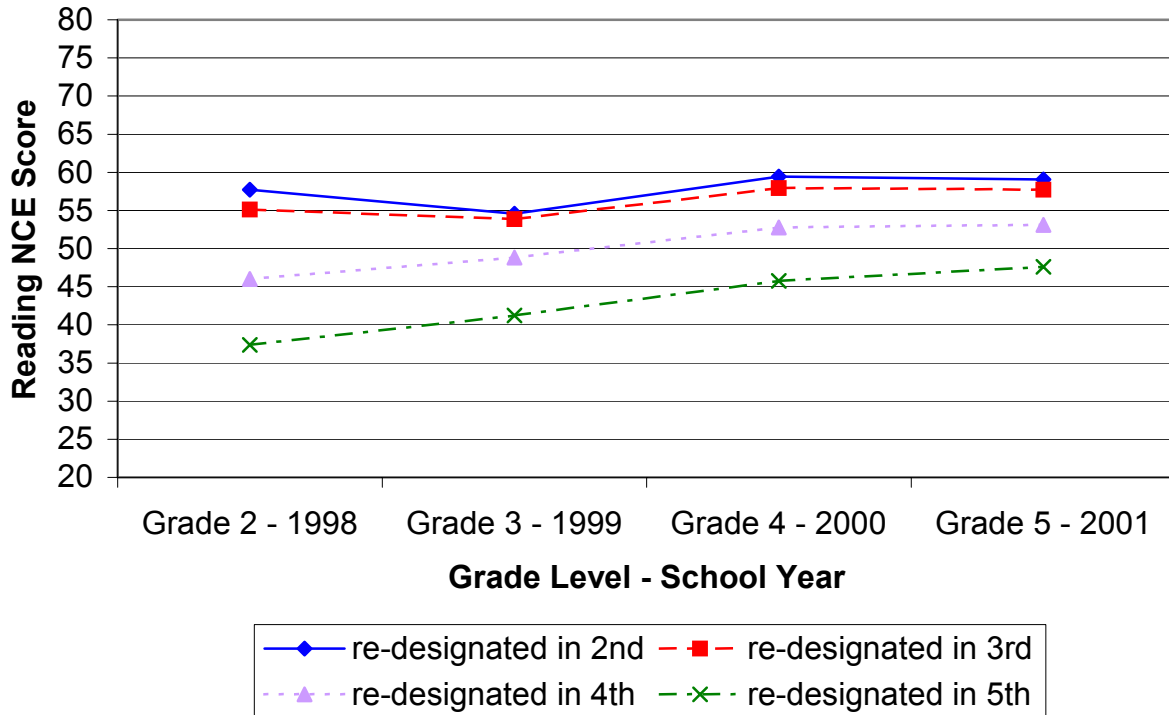
The downward trend in test scores for R-FEP students should not automatically be interpreted to mean R-FEP performance was declining. The lower scores indicate that each year less able students joined the R-FEP group. Even so, the R-FEP average in 2001 was above the 50th percentile of the norming sample.

The low EL test scores represent the opposite trend of R-FEP. Each year the most academically proficient students left this group and were reclassified R-FEP. The continuously low academic performance of EL students should not be interpreted to mean that EL students never improve or were failing to close the gap between themselves and the other language categories. Each year the EL group represented those students who were left behind after the most academically able were reclassified as R-FEP.

Test scores in Figure 10 are average scores. There was variance around these scores. In 2001 for each language designation, the individual NCE scores ranged from 1 to 99. The standard deviations for EO, FEP, R-FEP, and EL students scores were 19, 18, 15, and 14 respectively. The overall standard deviation for the 2001 grade 5 reading NCE scores was 21. Therefore, even though average EL scores were noticeably lower than EO, FEP, and R-FEP average scores there were EO, FEP, and R-FEP students scoring lower than the average for EL students.

Figure 11 shows the pattern of test scores across years by the grade in which students were reclassified. The number of students represents the total number of students who were reclassified in grades 2 through 5.

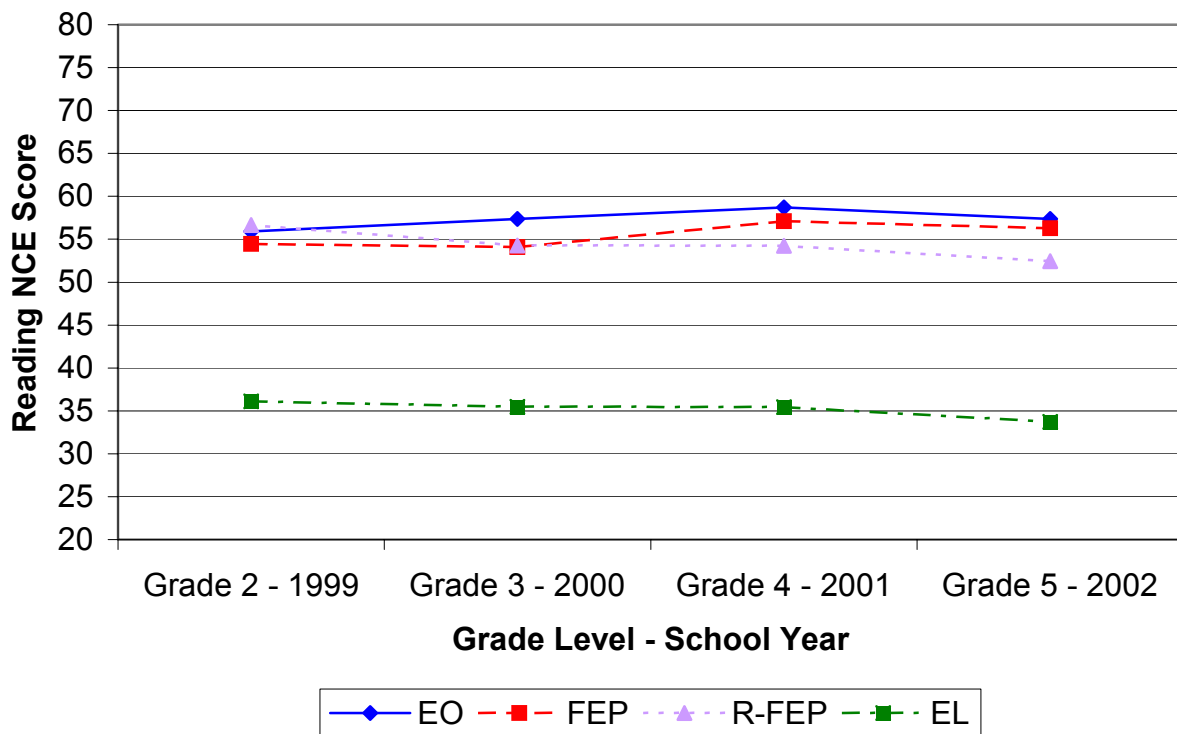
Figure 11. Average SAT/9 total reading NCE score for 1998-2001 cohort R-FEP students by the grade in which students were reclassified, n = 17,436



Students reclassified in 1998 had a pattern of highest test scores. Students reclassified in 2001 had the pattern of lowest test scores. These data support the contention that students reclassified in second-grade were more academically precocious than students reclassified in grade five. However, students reclassified in fifth-grade showed the most improvement over time. Average performance of students reclassified in second-grade had stabilized while students reclassified in the fourth and fifth-grades were closing the achievement gap.

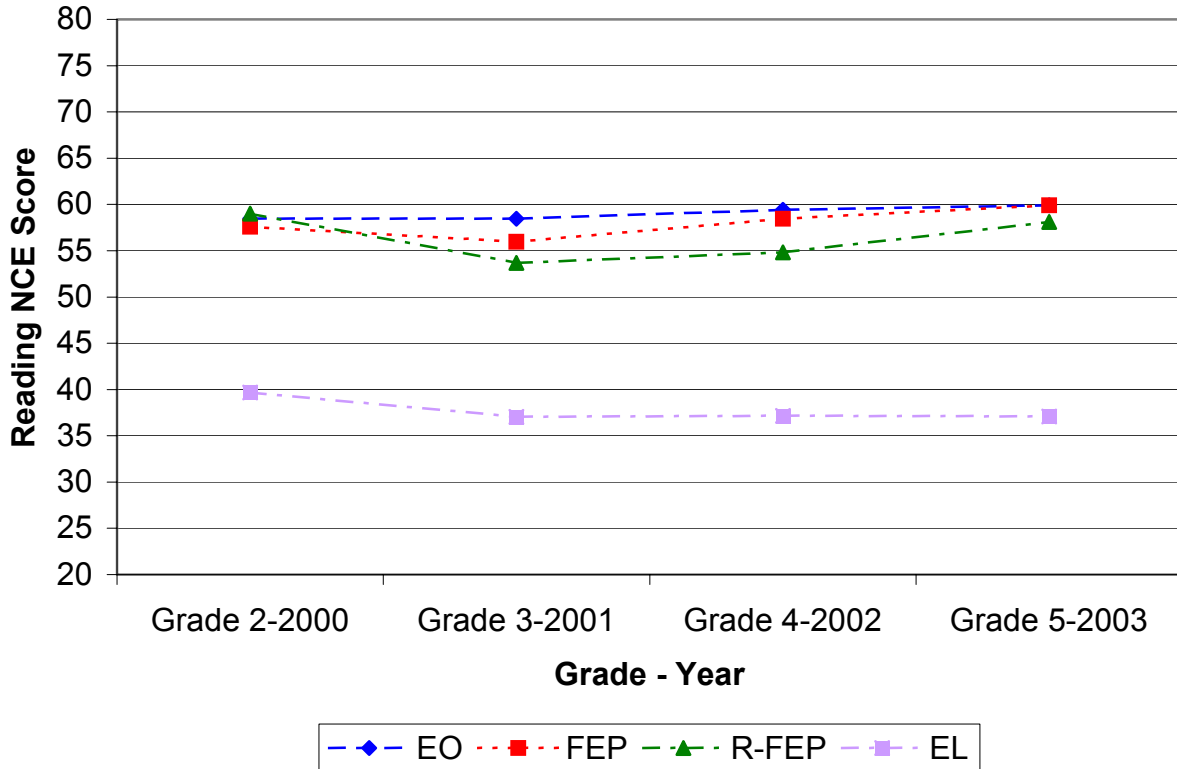
Figures 12 and 13 show achievement results for the 1999-2002 and 2000-2003 cohorts.

Figure 12. Average SAT/9 total reading NCE score by language fluency for the 1999-2002 cohort, n = 195,082



Results for the 1999-2002 cohort show the same pattern as the 1998-2001 cohort. However, the average score across language groups improved. This was not surprising since it has been widely reported that when the same test series is used year after year, test scores tend to improve as teachers become more aware of test content (Linn, Graue, & Sanders, 1990). For the 2000-2003 cohort, the average reading score across groups improved even more and the EO, FEP, and EL trends are comparable to the 1998-2001 and 1999-2002 cohorts. However, the R-FEP students in the 2000-2003 cohort did not demonstrate the downward trend in test scores seen in the other cohorts.

Figure 13. Average SAT/9 total reading NCE score by language fluency for the 2000-2003 cohort, n = 214,830



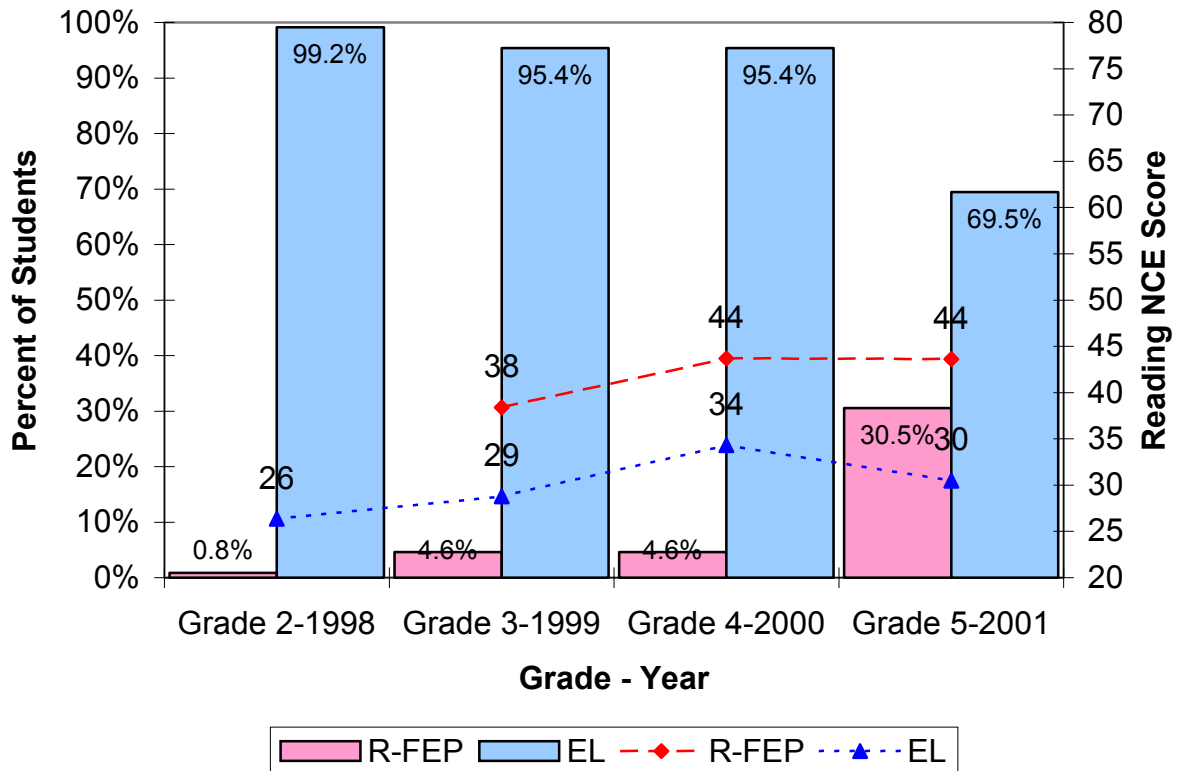
For the 2000-2003 cohort the STAR NRT changed from the SAT/9 to the CAT/6 in 2003. An equipercentile equating was done to make scores comparable over time. However, there might be error around the equating process that negates the downward trend. The downward trend was slight and slight error might disguise it. Or, it could be that the R-FEP average reading scores for the 2000-2003 cohort improved in fifth-grade and there was no longer a downward trend.

Academic Performance by Language Fluency in a Single District

Unz often references a particular school district in California as a model of the positive effects of Prop. 227 (Nishioka, 1999). Unz claims that the 50 percent rise in test scores was evidence that the English immersion practiced in this model district and Prop. 227 were working (Sailer, 2002).

Figure 14 shows average total reading NCE scores and reclassification rates for the 1998-2001 EL and R-FEP students in the model district.

Figure 14. Average SAT/9 reading NCE score for 1998-2001 R-FEP and EL students in the model district, n = 239

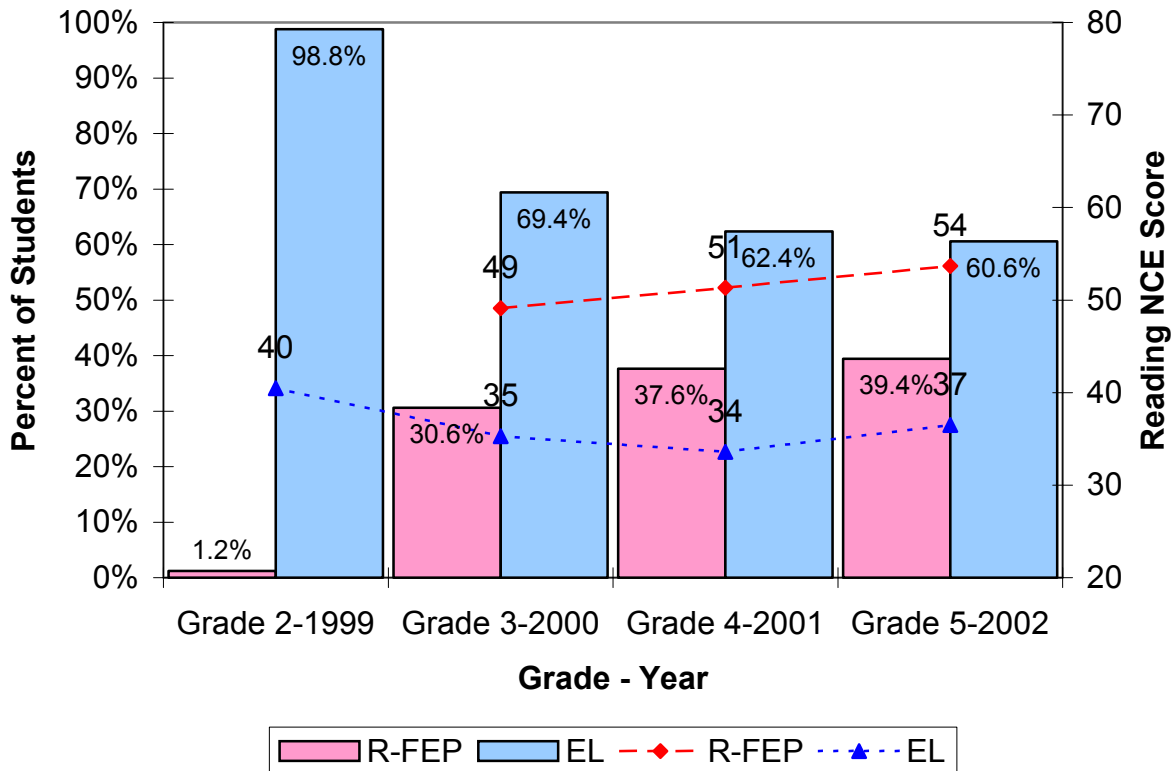


Test scores are not reported for grade two R-FEP students because CDE has a policy of not reporting scores for less than ten students. There were only two R-FEP students in grade two.

Data indicate that reclassification rates and test scores for this district’s EL and R-FEP students were lower than the state average. Test scores for EL students did not rise 50 percent between second and any of the other grade levels. Test scores did not rise 50 percent between third and any of the other grades for R-FEP students.

Perhaps it was too soon for Prop. 227 to affect the 1998-2001 cohort. Figure 15 reports results for the 2000-2003 cohort.

Figure 15. Average SAT/9 reading NCE score for 2000-2003 R-FEP and EL students in the model district, n = 327



Data in Figure 15 indicate that test scores were still a bit lower than the state average for EL and R-FEP students. Reclassification rates have improved over the district’s 1998-2001 cohort and the reclassification rate at the end of fifth-grade is higher than the state average. Even so, students in this model district take much longer than a year to be reclassified and test scores for their R-FEP and EL students were lower than the state average.

Discussion

To better estimate reclassification rates, cohorts were created so the same group of students could be tracked over time. Based on data from three cohorts, the probability that EL students would be reclassified R-FEP by the end of fifth-grade was 30 to 32 percent. Conversely, the probability that

EL students would not be reclassified R-FEP by the end of fifth-grade was 68 to 70 percent. The goal of reclassifying EL students as R-FEP within a year or two of entering the school system has not been achieved with the passage of Prop. 227.

It is unlikely Prop. 227, as written, had or will have any effect on reclassification rates.

Reclassification is dependent on the multiple criteria used in the reclassification process. These criteria existed before and after the passage of Prop. 227. One of these criteria, performance in basic skills, was reported by districts to be the biggest barrier to reclassification. Unz was critical of the basic skills requirement.

Children from immigrant or Latino backgrounds are categorized as not knowing English if they merely score below average on English tests, meaning that unknown numbers of children whose first and only language is English spend their elementary school years trapped in Spanish-only bilingual programs (Unz, 1997).

However, when Unz drafted Prop. 227, the reclassification criteria were not addressed in the new legislation.

Reclassification rates for the 1999-2002 and 2000-2003 cohorts were slightly higher than the 1998-2001 cohort. This slight improvement, if it is improvement rather than random year-to-year fluctuation, was more likely the result of better tracking at the local level. Rather than Prop. 227, CELDT testing and the requirement to include EL students in California's statewide accountability index have pushed districts to improve the tracking of EL students. Since it is less likely for students to fall through the cracks, reclassification rates improved.

There were differences in reclassification rates for subgroups. Females were more likely to be reclassified than males. Non-NSLP students were more likely to be reclassified than NSLP students, and other language EL students were more likely to be reclassified than Spanish EL students. However, regression analyses revealed when achievement was held constant these differences generally disappeared. Females were still more likely to be reclassified than males when achievement was held constant for the 1998-2001 and 1999-2002 cohorts but not for the 2000-2003 cohort. When achievement was held constant the difference in reclassification rates between non-NSLP students and NSLP students either disappeared or reversed (i.e., NSLP students are

more likely to be reclassified) and when achievement was held constant Spanish EL students were either more likely to be reclassified than other language EL students or there was no difference. The regression analyses further indicate that a major factor for reclassification was performance on standardized tests.

Multiple classification criteria exist to protect students from being reclassified too quickly. However, there may be an overprotected group of EL students. That is, by the end of fifth-grade 25 percent of EL students in the 1998-2001 cohort who were strong candidates for reclassification, based on standardized test scores, had not been reclassified. This finding warrants further study to better understand how LEAs use the reclassification criteria.

The 30 to 32 percent reclassification rate of EL to R-FEP after four or five years of schooling raises questions about the reclassification process itself. Educators of English learners need to evaluate whether students are being reclassified at an appropriate rate or too slowly. Are the safe guards to protect students from being reclassified too quickly helping or hindering the academic achievement of EL students? What are the advantages and disadvantages of long term EL designation?

Although Unz claimed a dramatic improvement in EL test scores after the passage of Prop. 227, his claims seemed questionable even before looking at the data. First, the dramatic improvement was based on the change in scores from 1998 to 1999 data. The initial CDE STAR report for 1999 had an error that was not caught until after data were released. The error consisted of combining EL and R-FEP scores and reporting the combined data as EL. At first, EL scores seemed to have improved dramatically. When the error was discovered and corrected by disaggregating the R-FEP from EL scores, the dramatic EL improvement disappeared. Even though Unz was well aware of the error in the initial 1999 report, he has failed to modify his statements about dramatically improved EL test scores. Second, when EL students demonstrate higher academic performance they are reclassified R-FEP. So, it is difficult to track improvement in EL scores because higher performing EL students would no longer be classified EL. Third, other large scale assessments such as NAEP do not support dramatic year-to-year change in student performance. It is very difficult to dramatically improve student achievement, even when that is the specific focus. Shepard, Flexer, Hiebert, Marion, Mayfield, & Weston (1996) found no achievement differences

between experimental and control subjects after a year-long project focused on modifying teacher pedagogy to improve student achievement.

After looking at achievement data, it appears Prop. 227 had no effect on student test scores. For EO and FEP students, there was a slight upward trend in the average reading score. Much of this change was likely due to using the same test year-to-year. For R-FEP students, there was a slight downward trend, except for the 2000-2003 cohort. The downward trend was likely due to less academically able EL students (i.e., less able than the already reclassified R-FEP students) being reclassified R-FEP. This should not be interpreted to mean that students who take longer to be reclassified are not academically capable. It simply means they tend to be less capable than students who have already been reclassified. For EL students, the average reading score remained consistently low over time. Scores remained low because the more academically proficient students were reclassified R-FEP.

There was no dramatic improvement in test scores across years within a cohort or from cohort to cohort for any of the language fluency categories. For example, for the 1998-2001 cohort there was no dramatic improvement in reading scores from grade two to grade three and there was also no dramatic improvement in reading scores from the 1998-2001 to the 2000-2003 cohort for any of the language fluency categories. Test scores changed in a manner that might be expected when the same test battery was administered year after year.

Data from Unz's model district do not support his claims that English immersion programs dramatically improved EL and/or R-FEP test scores. Test scores from Unz's model district did not show any dramatic upward trend. Scores were even lower than the statewide average. In addition, EL students in Unz's model district took considerably longer than a year to be reclassified. Hakuta (2002) reported comparable results.

Prop. 227 has had no effect on EL reclassification rates or test scores. Yet, a review of magazine and newspaper articles indicated that reporters generally accepted and reported Unz's data and anecdotal evidence without question. It is difficult to find a clear coherent criticism of Unz's statements in the press. For example, Unz's critics were correct when they said the annual reclassification rate was closer to 7 than 5 percent. Yet, Unz's 5 percent rate was reported over and

over again. Aryal (1998) reported there were specific reasons why Unz's message was more widely reported than his critics. During the Prop. 227 campaign, Unz repeated the same message, promptly returned phone calls, provided sound bites, and was the clear point person for the initiative. In contrast, opponents of Prop. 227 were a diverse group with a profusion of messages and difficult to reach. Even so, reporters could have verified or at least called into question Unz's statistics by visiting CDE's web site but failed to do so. The unfortunate aspect of not verifying data is that Unz has been given free reign to report misinformation that has influenced educational policy. The false claim, that there was a 50 percent improvement in EL achievement, has been reported so often in so many different sources that it has assumed a reality that this study is unlikely to undermine.

References

- Aryal, M. (1998). He says, she says: How California's major papers have covered prop. 227. *MediaFile*, 17 (3). Retrieved April 2, 2004, from <http://www.media-alliance.org/mediafile/17-3/mediafile.html>.
- California Department of Education (2004). Web site: <http://www.cde.ca.gov>.
- California Law (2003). Education Code. Web site: <http://www.leginfo.ca.gov/calaw.html>.
- Crawford, J. (2003). *A few things Ron Unz would prefer you didn't know about English learners in California*. Retrieved March 24, 2004, from <http://ourworld.compuserve.com/homepages/JWCRAWFORD/castats.htm>.
- Gandara, P. (2000). In the aftermath of the storm: English learners in the post-227 era. *Bilingual Research Journal Online*, 24(1&2). Retrieved May 14, 2003, from <http://brj.asu.edu/v2412/articles/ar2.html>.
- Gandara, P. & Merino, B. (1993). Measuring the outcomes of LEP programs: Test scores, exit rates, and other mythological data. *Educational Evaluation and Policy Analysis*, 15(3), 320-338.

Hakuta, K. (2002). *What Can We Learn About the Impact of Proposition 227 from SAT-9 Scores?*

Retrieved May 14, 2003, <http://www.stanford.edu/%7Ehakuta/SAT9/index.htm>.

Hakuta, K., Butler, Y., & Witt D. (2000). *How long does it take English learners to attain proficiency?*

Santa Barbara: University of California Linguistic Minority Research Institute (UCLMRI)

Policy Report 2000-1. Retrieved May 14, 2003, from <http://lmri.ucsb.edu/index.htm>.

Linn, R.L., Graue, M.E., & Sanders, M.N. (1990). Comparing district and state test results to national norms: Interpretations of scoring "above the national average." *CSE Technical Report 308*.

University of California, Los Angeles: National Center for Research, Standards, and Student Testing (CRESST). Retrieved April 2, 2004, from

http://cresst96.cse.ucla.edu/andmore_set.htm.

Linguanti, R. (2001). *The redesignation dilemma: Challenges and choices in fostering meaningful accountability for English learners*.

Santa Barbara: University of California Linguistic Minority Research

Institute (UCLMRI) Policy Report 2000-1. Retrieved May 14, 2003, from

<http://lmri.ucsb.edu/index.htm>.

Mora, J.K. (2000). *Proposition 227's second anniversary: Triumph or travesty?* Retrieved March 24, 2004,

from <http://coe.sdsu.edu/people/jmora/Prop227/227YearTwo.htm>.

Nishioka, J. (1999, June 25). Scores show Prop. 227 works, Unz says. *AsianWeek*, 20(3). Retrieved

April 2, 2004, from <http://www.asianweek.com>.

Parrish, T., Linguanti, R., Merickel, A., Quick, H., Laird, J., & Esra, P. (2002). Effects of the

implementation of Proposition 227 on the education of English learners, K-12: Year 2 Report.

Palo Alto and San Francisco: American Institutes for Research and WestEd. Retrieved May 14,

2003, from <http://lmri.ucsb.edu/index.htm>.

Ron Unz exposes folly of bilingualism. (2001, June 6). NewsMax.com. Retrieved March 29, 2004,

from <http://www.newsmax.com/archives.shtml>.

- Sailer, S. (2002, September, 25). Q&A with Ron Unz on bilingual education. United Press International. Retrieved March 29, 2004, from <http://www.upi.com>.
- Shepard, L.A., Flexer, R.J., Hiebert, E.H., Marion, S.F., Mayfield, V., & Weston, T.J. (1996). Effects of introducing classroom performance assessments on student learning. *Education Measurement: Issues and Practices*, 15 (3), 7-18.
- Unz, R. (1997, October 19). Bilingual is a damaging myth. *Los Angeles Times*.
- Unz, R. (2001, October 26). Rocks falling upward at Harvard University. *National Review Online*. Retrieved March 29, 2004, from <http://www.english4children.org/0110/102601.htm>.
- Weintraub, D.M. & Chey, E. (1999, July 1). Computer glitch clouds any gauge of Prop. 227's effectiveness. *Orange County Register*.

About the Author

James B. Grissom

Standards and Assessment Division
 California Department Education
 1430 N Street
 Sacramento CA 95814
 (916) 319-0361
 E-mail: jgrissom@cde.ca.gov

B.A. California State University, San Bernardino, 1970
 M.A. University of Colorado, Boulder, 1984
 Ph.D. University of Colorado, Boulder, 1988

Education Policy Analysis Archives

<http://epaa.asu.edu>

Editor: Gene V Glass, Arizona State University

Production Assistant: Chris Murrell, Arizona State University

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. The Commentary Editor is Casey D. Cobb: casey.cobb@uconn.edu.

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

Greg Camilli
Rutgers University

Sherman Dorn
University of South Florida

Gustavo E. Fischman
Arizona State University

Thomas F. Green
Syracuse University

Craig B. Howley
Appalachia Educational Laboratory

Patricia Fey Jarvis
Seattle, Washington

Benjamin Levin
University of Manitoba

Les McLean
University of Toronto

Michele Moses
Arizona State University

Anthony G. Rud Jr.
Purdue University

Michael Scriven
University of Auckland

Robert E. Stake
University of Illinois—UC

Terrence G. Wiley
Arizona State University

David C. Berliner
Arizona State University

Linda Darling-Hammond
Stanford University

Mark E. Fetler
California Commission on Teacher
Credentialing

Richard Garlikov
Birmingham, Alabama

Aimee Howley
Ohio University

William Hunter
University of Ontario Institute of
Technology

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

Heinrich Mintrop
University of California, Los Angeles

Gary Orfield
Harvard University

Jay Paredes Scribner
University of Missouri

Lorrie A. Shepard
University of Colorado, Boulder

Kevin Welner
University of Colorado, Boulder

John Willinsky
University of British Columbia

EPAA Spanish & Portuguese Language Editorial Board
Associate Editors

Gustavo E. Fischman
Arizona State University

&

Pablo Gentili
Laboratório de Políticas Públicas
Universidade do Estado do Rio de Janeiro

Founding Associate Editor for Spanish Language (1998—2003)

Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

Argentina

- Alejandra Birgin
Ministerio de Educación, Argentina
- Mónica Pini
Universidad Nacional de San Martín, Argentina
- Mariano Narodowski
Universidad Torcuato Di Tella, Argentina
- Daniel Suarez
Laboratorio de Políticas Públicas-Universidad de Buenos Aires, Argentina
- Marcela Mollis (1998—2003)
Universidad de Buenos Aires

Brasil

- Gaudêncio Frigotto
Professor da Faculdade de Educação e do Programa de Pós-Graduação em Educação da
Universidade Federal Fluminense, Brasil
- Vanilda Paiva
- Lilian do Valle
Universidade Estadual do Rio de Janeiro, Brasil
- Romualdo Portella do Oliveira
Universidade de São Paulo, Brasil
- Roberto Leher
Universidade Estadual do Rio de Janeiro, Brasil
- Dalila Andrade de Oliveira
Universidade Federal de Minas Gerais, Belo Horizonte, Brasil
- Nilma Limo Gomes
Universidade Federal de Minas Gerais, Belo Horizonte

- Iolanda de Oliveira
Faculdade de Educação da Universidade Federal Fluminense, Brasil

- Walter Kohan
Universidade Estadual do Rio de Janeiro, Brasil
- María Beatriz Luce (1998—2003)
Universidad Federal de Rio Grande do Sul-UFRGS
- Simon Schwartzman (1998—2003)
American Institutes for Resesarch—Brazil

Canadá

- Daniel Schugurensky
Ontario Institute for Studies in Education, University of Toronto, Canada

Chile

- Claudio Almonacid Avila
Universidad Metropolitana de Ciencias de la Educación, Chile
- María Loreto Egaña
Programa Interdisciplinario de Investigación en Educación (PIIE), Chile

España

- José Gimeno Sacristán
Catedratico en el Departamento de Didáctica y Organización Escolar de la Universidad de Valencia, España
- Mariano Fernández Enguita
Catedrático de Sociología en la Universidad de Salamanca. España
- Miguel Pereira
Catedratico Universidad de Granada, España
- Jurjo Torres Santomé
Universidad de A Coruña
- Angel Ignacio Pérez Gómez
Universidad de Málaga
- J. Félix Angulo Rasco (1998—2003)
Universidad de Cádiz
- José Contreras Domingo (1998—2003)
Universitat de Barcelona

México

- Hugo Aboites
Universidad Autónoma Metropolitana-Xochimilco, México
- Susan Street
Centro de Investigaciones y Estudios Superiores en Antropología Social Occidente,
Guadalajara, México

- Adrián Acosta
Universidad de Guadalajara
- Teresa Bracho
Centro de Investigación y Docencia Económica-CIDE
- Alejandro Canales
Universidad Nacional Autónoma de México
- Rollin Kent
Universidad Autónoma de Puebla. Puebla, México
- Javier Mendoza Rojas (1998—2003)
Universidad Nacional Autónoma de México
- Humberto Muñoz García (1998—2003)
Universidad Nacional Autónoma de México

Perú

- Sigfredo Chiroque
Instituto de Pedagogía Popular, Perú
- Grover Pango
Coordinador General del Foro Latinoamericano de Políticas Educativas, Perú

Portugal

- Antonio Teodoro
Director da Licenciatura de Ciências da Educação e do Mestrado Universidade Lusófona de Humanidades e Tecnologias, Lisboa, Portugal

USA

- Pia Lindquist Wong
California State University, Sacramento, California
- Nelly P. Stromquist
University of Southern California, Los Angeles, California
- Diana Rhoten
Social Science Research Council, New York, New York
- Daniel C. Levy
University at Albany, SUNY, Albany, New York
- Ursula Casanova
Arizona State University, Tempe, Arizona
- Erwin Epstein
Loyola University, Chicago, Illinois
- Carlos A. Torres
University of California, Los Angeles
- Josué González (1998—2003)
Arizona State University, Tempe, Arizona