Title
A Study of the Effects of Bilingual and Structured English Immersion Programs on the Oral English and Literacy Development of Students Learning English as a Second Language

Author
Wood, Danny Lee

Publication Date
2014

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

A Study of the Effects of Bilingual and Structured English Immersion Programs on the Oral English and Literacy Development of Students Learning English as a Second Language

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in Education

by

Danny Lee Wood

2014

ABSTRACT OF THE DISSERTATION

A Study of the Effects of Bilingual and Structured
English Immersion Programs on the Oral and Literacy Development
of Students Learning English as a Second Language

by

Danny Lee Wood

Doctor of Philosophy in Education

University of California, Los Angeles, 2014

Professor Concepción M. Valadez, Chair

This study had two purposes: (1) to compare two different English language learner (ELL) programs—transitional bilingual education (TBE) programs and transitional structured English immersion (SEI) programs—to identify which, if either, was superior in facilitating English language acquisition; and (2) to identify the relationship between English oral proficiency and English literacy proficiency. To answer these questions, this study analyzed English proficiency scores from the California English Language Development Test (CELDT) from a large school district in California from 2002 – 2007. The CELDT is an instrument that was aligned with the English Language Development (ELD) standards of California and first administered in 2001 to measure English language proficiency. The CELDT provided measurements in three English skill scales: oral (listening and speaking), reading,

and writing. The CELDT also produced a composite overall English proficiency score from the combined performance scale scores.

The students in the study were primarily Hispanic (98.26%) native Spanish speakers (99%) from lower socioeconomic backgrounds (91.4%). Of the 2,731 students whose scores were analyzed, 1,196 (44%) were in their respective SEI or TBE programs the entire 6-year period of the study.

For the program comparisons, there were four groups of students: (a) students who were in transitional bilingual programs for the entire length of the study, (b) students who were in transitional bilingual programs for various lengths of time before transferring to SEI programs, (c) students who were in the SEI program for the entire length of the study, and (d) students who were in the SEI program for various lengths of time before transferring to other programs or schools.

The bilingual and SEI programs in this study were analyzed in four different ways: raw CELDT score means, CELDT proficiency levels, CELDT highest scorers, and the comparison of oral to literacy skills. When CELDT means were compared, the SEI groups performed slightly better throughout the study period over all. However, the bilingual students narrowed the scoring gap each year across all skill areas, and in the case of oral skills, one of the bilingual groups scored higher than both SEI groups in 2007.

When CELDT overall and oral proficiency levels were compared, all groups finished in the same level by grade four (2006) for overall proficiency, and were the same for oral proficiency by grade five (except for the higher bilingual score just mentioned). By 2007, while SEI students slightly outperformed bilingual students each year, the proficiency scores for reading and writing were close between both programs. When scores were calculated for the percentage of students that scored at the highest levels between the groups, the SEI scores were almost always the highest. However, bilingual students out performed SEI students in 2007 on CELDT and academic measures.

The dissertation of Danny Lee Wood is approved.

Patricia Gándara

Nina M. Hyams

Reynaldo F. Macías

John D. McNeil

Concepción M. Valadez, Committee Chair

University of California, Los Angeles

2014

Dedication

I would like to dedicate this work to my wonderful family, without whose support it would have been difficult or impossible to complete! My dear wife, Naira, was a constant source of encouragement (even when I didn't want her to be!); my awesome son Armen and my amazing daughter Monja provided much needed fun diversions in numerous ways between keyboard strokes; and my super mother-in-law, Marietta, made it possible to work in a more relaxed atmosphere by helping with the children. I am particularly grateful to my late mother who always managed to convince me that I was worthwhile and capable whenever I began to think otherwise.

CONTENTS

Tables and Graphs

Chapter 5

Chapter 6

ACKNOWLEDGEMENTS

I would like to begin by thanking my advisor, Dr. Concepción Valadez, for taking an interest in getting me into the Ph.D. program at UCLA. This was a decision I made rather late in life, and her encouragement, together with the promptings of my good friend Jeff MacSwan, went a long way in convincing me to move forward with the application.

I am especially appreciative for the latitude Dr. Valadez granted me as I began taking classes. During my first quarter at UCLA I thought a useful cross-discipline department would be Applied Linguistics. It was there, under the auspices of Dr. John Schumann, and his protégé Don Favareau, that I was exposed to courses on how the brain processes language in monolinguals and multilinguals. This was my introduction to an exciting new dimension of language acquisition/learning where neuroscience intersected with the classroom. As I branched out to courses in brain science and fMRI technology, I more deeply appreciated the miracle of human language development in healthy individuals. Coursework with Dr. Susan Curtis in the Department of Linguistics also deepened my understanding of how various diseases, genetic disorders, and environment can inhibit normal language development. Dr. Nina Hyams, also in the Department of Linguistics, helped me to appreciate second and multiple language acquisition from a linguistic perspective. From these experiences I've grown to understand the need for an interdisciplinary approach to language learning and development.

I owe a special debt of gratitude to Dr. Macías for the scope and depth of his meticulous and insightful editing, and I thank Dr. McNeil and Dr. Gándara for their probing and thought-provoking questions as the drafts of this dissertation unfolded.

For financial support, I am grateful to the many stipends, grants, and loans provided through various programs at UCLA. My burden of debt would have been considerably higher without that assistance. The Office of Student Services, especially Harmeet Singh and Amy Gershon, have on numerous occasions been extremely helpful in providing information and moving behind the scenes to help meet various requirements and deadlines. Life would have been immensely more stressful without them.

For the data analyses I am deeply indebted to Dr. Phil Ender and Zachary Wagner-Rubin from the Institute for Digital Research and Education (idre) at UCLA. Dr. Ender spent countless patient hours helping me utilize and analyze the massive data sets from LAUSD that had proved to be quite unwieldy for inexperienced hands. Zachary helped me to refine Dr. Ender's work and analyze the data in additional directions. My objectives could not have been accomplished without them. In addition to being very encouraging, Dr. Gojko Vuckovic at LAUSD's Program Evaluation and Research Branch was very helpful in providing the necessary data for the research project, and helpful in making certain segments of the data more accessible.

For primal foundational support, I reserve the greatest thank you for my late mother, who wouldn't countenance self-doubt. Thanks to her I went back to finish high school, and began to set higher goals for myself in many areas. The runner up is

my great friend Randy Mundy, who over the years has been a source of inspiration on many levels, and has inspired me to think out of the box and create alternate realities for myself.

Steve and family in Chicago; Sara, Geda and Carlos; Jeff and family, and Phil and family continue to be fun refuges from the drudgeries of life.

My loving family, Naira, Armen, Monja, and Marieta have picked up the torch and made everything in life worthwhile. This is as much for them as it is for me. It's been a joy to feel the support of Naira's family here and abroad over the years: Levon, Eli, David, Lilit, Daryl, Edik, Kara (thanks for the yoga!), Armen, Nairine, Almalia, Kara, Ashot, Nairusik, Art and family, Tsolak and family, Chic and Yvonne and family, Gohar and everyone in Armenia! Thank you!

Vita

Danny Wood

EDUCATION HISTORY

Spring 1981	B.A. Near Eastern Studies Language concentration in Arabic and Hebrew Brigham Young University Provo, Utah
Spring 1885	TESL Certificate California State University, Long Beach Long Beach, California
April 1986	M.A. Linguistics (Emphasis in Applied Linguistics) Community College Teaching Credential: Humanities California State University, Long Beach Long Beach, California
December 1993	Secondary Teaching Credential English, Language Arts (with ESL add on for Pre K – Adult Ed) California State University, Dominguez Hills Carson, California

TEACHING EXPERIENCE: ESL/ENGLISH

Community College/University

1984-1991	Long Beach City College	Long Beach, California
	Cerritos Community College	Cerritos, California
1989-1990	UC Irvine Extension	Irvine, CA
Fall 1994	UCLA Extension	Los Angeles, California
June 1998	UCLA GSEIS, Center X	Los Angeles, California

Adult Education

1986-1987	Institute of Public Administration	Riyadh, Saudi Arabia
1987-1989	Cambria English Institute	Los Angeles, California

Secondary Education

1990-2014 Los Angeles USD Los Angeles, California

PROFESSIONAL ORGANIZATIONS

California TESOL
TESOL
AERA
Bilingual Education Research: A Special Interest Group of AERA
Brain, Neurosciences, and Education: A Special Interest Group of AERA
Second Language Research: A Special Interest Group of AERA

PRESENTATIONS

Spring 2003
Conference on Biosemiotics
Sponsored by the Department of Applied Linguistics and TESL
UCLA, Rolfe Auditorium
Topic: Language and the Brain

Fall 2003
Language Development Conference
Sponsored by Professor Susan Curtiss, Department of Linguistics
UCLA, Haines Conference Room
Topic: Early Language Development

Winter 2004
Bilingualism and Second Language Conference
UCLA, Haines Conference Room
Sponsored by Professor Nina Hyams, Department of Linguistics
Topic: SLI and Language Development

Winter 2006
UC Irvine
Sponsored by Professor Tugun Muftuler, Neuroimaging Lab
Presentation of research design and results of an fMRI experiment:
Topic: Nueural Substrate of Syntax in Multilinguals

PUBLICATION

Breakdown in the Semiosphere (under the pseudonym *Ak'bal Mol*)
in *Random Agenda,* vol 1, 2005
Arena Productions, Glendale, CA 91205

Chapter 1: Introduction to the Problem

The U.S. is a linguistically diverse nation and its public school systems have been charged with teaching the English language to children and youth who enroll with little or no proficiency in the language. The number of these English language learners enrolled in U.S. public schools has increased in recent years. *Education Week* (August 4, 2004, updated June 16, 2011) stated, "the number of English-language learners enrolled in public schools increased from 3.5 million to 5.3 million, or by 51 percent" in the decade between the 1997-98 school year and the 2008-2009 school year. During that same time period, the general population of students only increased 7.2 percent. While California had the highest number of and highest proportion of these students, they were distributed throughout the country.

> In 2010-11, states in the West had the highest percentages of ELL students in their public schools. In 8 states, 10 percent or more of public school students were English language learners—Oregon, Hawaii, Alaska, Colorado, Texas, New Mexico, Nevada, and California (California data were inputted from 2009-10 data)—with ELL students constituting 29 percent of public school enrollment in California. Thirteen states and the District of Columbia had percentages of ELL public school enrollment between 6 and 9.9 percent. In addition to the District of Columbia, these states were Oklahoma, Arkansas, Massachusetts, Nebraska, North Carolina, Virginia, Arizona, Utah, New York, Kansas, Illinois, Washington, and Florida. The percentage of ELL students in public schools was less than 3 percent in 13 states; this percentage was between 3 and 5.9 percent in 16 states. The percentage of ELL students in

> public schools was higher in 2010-11 than in 2009-10 in just over half of the states (28 states), with the largest increase in percentage points occurring in Nevada (3 percentage points) and the largest decrease in percentage points occurring in Minnesota (2 percentage points). (U.S. Department of Education, National Center for Education Statistics, 2013)

Clearly, many states that have not traditionally been associated with large-scale immigration are receiving immigrants in greater numbers, although all ELL students are not foreign born.

ELL students create a tremendous challenge for school systems across the nation. They must insure that these students learn English *and* the academic content of the school curriculum. How to do this has been the subject of controversy for over 100 years. Federal and state legislatures as well as federal and state courts have addressed the schooling of students whose first language was not English. The language medium of instruction and government policies of assimilation have been issues of public policy making and debates.

1.1 National, State, and Judicial Influences on English Language Learners

Federal and state laws, state propositions, and court rulings have governed the history of English language learner (ELL) instruction over the past 100 years, with laws adopted by a majority of states that made English the excusive language of instruction, and prohibited the teaching of foreign languages before puberty. The mid 20th century saw a liberalization of these policies. Congress passed the Bilingual Education Act in 1968, a discretionary program to experiment with bilingual forms

of instruction, to create equal education opportunity for ELLs (Wiese & Garcia, 1998).

In 1974, the US Supreme Court ruled on a California class action suit, *Lau v. Nichols* (Crawford, 1999), filed on behalf of Chinese-speaking ELL students in the San Francisco Unified School District. The Court ruled on behalf of the plaintiffs stating that students who were not proficient in English did not have equal access to the curriculum and therefore were being discriminated against. Also, in 1974 Congress passed the Equal Educational Opportunities Act (Crawford, 1999), which required school districts to take action to overcome barriers to students' equal participation in the curriculum.

In 1981, the U.S. Fifth Court of Appeals, in *Castañeda v. Pickard* (Hakuta, 2011), set criteria to determine the adequacy of program services to students with limited English proficiency. The Court established three criteria for review: (1) the program must be based on sound educational theory or principles, (2) it must effectively implement this program, and (3) it must produce results indicating that it is working. This language would later resurface in the federal legislation known as the No Child Left Behind Act of 2002.

In 1983, in *Keyes v. School District No. 1* (Alexander & Alexander, 2004), a federal court ruled that teachers who don't have the necessary bilingual skills to communicate with their students were not in compliance with Castañeda's second criteria. And in 1987, a federal court stipulated in *Gómez v. Illinois State Board of Education* that a state education agency could be sued for not taking the "appropriate action" required by the *Castañeda* decision (Alexander & Alexander,

2004).

1.1.2 California

In 1976, the California legislature passed the Chacone-Moscone Bilingual-Bicultural Education Act. All school districts were required to offer services to all ELLs, which was most commonly, but not exclusively, primary language instruction (Crawford, 1999).

California passed two laws in the 1980s that provided assistance to school districts with large ELL populations. The Bilingual Teacher Training Assistance Program (Crawford, 1999) was passed in 1981 to provide training for teachers who had been granted bilingual teacher waivers. In 1984 the Impacted Languages Act (Crawford, 1999) was enacted to provide assistance for districts impacted by large refugee and ELL populations.

During the latter 1980s and 1990s California experienced increased political controversy over education and immigration issues. In 1986 Proposition 63 (Cal Prop 63, 2014) was enacted, making English the "official language of California." In 1994 California voters passed Proposition 187 (Cal Prop 187, 2014), which was later overturned, making it illegal for undocumented students to attend public schools. In 1996 Proposition 209 (Cal Prop 209, 2014), which opponents argued would discriminate against the disadvantaged and subvert affirmative action policies in California, was passed. Proposition 227 (Crawford, 1999), seen basically as an anti-bilingual education measure, was passed two years later in 1998.

In the 1990s and early part of the 2000s, school accountability and school

reform were placed high on state and national policy agendas, which would effect changes in ELL education. One important reform in California schools was the state's class size reduction initiative, the philosophy being smaller class sizes would make it possible for teachers to have a greater influence on their students.

To monitor academic progress across the state and to promote school accountability, an annual spring standardized assessment was started in 1998. The Stanford Achievement Test 9 (SAT-9), a standardized test of student academic achievement in the core subject areas of math, English, social studies, and science, etc., was administered starting in the spring of 1998 to every student in the state.

The introduction of the SAT-9 was followed in 1999 by California's Public Schools Accountability Act (PSAA). The Act was designed to enhance and more clearly define school responsibility regarding the performance of their students on statewide tests (initially the SAT-9). The PSAA required schools to demonstrate that their students were making academic progress, with the individual school as the primary unit of accountability.

The PSAA had three components. First, the Academic Performance Index (API), a composite score of student performance for the school on several standardized tests. The second component was the Intermediate Intervention/Underperforming Schools Program (II/USP), which provided funds to support low-performing schools. The third component was the Governor's Performance Award (GPA), which provided financial rewards to schools that met API improvement targets and other performance goals.

Also in 1999 California adopted English Language Development (ELD)

standards, which were guidelines defining skills ELLs needed to acquire to become age and grade proficient in English (California Department of Education, Standardized Testing and Reporting (STAR) Results, 2013). Related to the ELD standards was the development of the California English Language Development Test (CELDT). First administered in 2001, The CELDT was used to identify English learners, assess their level of English proficiency, and monitor their progress toward becoming fluent English proficient.

In 2003, the SAT was replaced by the California Achievement Test (6th Edition) or CAT/6. Other significant academic standardized tests were the California Standards Tests (CST) and the California High School Exit Examination (CAHSEE), which appeared in 2001. Replacing the SAT, the CST was basically an updated academic content area state standardized test. The CAHSEE was created to ensure that students had attained basic proficiency in English reading, English writing, and mathematics before they were allowed to graduate from high school (this was in addition to curricular graduation requirements).

In addition, while California schools were still adapting to the PSAA, the federal *No Child Left Behind Act* (NCLB) (Alexander & Alexander, 2004) was signed into law in 2002, creating similar national accountability standards as a condition of federal funding. While holding schools accountable, NCLB also made local districts and states accountable for student performance. A significant provision of the Act was the requirement of "adequate yearly progress" (AYP), an annual measure of student participation and achievement on state-wide assessments and other academic achievement indicators to which districts and schools were held

accountable. Like PSAA, NCLB called for public reporting of schools' student performance.

This legislation accentuated the focus and heightened attention on the academic performance of ELLs. NCLB mandated reporting on the performance of specific subgroups within a school, including English learners. Under NCLB's Title III (the revised version of the Bilingual Education Act), states were required to set specific goals, called annual measurable achievement objectives (AMAOs). English learners had to make annual progress toward English proficiency, and an increasing percentage of a district's ELLs had to become proficient in English each year. California's AMAOs focus on the percentage of ELLs moving up on level of proficiency on the CELDT, the percentage attaining English proficiency (as measured by the CELDT), and the percentage of ELLs meeting AYP requirements at the school level (California Department of Education, Standardized Testing and Reporting (STAR) Results, 2013).

Clearly, many political forces have come to bear on ELL education in the nation. While bilingual education has been a factor in educating ELLs, more recent political trends have begun to call for a stronger role for the use of English to teach English to ELLS, with a strong de-emphasis in bilingual programs (discussed more fully in Chapter 2 and Chapter 5) to do so. The following section emphasizes ELL challenges in California.

1.2 Programs for ELLs

Over the years there has been a persistent achievement gap on standardized tests between the academic performance of ELLs and their monolingual English-speaking peers. Understandably, most ELLs perform below grade level on these English standardized tests. A Pew Hispanic Center (June 6, 2007) report stated "that nearly half (46%) of 4th grade students in the English language learner (ELL) category scored "below basic" in mathematics in 2005—the lowest level possible. Nearly three quarters (73%) scored below basic in reading. In middle school achievement in mathematics was lower still, with more than two-thirds (71%) of 8th grade ELL students scoring below basic. Meanwhile, the same share (71%) of 8th grade ELL students scored below basic in reading."

Some researchers (Shaftel, Belton-Kocher, Glasnapp, & Poggio, 2006; Solano-Flores, 2008; Martiniello, 2009; Wolf & Leon, 2009) have suggested that since these tests were in English, a language in which these students were not proficient, that the achievement gap exists at least in part because of the language complexity of the tests. In other words, if the language of the tests were simpler in English structure, ELL students would likely perform at a higher level. This introduces a dichotomy in learning development. On the one hand, students are expected to learn English, the language of instruction in U.S. schools and the language of communication in the broader U.S. society. On the other hand, they are also required to learn academic content in math, science, and social studies, as well as developing academic skills in English literacy.

While they cannot be completely separated, there is clearly a difference between language ability and academic learning. For native speakers of a language, all normal children of a given society will become age competent in that language by the age of five, knowing its basic structures, a useful vocabulary, and the appropriate social rules of interaction. However, not all children who enter school with native language competence end up with the same academic proficiency over time. Factors such as I.Q., the type of schooling opportunities they are exposed to, family wealth, individual interests, and individual motivation, together with other factors, all contribute to an individual's level of academic growth (Cummins 2000a). Under NCLB all states provide assessments for ELL English language growth and academic achievement growth.

In California the instrument that measures English language growth is the California English Development Test (CELDT). It has been developed to measure how native-like a student's language has become (California Department of Education, CELDT, 2013). Students who take this exam are placed in one of five levels: (1) beginning, (2) early intermediate, (3) intermediate, (4) early advanced, or (4) advanced. These levels are developed from a composite score that measures listening (aural) skills; speaking skills; reading skills; and writing skills. Once students obtain an overall "early advanced" level on the CELDT, the state of California considers the students "proficient" in English. This means that students are thought competent to do as well as native speakers when in a regular classroom that is taught in English. However, even when students achieve the "advanced" level on the CELDT, it is recognized that their English skills will not be completely native-

like, and that students' linguistic competence will increase as they continue to develop English skills (California Department of Education, CELDT, 2013).

California measures academic growth annually for all students by administering the standardized California Standards Test (CST). Both native English-speaking students and ELLs take this test in English. A central question for policy makers, administrators, and teachers has been: How *much* English do students need to be prepared for regular academic English instruction in the classroom? Another question is: *How much time* does it take a non-English proficient student to learn sufficient English for them to be prepared to perform successfully academically in an English language classroom? Central to these two questions are: What is the best way to teach English to non-English proficient students? Is it better to teach students academic content in their native language while simultaneously teaching them English (a bilingual approach)? or Is it better to instruct students in a structured English immersion medium to accomplish these two ends?

Many states employ some type of bilingual education to teach students academic content in their native language while at the same time teaching them English. These programs tend to be of two types: transitional bilingual education instruction, and long-term bilingual programs. Early exit transitional bilingual education programs tend to use only the student's native language in the early years of instruction, then gradually introduce more English instruction. Early exit bilingual programs may last two to three years. Long-term bilingual education programs, or "late exit bilingual transitional programs" are from 4 to 5 years. Two-way dual language programs involve students who speak a minority language and

students who speak the majority language of a given community. These programs seek to provide language and content academic instruction in both groups' native language. Students who participate in either dual language program will ideally be equally competent in two languages for everyday communication skills and for academic learning by the end of elementary school (when starting the program in Kindergarten).

Many states also offer some type of structured English immersion instruction designed to take into account the needs of ELL students. Structured English Immersion (SEI) is one approach to teaching ELLs English. Typically, these programs use only English in English language instruction while sometimes also teaching content area subject matter (i.e. Math, Science, Social Studies, etc.) in "sheltered English." Sheltered English means teaching in English that is (1) highly contextualized, (2) contains many visual examples, (3) has simpler vocabulary, (4) increased redundancy, (5) adjusts to the levels of proficiency of the student, and (6) otherwise contains "comprehensible input" (or I + 1, see Krashen, 2004).

Other programs remove students for alternate instructional periods (ESL pullout) during the day to teach English language skills. The regular instructional periods generally provide no special language assistance, meaning they are with "mainstream fluent English language students" (generally students whose native language is English). ELLs who receive no language services at all during the day are said to be in a "submersion" environment.

District X, which provides the data base for this study, is a large urban school district in California which offers four possible programs (or options) to help ELL

students learn English. One is the dual language immersion program, which is offered in Chinese, Korean, and Spanish. Another is the transitional bilingual program, which is mostly available in Spanish. Structured English immersion is one English-only program available to ELLs. The other English–only option available to students is what is often termed "submersion." LAUSD considers them "mainstreamed." This is when an ELL student, at the parent's request, is placed in a classroom designed for native English speakers (or children who have demonstrated a high level of fluency in English). It is also sometimes referred to as the "sink-or-swim" method. Ostensibly, like other "mainstream" (native English, or highly fluent English) students, submersion students receive no special language program services. However, in District X all teachers must have some training to enable them help students with low English proficiency, and all students who are not proficient in English must take the CELDT annually.

For purposes of long-term language program evaluation in District X, the transitional bilingual program and the structured English immersion program were the only possibilities for comparison study. The numbers of students in the dual language bilingual programs and the numbers of students who receive no language program services (submersion students) were too small. This study therefore seeks to identify which program type, transitional bilingual education (TBE) or structured English immersion (SEI), is more effective in helping ELL students acquire English proficiency.

1.3 Focus of This Study

Noting the need for programs that can best help the burgeoning number of ELLs in this country, and the need for more *scientific* research, this study explored the relevant and ongoing question, "Which language program type is more beneficial for English second language acquisition: transitional bilingual education (TBE) or structured English immersion (SEI)?" Another question this study seeks to answer is, "What is the relationship of oral language proficiency to the development of English literacy skills?" As Chapter 2 shows, this second question has received little attention in the research literature, and may be important (Chapter 5) in developing programs to help ELLs.

To answer these questions, this study sought to avoid the pitfalls of earlier research. This study was of a single site; it was self-contained and relied on its own results, so it avoided any dangers of *synthesizing results from disparate studies*.

While the atmosphere surrounding these programs in District X was *politically charged* following the passage of the controversial Proposition 227 (1998), District X decided to comply with the law and implemented its directives. Local school administrators began to enroll ELL students in SEI programs as the default program for ELL students unless parents waived this option and requested bilingual programs at local school sites. Compliance was district-wide.

By using only data from one site, this study avoided *inconsistently applied program labels*. District X's master plan for language program implementation was well defined and implemented consistently across the district (see Los Angeles Unified School District, English Learner Master Plan, 2013) for similar policy

objectives. There is a clear difference in expectation between transitional bilingual programs, and English immersion programs. Bilingual programs are to teach academic content subject matter in the students' native language in the initial years of the program and gradually introduce more English into the academic curriculum. Table 1 below illustrates how an *early exit* transitional bilingual education program would function. All of the students in one analysis cohort for this study were in a *late exit* transitional bilingual program, and many of the students in the other analysis cohort for this study were in a *late exit* transitional bilingual program for the entire six-year period of the study (2002-2007).

Table 1 Idealized Schedule for an Early Exit Transitional Bilingual Education Program

Subject Taught in the Primary Language				**Subjects Taught in English**			
Grade	Primary Language Arts Reading Writing	Math	Science Social Science	English Language Develop-ment (ELD)	English Language Arts Reading Writing	Math	Science Social Science
Kinder	130	60	40	60	0	0	0
1st	100	45	40	60	30	15	0
2nd	90	20	10	60	60	30	20
3rd	60	0	0	60	90	50	30
4th	0	0	0	45	150	50	45

Source: LAUSD

Students in the structured English immersion programs may receive some help as needed in their native language, but nearly all instruction is to occur in English.

Several factors helped control for *pre-existing differences between students* and *appropriate comparisons between groups*. First, each student began their

respective programs at the same level of non-proficiency in English as measured by the CELDT. Secondly, all students were enrolled in Kindergarten in the same school year, which closely controlled for age. All students in the bilingual program were Hispanic and native Spanish-speakers, and 98% of students in the SEI programs were Hispanic native Spanish-speakers. All students in the study started schooling in District X, and had experienced no other formal schooling in the United States or another country. Also, most of the students were from a similar low socioeconomic class.

To ensure that *standardized tests* were used, the CELDT was determined to be the instrument to measure English proficiency across program types. All students took the same test each year, and at the same time of the year, for the full six years of the study.

The students in this study were all in the same school district, and therefore had access to the *same instructional materials* and the *same district mandated expectations in their relative programs*. Therefore, the *schooling environment* was fairly similar between groups. Other control variables used included the local mini-district (a proxy for "best practices") where the students attended school, and the type of school the student attended (i.e., magnet, elementary, span magnet, etc., also a proxy for "best practices").

Some studies have been criticized for *mixing data*. Only CELDT data was used in this study, and all data that was analyzed and presented was explicitly identified. For example reading and writing data would not be presented as a combined variable unless explicitly stated.

This study used *sufficient numbers of students* to generate *statistically significant results*. Also, the student scores were taken over a six-year period of time (2002-2007) *controlling for length of study*, sufficiently long to have confidence in the validity of the results. The study also ran regressions on the student data to compensate for the unavailability of some data, such as *teacher program implementation* and *language use* both in and out of the classroom.

By addressing the concerns listed above, this study met the stringent requirements for scientific research set forth by the National Research Council (NRC) and others. It is however, delimited by several factors. It is a study of only one school district in the state, so it does not pretend to speak descriptively on a general basis, but to explore the research question on program effectiveness.

This study was also an analysis of aggregate statistical data and does not speak to the details or relationships or interactions that qualitative studies appropriately address. It only analyzes data from students in Kindergarten through grade 5, but doesn't represent the secondary years, which implicates a presumption of length of time for program effects. This might apply to students who enter U.S. schools after K or 1st grade, who may, or may not, have had prior schooling.

The dissertation is organized into six chapters. Chapter 1 is the introduction. Chapter 2 reviews research literature on program effectiveness and the relationship between orality and literacy. Chapter 3 lays out the research design, methods, subjects, instrumentation, and analysis. Chapter 4 presents the results to the two research questions. Chapter 5 discusses their salience within the literature discussed in Chapter 2. Finally, Chapter 6 summarizes the study and its conclusions.

Chapter 2: Literature Review

As the purpose of this study is to compare the effectiveness of transitional bilingual programs to structured English immersion programs, this chapter will review previous research studies with similar goals. In the 1970s Federal/National Policies pushed for a *magic bullet* that would identify the best programs to teach English to ELLs. They wanted *proof* that bilingual education worked, and they wanted to see theories developed to that end. This took the form of program evaluations, literature reviews, research studies, meta-analyses, and secondary analyses of national and state data sets, e.g., the National Assessment of Educational Progress (NAEP). The following section outlines some of the concerns that research reviewers and researches have with some studies.

2.1 Problems Encountered in the Research on Program Effectiveness

As policy makers look to researchers for guidance in establishing education guidelines, they have met discrepancies in the results of studies that have tried to identify language programs that can best facilitate academic success with ELL students. Much of the problem has been leveled at the quality of research being done. According to de Cos (1999), few studies of services for English learners in the United States are considered scientific (i.e., have a good experimental design), and few provide conclusive information on which instructional programs serving English learners are effective (see also Greene, 1998; Rolstad, Mahoney, & Glass, 2005b; Rossell & Kuder, 2005).

The National Research Council (NRC), in its 1997 review of the research of

programs serving English learners, acknowledged the limitations of the research conducted in the field (August & Hakuta, 1997). The NRC report discussed the difficulties involved in synthesizing results across studies, stating that this is partly due to the highly politicized character of the field resulting in inconsistently applied program labels. Of particular concern were program evaluation studies that lacked random assignment of subjects or controls for pre-existing differences, made difficult after 1974 when standards of adequate services were made legally mandatory.

Another NRC concern was that some studies failed to compare appropriate groups or programs together. Slavin and Cheung (2003) echoed this in their criticism of Rossell and Baker (1996), for comparing the results from Canadian immersion programs to US immersion programs, which are quite different. In a 2005 meta-analysis, Rolstad, et al. (2005b) discovered that one study they considered (Rossell, 1990), compared the results of a mixed Asian group to a Hispanic group.

Krashen (2005) summarized 16 studies that compared various types of bilingual programs and English programs for ELLs. Most of the studies showed results more favorable to bilingual than to English programs, but each of the studies had methodological flaws. The most pervasive problem, found in eleven of the studies, was not establishing the initial English competence of the students, making it difficult to measure progress. The next most prevalent problem, found in half of the studies, was the small number of subjects observed, making it difficult to generalize the results to a wider population. Five of the studies only monitored

progress for a short time, 1 – 2 years (or less), again making it difficult to make long-term predictions or program recommendations. The other types of methodological problems included (a) the use of non-standardized tests, (b) dissimilar schooling environment between groups (i.e., supplies, materials), (c) lack of control variables, (d) no controls for program integrity, and (e) the mixing of data from different studies. Half of these studies had from 3 to 5 of these methodological problems, while the other half had from 1 to 2 methods flaws. Krashen points out in this paper that these types of errors also plague research that is used to promote structured English immersion ELL programs.

2.2 Studies that Measure Levels of English Acquisition or Proficiency Based on Standardized Academic Tests Designed for Fluent English Speakers

In an effort to review the literature on bilingual and English programs for the Department of Education, Baker and de Kanter (1981) examined over 300 studies. To ensure that a strict set of criteria was met, they selected only 28 of those for review. The authors compared transitional bilingual (TBE) programs to submersion, English as a second language (ESL), and structured immersion (SEI) programs. Students in ESL received daily class time instruction to learn the English language, while the rest of the day was spent in mainstream classes. In the SEI classes, students received sheltered English content instruction. Their review of these carefully selected studies did not produce a clear winner. Some of the studies indicated that TBE was superior to or as good as SEI. Other studies showed ESL to be superior to, or equal to both TBE and SEI.

A longitudinal study by Gersten and Woodward conducted between 1985 and

1997 in El Paso, Texas, compared the outcomes of English learners in transitional bilingual programs (i.e., those that provide initial instruction in the students' home language, with rapid transition into English exclusively) and bilingual immersion programs (i.e., those integrating English instruction with some maintenance of Spanish content and language instruction). Initial differences found in English reading and English language favoring the bilingual immersion program disappeared by the seventh grade. In fact, by seventh grade many English learners in both program models were not meeting grade-level achievement, as measured by the Iowa Test of Basic Skills, in either English reading comprehension or English vocabulary. A follow-up at the high-school level indicated high attrition rates for students in both programs and comparable low achievement rates (de Cos, 1999).

Ramírez, Yuen, and Ramey (1991) compared Structured English immersion programs with early exit (1–2 years) and late-exit (4–6 years) transitional bilingual education programs. They followed 2,352 Spanish-speaking students over a four-year period from Kindergarten through fourth grade. The study found that while early-exit transitional bilingual students initially outperformed English immersion students in mathematics and reading in English, by the end of the third grade their advantage had essentially disappeared. The authors were unable to directly compare the late-exit transitional bilingual education programs with the early-exit transitional bilingual education programs and English immersion programs, and they therefore relied on indirect comparisons which have since been questioned by the National Research Council (Meyer & Feinberg, 1992). Hakuta (2011) maintained

that the comparison of the early exit transitional bilingual education programs to the structured English immersion programs were valid.

In 1996 Rossell and Baker conducted a large review of program effectiveness studies. They chose 72 studies comparing bilingual to Structured English immersion programs. This was an update of the 28 studies evaluated by Baker and de Kanter (1981), plus an additional 44 studies. The results of this review led them to conclude, as did Baker and de Kanter, that the research did not clearly support one program over another.

Long-term research by Thomas and Collier from 1985-2001 highlights possible shortcomings of research examining the effectiveness of program models (Thomas & Collier, 2002, 2004). They analyzed the education services provided to over 210,000 language minority students in US public schools and the resulting long-term academic achievement of these students. This was accomplished by examining in depth a review of their research on five urban and rural sites from throughout the USA for 5 years, from 1996 to 2001. The school bilingual program types examined within these contexts varied widely—they included full immersion programs in a minority language, dual-medium or two-way programs, where both minority and majority languages (usually Spanish and English) were used as mediums of instruction, transitional bilingual education programs, ESL (English as a second language) programs, and English submersion (English-only) programs.

The authors maintained that examination of language minority students' achievement over a one to four year period is too short and leads to an inaccurate perception of actual long-term performance. Through their long-term approach to

examining the English reading and math achievement of K-12 English learners, they found that when exiting ESL or bilingual services, students who had been submersed in the English mainstream performed better than those that received bilingual instruction. This trend reversed by the time students reached high school, with formerly bilingual education students showing higher performance than English-only instructed students. The strongest gains were most evident in those programs where the child's native language (first language, or L1) was a language of instruction for an extended period of time. In other words, Thomas and Collier (2004) found that the strongest predictor of student achievement in English (the students' second language, or L2) was the amount of formal L1 schooling they experienced. Only one-way and two-way or dual immersion programs achieved the highest results. Length of L1 education turned out to be more influential than any other factor in predicting the educational success of bilingual education students, including socioeconomic status.

Slavin and Cheung (2005) also reviewed a large number of studies. They chose 16 studies, which met strict criteria spanning a 29-year period (1971-2000) that compared various bilingual program types and English-only reading scores. While in some instances the English immersion results were about the same or slightly higher than the transitional bilingual results, in the majority of the cases the bilingual programs produced higher scores.

2.3 Meta-analyses

Meta-analyses are statistical procedures that researchers use to combine and

analyze data from several different studies using stringent criteria for the combining of data to be able to do the new analysis of the lager aggregate data set. The first meta-analysis on bilingual education program effectiveness was conducted in 1982 by Okada et al. at the behest of the United States Department of Education (cited in Baker, 1987; Lam, 1992; and Rolstad, Mahoney, & Glass, 2005b). Rolstad et. al, (2005b) state that in 1982 and 1983 Okada, Besel, Glass, Montoya-Tannatt, and Bachelor conducted a meta-analysis of more than 1,000 evaluations of bilingual education projects and concluded that inferences as to the effectiveness of bilingual education programs could not be made. According to Baker (1987), who worked in the Department of Education at the time, Okada et al. identified two major problems in the bilingual evaluation literature that made it impossible to estimate program effects. First, extensive data were missing, especially with respect to the nature of the treatment. Second, the nature of bilingual programs made it difficult to disentangle program effects from other sources of learning English (for details, see Baker & de Kanter, 1983; Baker & Pelavin, 1984).

However, according to Rolstad et. al (2005b) Okada et al. revealed that children who received bilingual instructional support progressed at nearly twice the national norm in English reading, math, and English language arts, with the strongest effects found in the early grades. Moreover, progress was greater when teachers were bilingual rather than monolingual speakers of English. Lam (1992) concurred with Baker's version of the report.

In 1985 Willig wanted to see if Baker and de Kanter's (1981) literature review would yield the same results when applied to meta-analysis. To make the

analysis more applicable to students in the United States, Willig (1985) decided not to include foreign studies. This reduced the number of studies to 23 from 28 in Baker and de Kanter's (1983) literature review. Whereas Baker and de Kanter (1981) couldn't make an overall judgment as to which kind of program was superior, Willig's (1985) results showed a positive effect for bilingual programs over Structured English immersion programs.

Greene (1998) subjected Rossell and Baker's (1996) review to a meta-analysis. However, in an effort to produce even more accurate results, Greene heavily pared back the number of studies to be analyzed (from 72 to 11). Unlike Rossell and Baker's conclusion that a clearly superior program could not be identified, Greene asserted that bilingual programs were more effective than English-only programs.

Rolstad, Mahoney, and Glass, (2005b) presented the next major meta-analysis in the field. They searched for studies that compared bilingual and Structured English immersion programs from 1985 to 1996. Wanting to cast a wider net than Willig and Green, they sought to be more inclusive in their selection of studies. Nevertheless, after identifying 300 studies, they chose only 17 to be in their analysis. The comparisons were between dual language, transitional bilingual, structured English immersion and/or English as a second language (which were similar), and submersion programs. Their findings were that long-term bilingual education programs were superior to short-term bilingual programs, and that all bilingual programs were superior to Structured English immersion programs.

In 2008, Rolstad et al. revisited their 2005 analysis when it became apparent

that there were errors in one of the studies that they had encountered, Gersten, 1985, which they had considered an outlier. Reanalysis confirmed and strengthened their earlier conclusions.

Rolstad et al. (2005b) published the results of another meta-analysis that focused primarily on transitional bilingual and SEI programs in the Arizona school systems. While their corpus of four studies was much smaller, the results reflected their nationwide review: bilingual programs were better than Structured English immersion programs. More recently the authors (Rolstad et. al, 2012) reaffirmed these studies with an additional emphasis that Structured English immersion programs were not able to prepare the majority of ELLs for mainstream instruction within a year of initial exposure to English.

Two meta-analyses were published in 2006. One was by the National Literacy Panel (NLP) (Goldenberg, 2008) sponsored by the U.S. Department of Education's Institute of Education Sciences, which reviewed nearly 300 research documents. The other was by the Center for Research on Education, Diversity, and Excellence (CREDE) (Goldenberg, 2008), which included nearly 200 research documents. Goldenberg (2008) stated the research supported the concept that L1 instruction for reading in Spanish was more effective in L2 reading development for ELLs than L2 reading instruction without L1 literacy development.

2.4 Post Proposition Studies

Sentiment against bilingual programs has been building over the last several years. Its roots are in earlier studies on the effects of bilingualism. Perhaps the first

of these, the American Institutes for Research's (AIR) evaluation of bilingual education programs, was commissioned in the 1970s by the United States Office of Education (Danoff, Coles, McLaughlin, and Reynolds, 1978). It provided an overview of US federally funded bilingual programs operating at the time and found that such programs had no significant impact on educational achievement in English, although they did enhance native-like proficiency. It furthermore suggested that pupils were being kept in bilingual programs longer than necessary, thus contributing to the segregation of such students from mainstream English classes.

Despite concerns about its methodology, the conclusions of the AIR study were seemingly replicated by a second US federally commissioned research by Baker and de Kanter (1981, 1983; see also Rossell and Baker, 1996). They reviewed the literature and likewise concluded that bilingual education was not advancing the English language skills and academic achievements of minority language students, predominantly Spanish-speaking L1 students. In short, Baker and de Kanter argued that students in bilingual programs demonstrated no clear educational advantages over those in English only programs.

To date, three states have adopted anti- or reduced-bilingual education measures: California (Proposition 227 in 1998), Arizona (Proposition 203 in 2000), Massachusetts (Question 2, in 2002). Some researchers have examined the effects of these policies on English language learning. Once these policies were implemented, it became possible to compare the rates of English language acquisition in states that maintained bilingual education with the rates of English acquisition in states that abolished bilingual education.

According to Jepsen (2009) the forced change from bilingual education following Proposition 227 seems to have slightly decreased average academic performance for ELL in grades 3 and 4 and not affected ELL academic achievement in grades 5 through 8, except for a negative effect on math in grades 5 and 6. Lindholm-Leary and Genesee (2010) declared that more exposure to English in California schools since Proposition 227 has produced no improvement in English achievement levels.

Rumberger and Tran (2010) conducted an analysis of National Assessment of Educational Progress (NAEP) data across 50 states to assess the achievement gap between EL and EO students using the state level data from the 2005 NAEP. They found that the state and its policies towards ELLs had a significant effect on the achievement gap. Results from their study showed that states with bilingual instruction (New Mexico and Texas) tended to have smaller achievement gaps than those states that had implemented Structured English immersion instruction (Arizona, California, and Massachusetts).

Another analysis of NAEP data Losen (2010) looked at the average math scores for ELL students in those states that had adopted English only policies and for ELLs in the nation as a whole, as well as all other states except the Structured English immersion states (since those states combined accounted for about 40% of all ELL students nationally). Losen (2010) found that English learners in the English only, or SEI states, on average performed worse than ELLs across the nation, and that Arizona ELLs performed significantly worse than both all states and even the other English only states on the national metric (NAEP) available for comparison.

2.5 Studies Using Oral Proficiency to Assess English Acquisition

The second research objective of this study is to link the acquisition of oral skills to the acquisition of literacy skills (reading and writing) across program types. It has been difficult for researchers to establish clear links between oral English proficiency and academic proficiency.

One difficulty is in finding agreement between the different instruments that evaluate oral skill level. Much of the research has centered on oral skills as the main construct for measuring students' growth in English proficiency. Katz, Low, Stack & Tsang (2004) report that an examination of oral English language proficiency tests (Graham and Acosta, 1979) found a lack of consistency in the results from different measures of oral English ability. They found correlations ranging from .39 to .82 between the Bilingual Inventory of Natural Language (BINL), the Language Assessment Battery (LAB), the Language Assessment Scales (LAS), and the Bilingual Syntax Measure (BSM). When they examined whether there was consistency in the way pairs of tests classified students, they found averages of 56% for the BINL and BSM, 65% for the BINL and LAS, and 77% for the BSM and LAS. Also in Katz et al. (2004) Gilmore and Dickerson (1979) compared the classification of students among five tests used in Texas - BINL, BSM, LAS, the Primary Acquisition of Language (PAL), and the Schutt Primary Language Indicator Test (SPLIT) and also found lack of agreement in the classification of students.

Not only is the consistency too low between instruments that measure oral proficiency, but oral English language ability seems only mildly related to academic achievement. Katz et al. (2004) cite DeAvila, Cervantes, and Duncan (1978), Ulibarri,

Spencer, and Rivas study (1981), and Saville-Troike (1984) attempting to link oral proficiency from the BSM, LAS, and BINL to various other standardized achievement tests, such as the Comprehensive Tests of Basic Skills (CTBS), California Achievement Test (CAT), and the Stanford Achievement Test (SAT), but with weak correlations. Saville-Troike (1984) linked oral verb use, the length of T-units, and grammatical accuracy to academic achievement, and she was surprised to discover that some of the students that had high grammar skills received the lower scores on the tests and some of the students who had lower grammar skills received higher scores on the tests.

Katz et al. (2004) cite a 1981 study by Tregar and Wong (1981) who used oral proficiency to examine the relationship between L1 reading comprehension and L2 reading comprehension. They used the Oral Dominance Test of the Boston Public Schools as a measure of oral English language proficiency, and they used cloze English reading tests of the same school district to measure English reading comprehension. In grades three through five they noted a higher correlation between L1 reading comprehension and L2 reading comprehension than between L2 oral ability and L2 reading comprehension, but for students in the middle grades (6, 7, and 8), there was a higher correlation between L2 oral ability and L2 reading comprehension than between L1 reading comprehension and L2 reading comprehension.

While it has been difficult for research to establish strong links between oral language ability and its influence on academic work, several researchers (Peregoy & Boyle, 2001; Roberts & Neal, 2004; August & Shanahan, 2006; Saunders & O'Brien,

2006) have posited such connections.

Cummins (1981, 2000a) has stated that oral language is part of the language continuum from basic interpersonal interactions to academic language, which is more complex in vocabulary and grammar. He has also argued (Cummins, 2000a) that the development of basic interpersonal communicative skills (BICS) and the acquisition of cognitive academic language proficiency (CALP) may be largely unrelated. That is, conversational English skills do not predict achievement in academic English. He maintained that since they follow separate developmental paths, children's conversational English skill do not play a role in their acquisition of academic English language skills. According to Cummins' theoretical framework, English immersion programs that stress mainly oral English ability, therefore, are not necessarily effective in the development of academic skills in English. August (2003), Durgunoglu & Oney (2001), Escamilla (2000), and Purcell-Gates (2001) lend support to Cummins' distinction between BICS and CALP by reporting that children's oral language skills do not predict future reading achievement. This would seem self-evident in that monolinguals start school with presumably equal L1 oral proficiency but do not perform equally on academic measures.

On the other hand, proponents of structured English immersion programs make no distinction between conversational English skills and academic English language skills (Baker, 1998; Gersten & Baker, 2000; Gersten & Woodward, 1995; and Rossel & Baker, 1996).

Crawford (1999) suggests that according to Rossell and colleagues, ELL students are simultaneously acquiring both *basic interpersonal communication skills*

(BICS) and *cognitive academic English proficiency* (CALP), and that the acquisition of conversational English skill may be taken as evidence for emergence of academic English skills. The implication is that the development of conversational English and acquisition of academic skills are one and the same. When ELL students attain fluent English oral ability, many proponents of Structured English immersion programs declare the students are ready for academic instruction in English (Krashen, 1996).

A few studies attempting to establish the relationship of oral language competency to program type, and from oral proficiency to academic proficiency have been cited. In their review of U.S. research during the last 20 years, Saunders & O'Brien (2006) point out only one study (Hakuta, Butler & Witt, 2000) that directly examined data on the rate of oral English language proficiency. Hakuta's study showed that it took up to six years or longer for students to achieve oral English fluency. Later, Lopez and Tashakkori (2006) compared ELLs placed in a two-way immersion program to ELLs placed in an early-exit transitional bilingual program and found students in the two-way program acquired oral English at a faster rate over a two and a half year span.

More recently Tong et al. (2008) have demonstrated in a 2-year study that oral language development can be improved regardless of program type (Transitional Bilingual vs. SEI) and that the rate of oral proficiency can be accelerated with intervention regardless of program type.

2.6 Using the CELDT as a Predictor of Academic Proficiency

Following their review of previous oral studies, Katz et al. (2004) introduced

their own project to try to link English proficiency to academic proficiency. To do so they used the state standardized California English Language Development Test (CELDT), which measures English oral and literacy ability, to predict performance on the Stanford-9 (SAT-9), a national standardized subject matter test. Their student population was from the San Francisco Unified School District, one-third of which were ELL students. Chinese- and Spanish-speaking students make up the majority of their ELL population, but there were also enough students of Korean-, Filipino-, and Japanese-speaking background to offer bilingual programs in those languages as well. This study spanned from the 2000-2001 to the 2002-2003 school year.

Rather than distinguish between bilingual and Structured English immersion programs, this study pooled all ELLs (including reclassified fluent English speakers, RFEPs) together and used their individual CELDT scores to try to establish correlations with the SAT-9. The overall score on the CELDT, a combined score for English listening, speaking, reading, and writing, was used to predict individual math and reading scores on the SAT-9. There was a low correlation between the CELDT and the math scores, but the CELDT and English reading scores on the SAT showed a stronger relation. See Table 2.1 and Table 2.2 below for examples.

Table 2.1 Correlation of Overall CELDT and Stanford 9 Math

Elementary			
Gr 2	Gr 3	Gr 4	Gr 5
.58	.53	.46	.39
Middle			
Gr 6	Gr 7	Gr 8	
.34	.23	.25	
High			
Gr 9	Gr 10	Gr 11	
.27	.16	.27	

Source: Katz et al., 2004 page 65

Table 2.2 Correlation of overall CELDT and Stanford 9 Reading

Elementary			
Gr 2	Gr 3	Gr 4	Gr 5
.58	.53	.46	.39
Middle			
Gr 6	Gr 7	Gr 8	
.34	.23	.25	
High			
Gr 9	Gr 10	Gr 11	
.27	.16	.27	

Source: Katz et al., 2004 page 65

For elementary students, there was a stronger correlation between both math and English reading than for secondary students. Then the authors examined the scores of students who had a CELDT overall score of "early advanced" with at least a reading and writing subscale score of "early advanced" and compared them to their SAT-9 English reading scores. The elementary students scored similar to the national norm group, so the CELDT was a predictor of success at that level. However, there was still a greater disconnect for middle and high school students. This proved true even when higher proficiency CELDT scores were compared to

SAT-9 scores.

Hakuta (2011) also used the CELDT to show the relationship between English proficiency and academic performance. His students were from Sanger Unified School District, a rural area just outside Fresno, California. This is a district that was low performing, but through interventions became a high performing school. The data Hakuta (2011) used was from 2003 to 2009. Like the previous study, Hakuta (2011) did not try to distinguish between program types, but matched the CELDT scores of all ELLs (and former ELLs) with their corresponding California Standards Test (CST) scores. The CST scores of native English speakers were also included for comparison. Once ELLs attained CELDT proficiency, they performed at the same level as native English speakers on the CST. This is similar to Katz et al.'s (2004) data for elementary students. However, for most students, this can take seven years.

Hakuta (2011) also found that students who scored between the CELDT overall "Beginning" and "Intermediate" levels in the 4th grade have an 81% chance of passing the high school exit exam. That chance drops to 48% if students are at that CELDT level in the 7th grade. Hakuta (2011) refers to this group as the "long-term English learners." In this district, upwards of 80% of all ELLs will score in the CELDT "Intermediate" range in two years, and almost 80% will score in the "Early Advanced" or "Advanced" range in seven years. Even more time will be required to reclassify as "Reclassified Fluent English Proficiency" (RFEP – meaning the students are no longer considered ELLs in need of special language services).

2.7 Studies Using the CELDT to Identify English Proficiency Across Program Types

Jepson (2009) used a non-academic instrument, the CELDT to compare bilingual with Structured English immersion programs. His database was the CELDT scores of most of the 500,000 English Leaners (ELLs) in California. He uses only Spanish-speaking students from grades 1-5. He excludes students from charter schools, nontraditional schools, special education schools (also individuals with special education services were not included), county offices of education, and students who repeated or skipped a year. Most of his student pool are in schools that have little language diversity and have a majority of ELLs. Students in any bilingual class were compared to students in any structured immersion class.

The results of Jepson's (2009) study were the same across each program type. Bilingual education has a large negative effect on English proficiency for students in grades 1 and 2, meaning (as measured by the CELDT) the students in bilingual programs did not learn English as rapidly as students who were in English-only programs. However, in grades 3-5 the negative effect of bilingual education is negligible or non-existent. For the CELDT listening and speaking sub-skills, bilingual education had a large negative effect in grades 1 and 2, but by grade 3 the effect was extremely small. In grade 3 bilingual education had a small positive effect on the CELDT reading sub-section, but in grades 4 and 5 there seems to be no association. For the writing CELDT sub-skill, bilingual education exhibited a slight negative association with English proficiency for each year. In other words, students in Structured English immersion programs scored higher overall in each CELDT category than students in bilingual programs.

2.8 Background of Program Types in District X Pre- and Post Proposition 227

By far, the largest recent change influencing ELL participation in program type in District X was the passing of Proposition 227 in 1998, which was adopted in a June election to be implemented fully in fall 1998. While the intent of the proposition was to eliminate bilingual education from California, it contained provisions for school waivers to continue instruction in a student's first, or native language (L1). About 70% of ELLs in District X received Structured English Immersion after Proposition 227, which contrasts with only 25% of students receiving Structured English Immersion before Proposition 227. In 1998, the bilingual program consisted of dual language (designed to support and sustain the L1 over time) and transitional types (designed to temporarily support content learning in L1 until the student was competent to function in English coursework). In 1998, structured English immersion existed in two different models, labeled "A" and "B". In both models the instruction was primarily in English. Model "A" allowed for some primary language support using a paraeducator, while Model "B" allowed for some primary language support from the teacher, who was fluent in the primary language. In 2003 those models were combined into one as there was found to be no effective difference in student outcomes between them (Parrish et al., 2006)).

About 8% of English learners in grades 5 and below in District X were enrolled each year in a bilingual instructional setting after the implementation of Proposition 227, and 141 elementary schools (out of some 600) maintained a bilingual program in grades 5 and below as of 2003 (Parrish et al., 2006).

When students enroll in District X they take the Home Language Survey. This

survey identifies the (1) student's first language, (2) the language the student uses most frequently in the home, (3) the language in which the parent communicates with the student most frequently, (4) the language the adults communicate in most frequently in the home, and (5) whether or not the student has received any formal English language instruction. If a parent indicates on the survey that a language other than English is used at home, or if the student has received English as a Second Language (ESL) training, the student will then be tested to determine their English ability level. Their English level will determine the appropriate program for the student.

The initial testing instrument is the CELDT (referred to as the Initial CELDT). Students who score at the "Advanced" level on the CELDT and show acceptable CST scores may be deemed "Initial Fluent English Proficient" (IFEP), and placed in "regular" (or, mainstream) English-only classes without special language instruction. Elementary students who score as "non-proficient" in English are then placed in a program according to their proficiency level. In secondary school, students who score non-proficient on the initial CELDT are also given a Diagnostic Proficiency Inventory (DPI – developed by the High Point publishers McGraw Hill) test to determine which English Language Development level (ELD), or English as a Second Language level (ESL) they should be placed in. For elementary (K-5) schools, the ELD levels are ELD 1 (lowest) – ELD 5 (most advanced). For secondary schools (grades 6 –12), lower level placement begins at a "Newcomer" level, which can last a year, then Beginning ESL 1A/1B (2nd year), Intermediate ESL 2A/2B (3rd year), and ESL 3/4 (4th year). Both elementary and secondary students are expected to

advance one English language level per year.

The goal of District X is for ELL students to be designated as Reclassified Fluent English Proficient (RFEP). This is expected to occur within five or six years from an initial evaluation of "non-proficient" regardless of grade level at the time of the initial assessment. English proficiency is defined by District X as students no longer being in need of special English language instructional services to be successful as they participate in English-only classrooms. This is established by multiple criteria, including acceptable scores on the CELDT and the Language Arts portion of the CST, and passing with a "C" or higher in ELD/ESL and English language arts classes.

The CELDT has listening (K-12), speaking (K-12), reading (grades 2-12) and writing (grades 2-12) components. Scoring levels are (a) Beginning, (b) Early Intermediate, (c) Intermediate, (d) Early Advanced, and (e) Advanced. Students are expected to advance one CELDT level per year. To score English Proficient on the CELDT, a student's overall CELDT score must be Early Advanced or higher, and each CELDT sub-skill must be Intermediate or higher.

The CST Language Arts and Math levels are (a) Far Below Basic, (b) Below Basic, (c) Basic, (d) Proficient, and (e) Advanced. A passing score on the CST is Basic or higher. Students are also expected to advance one CST level per year. Once students have passed their ELD/ESL courses, they must have a mainstream English language arts grade of "Proficient" or "Advanced" (for elementary), or a "C" or higher (for secondary). Once these criteria have been met, students can be re-classified (or re-designated) as Reclassified Fluent English Proficient (RFEP). It is at

this point that students are deemed fully proficient in English and no longer in need of special ELD services.

In District X students identified as non-English proficient are placed into one of four programs: (1) Structured English Immersion, (2) Basic Bilingual, (3) Dual-Language Bilingual, or (4) Mainstream English.

In the SEI program students receive all, or nearly all instruction in English. The emphasis is on the acquisition of academic English to enable students to be successful in English language arts and the other subjects across the curriculum. In District X the Structured English Immersion curriculum is a program put together by the publisher Hampton- Brown's *Into English* series. For secondary it is *High Point* (developed by publisher McGraw Hill). SEI students progress through ELD/ESL coursework and receive *sheltered instruction* (instructional language that is modified to facilitate comprehension) in their academic content classes. Once students in any program pass their ELD/ESL classes (ideally in 5 years) and begin taking regular, or mainstream, English language arts classes, they receive the label of Preparing to Reclassify as Fluent English Proficient (PRP) until they are reclassified as RFEP.

There are still two basic types of bilingual programs in District X: (a) the Transitional Bilingual program (TBE) and (b) the Dual Language Immersion (DLI) program. Students in the TBE receive academic content in their native language while they learn English in the same ELD/ESL program as the SEI students. Once students obtain proficiency in English, they transfer into the mainstream English program. The DLI program offers content and language instruction in both the

students' native language and English – the goal being to develop bilingual oral proficiency and literacy proficiency in both languages.

While the Mainstream Program is designed for monolingual English speakers, parents may request their ELL child to be placed in a mainstream classroom at any time. The term "Submersion" is sometimes used to characterize ELLs who are placed in regular, or mainstream classes, that are not specifically designed for language learner assistance. However, ideally, all District X teachers are required to have some training in sheltered English (SH) techniques (or, alternatively, "specially designed academic instruction in English" (SDAIE) methodology) to help ELLs who may be enrolled in their classes. In addition, the English proficiency progress of these ELL students continues to be monitored via the CELDT and the Language Arts sections of the CST. Thus, even ELL students in the Mainstream Program should receive sheltered some instruction as needed.

2.9 Program Results in District X

In 2000, the California Department of Education contracted with the American Institutes for Research and WestEd to conduct a five-year study (mandated by the California Legislature) to evaluate the effects of Proposition 227 on the education of ELLs in California. In their data collection, District X was chosen as a significant part of the study because of its huge student-level database (the DSS system discussed below). Part of their analyses examined student achievement across program type. Comparing the various ELL programs, Dual Language, Transitional Bilingual, Structured English Immersion, and Submersion (ELLs placed in regular, or

mainstream English classes) over this period from 1998 to 2003, they found no significant difference in overall student achievement on the California standardized SAT-9 test (Parrish et. al, 2006). While the achievement gap in standardized test scores persisted in District X between native English-speaking students and ELLs, ELLs continued (pre- and post Proposition 227) to show greater gains overall on standardized tests than native English-speaking students (Salazar, 2007).

Mixed results on the effectiveness of ELL program types have been found in District X's internal program monitoring data collection. District X's Master Plan Report (Salazar, 2007) gives a yearly summary of ELL student achievement across programs.

Regardless of their program, each ELL must take the CELDT annually. The District X goal is that students will advance one CELDT level annually, meaning that in 5 or 6 years the lowest level non-English proficient student should reclassify (no longer need to be labeled an ELL). The results, however, fall short of that goal. In the 2006/7 school year, only half of the ELL population in elementary schools met that objective (Salazar, 2007). In middle school 41.4% reached the target, and in high school, only 36.8% advanced one CELDT level per year. In the five-year span between 2002-03 and 2006-07, the percentage of students who reclassified increased at all schooling levels. The increase has been greatest for students in middle school, followed by students in elementary school.

In line with the studies cited above, there was some academic parity between programs. In the 2006/07 school year, students in the SEI programs narrowly outperformed their transitional bilingual program peers on the English Language

Arts sections of the CST, while students in the transitional bilingual program slightly outperformed their English immersion program counterparts on the Math section of the SAT-9 (Salazar, 2007).

Also in line with other studies (Hakuta, Butler, and Witt, 2000), performance in bilingual programs is not equal across ethnic groups and SES status. The Dual Language Program (DLP) CST results showed that the Korean/English DLP students in District X generally outperformed their SEI program peers from the same school in English Language Arts, and somewhat outperformed them in Math. However, all students participating in the Spanish/English DLP tended to score slightly lower in English Language Arts in the elementary grades, but higher in the middle school grades, than their structured English immersion/mainstream peers on math, except for Algebra 1 (Salazar, 2007).

There are a large percentage of students who have been enrolled in District X for six years or more that have not reclassified as English proficient. The 213,325 students who enrolled as English learners in any grade in District X six years prior or more, 41.5% were still classified as ELL (Salazar, 2007). For students who received their entire elementary (K-5) education in District X, nearly six- in-ten (58.4%) matriculated from elementary into middle school without reclassifying to fluent-English proficiency. These students spent their entire six-year elementary career as English learners. Four-in-ten students (39.5%) who received their entire K-8 instruction in a District X classroom matriculated from middle school into high school without being reclassified. More than one-third (35.2%) of the English learners who received eleven years of instruction in District X (K-10 cohort) had

still not reclassified. By the 11th grade, total student enrollment begins to significantly decline, making it difficult to interpret the results.

Two-thirds (66%) of the English learners who received their entire K-5 instruction in a District X school lacked sufficient skills to attain "Early Advanced" English proficiency on the CELDT. Previously, District X administrators believed that English learners did not reclassify due to a low report card grade-mark in their English or English language development class (the grade must be a "C" or higher), or for not scoring high enough on the language arts portion of the CST (a score of "Proficient", or higher). In actuality, the CELDT plays a bigger role in these students' non-reclassification. More specifically, most students (59.4%) who did not test English proficient "failed" either the English writing or reading component of the CELDT; that is, they did not attain a score of at least Intermediate in these sub-skills.

District X assumes that there were at least two contributing factors in low CELDT scores. One is the lack of SEI implementation in ELD/ESL classrooms. In the 2001/02 school year only 20% of students were receiving daily ELD instruction. As the District has sought to correct this, by Spring of 2007, 74% of English learner students were receiving daily ELD instruction (Salazar, 2007). Another contributing factor has been the presence of more than one or two ELD/ESL levels in the classroom. There is a strong correlation between the numbers of arbitrary levels in a classroom and the likelihood of high student achievement. While not mentioned in the Master Plan, other research has noted that weak program integrity is not limited to SEI. Researchers have pointed out that there is often more differentiation in L1 use between bilingual program types than between bilingual and SEI programs

(Rossell & Kuder, 2005).

2.10 What do we know and what needs to be done?

In Chapter 1 a range of concerns that contributed toward weak program evaluation comparison studies were listed. One was *inconsistently applied program labels.* For example, some studies (Rolstad et al., 2005a) claimed to compare bilingual to English-only programs, but upon closer scrutiny some of the English-only programs contained bilingual instruction components. A major concern was that *non-randomization* of students placed in programs wasn't controlled for. Several studies were criticized for *not controlling for pre-existing differences* (i.e., initial English proficiency levels). Various studies *compared inappropriate groups*—for example, comparing Asian populations to Hispanic populations, or middle class students to lower socioeconomic students. Some comparison studies compared groups that were exposed to *different school environments.* The *use of non-standardized tests* in comparing programs made it difficult to generalize some studies. In other instances *data were mixed from different studies,* or *studies were synthesized that were not similar* making it difficult to draw conclusions.

In several research studies listed above various bilingual and English-only programs were reviewed. Most of them compared bilingual and English-only programs to see which programs were more effective in producing academic English language acquisition. Some of these academic studies compared types of bilingual programs (dual language vs. transitional bilingual) to each other, and compared different English-only programs (ESL vs. structured English immersion

vs. submersion [mainstream]). These program comparisons produced mixed results.

Other programs were reviewed that sought to link oral English competency to academic performance on English tests. The instruments used to measure oral competency showed weak correspondences to academic performances. However, some studies were able to make predictions for the length of time it takes English language learners (ELLs) to develop academic oral competency and academic competency in general.

Finally, studies were reviewed that used the California English Language Development Test (CELDT) as an instrument to predict academic proficiency or predict literacy skills. One study Katz et al. (2004) indicated the CELDT had some predictive ability for reading success in the early grades. Another CELDT-based study, Jepsen (2009), sought to identify English proficiency divorced from academic growth.

There is a need for a study that controls for the weaknesses outlined above and in Chapter 1 when comparing bilingual programs to English-only programs. There is also a need for further studies that identify the development of English proficiency across program types independent of academic achievement. To date there has only been one study to do that (Jepsen, 2009), which has the limits of being a cross-sectional study. Finally, there is a need for a study that can link oral English proficiency with literacy skills. This study seeks to accomplish these objectives.

The current study seeks to add controls that other studies have not included.

One of the most important elements of the study is ensuring that each student is starting from the *same level of English proficiency*, which is absolutely no proficiency. Each student in this study scored at the lowest possible CELDT level with a cut score of 220. Therefore, each student begins with the same prior effects, not just somewhere in the continuum of the CELDT beginning level. No other study ensures this.

Another important set of factors for the study is *student identity*. All of the students in the bilingual program are Hispanic and are native Spanish-speakers. Nearly all of the students in the structured English immersion program are Hispanic and are native Spanish-speakers. Each student is part of District X. While admittedly District X is a large district, an analysis of the various sub-regions of District X indicate that the geographic location of the student does not present a significant advantage, or disadvantage for test scores. Also, the *type of school* the students in this study attend do not place the students at a significant advantage or disadvantage. The great majority of students are from a *low socioeconomic background*.

This study is a *longitudinal* six-year study that begins in the 2000-2001 school year and continues through the 2006-2007 school year and tracks the test scores of the *same students*. All students begin in Kindergarten, which controls for *age effects*.

Another unique feature of the current study is that it compares *limited transitional bilingual treatment to more long-term bilingual treatment*. The effect of being in a bilingual program for 1-3 years and 4-5 years before transferring to a structured English program is compared to only being in a structured English

immersion (SEI) program.

There are other important characteristics about this program comparison. All students in the study enrolled in District X *soon after the passage of Proposition 227* and therefore experienced the same collective upheaval that District X (and all California schools) passed through when undergoing a major policy shift. Each student in the study took the *same standardized English proficiency test* (the CELDT) every year at the same time of year. No data from other school districts or other programs were included in the study.

This study is also unique in that it tracks the relationship of student CELDT oral proficiency to CELDT reading and writing skills over time. By controlling for the elements listed above, this study is able to overcome many of the experimental features not present in previous program comparisons.

Chapter 3: Research Design and Methods

The following research design was developed to answer the two research questions (1) "Which language program type is more beneficial for second language acquisition: transitional bilingual education (TBE) or structured English immersion (SEI)?" and (2) "What is the link between oral proficiency and literacy skills?"

To accomplish these objectives it was decided to perform a secondary data analysis on a large urban public school district data set on consistent measures of English language development and subject matter achievement of students who were enrolled in various instructional programs to meet their language needs. The data was obtained from District X in California for the years 2002 to 2007. District X offered transitional bilingual programs (TBE) and structured English immersion (SEI) programs to English language learners (ELLs). District X in California has an extensive database that includes annual CELDT test scores for students from the CELDT's first implementation in 2001. Therefore a quantitative approach was chosen to compare programs. There were no known qualitative studies that paralleled this test score data for the time period of interest (2002-2007).

3.1 The School District

District X in the 2006-2007 school year enrolled over 700,000 students in grades K – 12 at over 900 schools and over 100 charter schools. It's area covers over 720 square miles and includes all or parts of 31 smaller municipalities plus several unincorporated sections of Southern California. (California Department of Education). Because of its size, the district was divided into 8 smaller administrative

units, or mini-districts. About 38% of the total student body were ELL students in 2005-2006. If the RFEP students (about 21%), and the IFEP students (about 10%) are included, about 70% of the entire student body comes from homes where a language other than English is spoken. The following Table shows the ELL student enrollment in elementary grades by the district's master plan program schedule from the 1998-1999 school year through the 2006-2007 school year.

Table 3.1 Program Placement of ELL Students in District X in K – 5

Elementary Grades

		Programs Taught Mainly in (SEI) or Totally in (Mainstream) English				Alternative Bilingual Programs			
	Total*	**SEI**	**%**	**Main**	**%**	**Bilingual**	**%**	**Dual Lang.**	**%**
1998-99	215,954	190,876	88.4	10,673	4.9	13,516	6.3	500	0.2
1999-00	221,053	186,221	84.2	11,053	5.0	22,028	10.0	494	0.2
2000-01	219,427	190,462	86.8	8,788	4.0	18,503	8.4	516	0.2
2001-02	210,478	188,275	89.5	4,051	1.9	15,243	7.2	569	0.3
2002-03	215,556	192,169	89.2	7,307	3.4	14,180	6.6	886	0.4
2003-04	218,374	188,983	86.5	12,987	5.9	13,328	6.1	1,284	0.6
2004-05	198,584	173,099	87.2	9,353	4.7	13,456	6.8	1,132	.08
2005-06	153,886	134,165	87.2	8,083	5.3	9,428	6.1	2,083	1.3

*Includes students with no Master Plan Program.
Scource: Salazar 2007

In Table 3.1 it can be seen that from the 2001-2002 school year to the 2006-2007 school year the percentage of students that were enrolled in structured English immersion programs and the percentage of students that were enrolled in the transitional bilingual education programs (Bilingual) was somewhat stable (with a steady decrease in the transitional bilingual programs). Accompanying the gradual decrease in transitional bilingual education enrollment was gradual but consistent enrollment of students in Dual Language programs.

Because of its size, District X is divided into smaller administrative units, called local districts, or mini-districts. The ELL population for this study is spread throughout the district. All school types that the district offers are included within in mini-district. The elementary school category options are: Elementary School (with no special programs for gifted or highly gifted students), Magnet Center on Regular Campus (which has some students enrolled in the Magnet Program), Magnet Self-Contained (which are full magnet schools), Span Magnet (span schools are schools that have a shorter or longer number of school years than the traditional elementary school six-year-term of grades K – 5), and Span (not Magnet). As with other students, ELL students are encouraged to apply to magnet programs.

3.2 Background of Program Types in District X Pre- and Post Proposition 227

By far, the largest recent change influencing ELL participation in program type in District X was the passing of Proposition 227 in 1998. While the intent of the proposition was to eliminate bilingual education from California, it contained provisions for school waivers to continue instruction in a student's first, or native

language (L1). About 70% of ELLs in District X received structured English immersion after Proposition 227, which contrasts with only 25% of students receiving English immersion before Proposition 227. Before 1998 there were two types of bilingual programs, one which included only native L1 speakers, and one which included a mix of L1 and native speakers of L2 (English). In 1998, the bilingual program consisted of dual language (designed to support and sustain the L1 over time) and transitional types (designed to temporarily support content learning in L1 until the student was competent to function in L2 coursework). In 1998, structured English immersion existed in two different models, labeled "A" and "B". In 2003 those models were combined into one as there was found to be no effective difference in student outcomes between them (Parrish et al., 2006)).

About 8% of English learners in grades 5 and below in District X were enrolled each year in a bilingual instructional setting after the implementation of Proposition 227, and 141 schools maintained a bilingual program in grades 5 and below as of 2003 (Parrish et al., 2006).

Currently when students enroll in District X they take the Home Language Survey. This survey identifies the (1) student's first language, (2) the language the student uses most frequently in the home, (3) the language in which the parent communicates with the student most frequently, (4) the language the adults communicate in most frequently in the home, and (5) whether or not the student has received any formal English language instruction. If a parent indicates on the survey that a language other than English is used at home, or if the student has received English as a Second Language (ESL) training, the student will then be

tested to determine their English proficiency level. Their proficiency level will determine the appropriate program for the student.

The initial testing instrument is the CELDT (referred to as the Initial CELDT). Students who score at the Advanced level on the CELDT and show acceptable CST scores may be deemed Initial Fluent English Proficient (IFEP), and placed in regular (mainstream) English classes without the need of special language instruction. Elementary students who score as "non-proficient" in English are then placed in a program according to their proficiency level. In secondary school, students who score non-proficient on the initial CELDT are also given a Diagnostic Proficiency Inventory (DPI – developed by the High Point publishers McGraw Hill) test to determine which English Development Level (ELD), or English as a Second Language level (ESL) they should be placed in. For elementary (K-5) schools, the ELD levels are ELD 1 (lowest) – ELD 5 (most advanced). For secondary schools (grades 6 –12), lower level placement begins at a newcomer level, which can last a year, then Beginning ESL 1A/1B (2st year), Intermediate ESL 2A/2B (3rd year), and ESL 3/4 (4rd year). Both elementary and secondary students are expected to advance one ELD level per year.

The goal of District X is for students to reclassify their status, or designation, from ELL to Reclassified Fluent English Proficient (RFEP). This is expected to occur within five or six years from an initial evaluation of "non-proficient" regardless of grade level at the time of the initial assessment. English proficiency is defined by District X as students no longer being in need of special ELD instructional services to be successful as they participate in mainstream academic content courses. This is

established by acceptable scores on the CELDT, the Language Arts portion of the CST, and passing grades in ELD/ESL and English language arts classes (with a "C" grade[or its equivalent], or higher).

The CELDT has listening (K-12), speaking (K-12), reading (grades 2-12) and writing (grades 2-12) components. Scoring levels are (a) Beginning, (b) Early Intermediate, (c) Intermediate, (d) Early Advanced, and (e) Advanced. Students are expected to advance one CELDT level per year. To score English Proficient on the CELDT, a student's overall CELDT score must be Early Advanced or higher, and each CELDT sub-skill must be Intermediate or higher. The CST Language Arts and Math levels are (a) Far Below Basic, (b) Below Basic, (c) Basic, (d) Proficient, and (e) Advanced. A passing score on the CST is Basic or higher. Students are also expected to advance one CST level per year. Once students have passed their ELD/ESL courses, they must have a mainstream English language arts grade of "Proficient" or "Advanced" (for elementary), or a "C" or higher (for secondary). Once these criteria have been met, students can be re-classified (or re-designated) as Reclassified Fluent English Proficient (RFEP). It is at this point that students are deemed fully proficient in English and no longer in need of special ELD services.

In District X students identified as non-English proficient are placed into one of four programs: (1) Structured English Immersion, (2) Basic Bilingual, (3) Dual-Language Bilingual, or (4) Mainstream English.

In the SEI program students receive all, or nearly all instruction in English. The emphasis is on the acquisition of academic English to enable students to be successful in English language arts and the other subjects across the curriculum. In

District X the Structured English Immersion curriculum is a program put together by the publisher Hampton- Brown's *Into English* series. For secondary it is *High Point* (developed by publishers McGraw Hill). SEI students progress through ELD/ESL coursework and receive *sheltered instruction* (instruction that is modified to facilitate comprehension) in their academic content classes. Once students in any program pass their ELD/ESL classes and begin taking regular, or mainstream, English language arts classes, they receive the title of Preparing to Reclassify as Fluent English Proficient (PRP) until they are reclassified as RFEP.

There are still two basic types of bilingual programs in District X: (a) the Transitional Bilingual program (TBE) and (b) the Dual Language Immersion (DLI) program. Students in the TBE receive academic content in their native language while they learn English in the same ELD/ESL program as the SEI students. Once students obtain proficiency in English, they transfer into the mainstream English program. The DLI program offers academic and language instruction in both the students' native language and English – the goal being to develop bilingual oral proficiency and literacy in both languages.

While the Mainstream Program is designed for monolingual English speakers, parents may request their ELL child to be placed in a mainstream classroom at any time. The term "Submersion" is sometimes used to characterize ELLs who are placed in regular, or mainstream classes, that are not specifically designed for language learner assistance. However, ideally, the English proficiency progress of these ELL students continues to be monitored via the CELDT and the Language Arts sections of the CST. In addition, all District X teachers are required to have some

training in sheltered English (SH) techniques (or, alternatively, "specially designed academic instruction in English" (SDAIE) methodology) to help ELLs who may be enrolled in their classes. Thus, even ELL students in the Mainstream Program should receive special ELD instruction as needed.

District X was chosen because of the large data sets they have for CELDT test scores from students in various ELL programs. In the post Proposition 227 environment when bilingual programs were being de-emphasized and phased out in California schools, District X still had a sizable population of students in transitional bilingual programs, making it a good candidate for program comparisons. Unfortunately, while District X still offered dual language bilingual programs, the numbers of students in those programs were not large enough to be included in the study. There were also too few ELL students who received no special English language learning support (by parent request) to include in this analysis.

All of the students who participated in this study enrolled in District X in Kindergarten in the 2001-02 school year. The students were divided into two main cohorts named: Analysis 1 and Analysis 2. A description of the cohort characteristics follows.

3.3 Program Results in District X

In 2000, the California Department of Education contracted with the American Institutes for Research and WestEd to conduct a five-year study (mandated by the California Legislature) to evaluate the effects of Proposition 227 on the education of ELLs in California. In their data collection, District X was chosen as a significant part

of the study because of its huge student-level database (the DSS system discussed below). Part of their analyses examined student achievement across program type. Comparing the various ELL programs, Dual Language, Transitional Bilingual, Structured English Immersion, and Submersion (ELLs placed in regular, or mainstream English classes) over this period from 1998 to 2003, they found no significant difference in overall student achievement on the California standardized tests (Parrish, 2006). While the achievement gap in standardized test scores persisted in District X between native English-speaking students and ELLs, ELLs continued (pre- and post Proposition 227) to show greater gains on standardized tests than native English-speaking students (Salazar, 2007).

Mixed results on the effectiveness of ELL program types have also been found in District X's internal program monitoring data collection. District X's Master Plan Report (Salazar, 2007) gives a yearly summary of ELL student achievement across programs.

Regardless of their program, each ELL must take the CELDT annually. The District X goal is that students will advance one CELDT level annually, meaning that in 5 or 6 years the lowest level non-English proficient student should reclassify (no longer need to be labeled an ELL). The results, however, fall short of that goal. In the 2006/7 school year, only half of the ELL population in elementary schools met that objective (Salazar, 2007). In middle school 41.4% reached the target, and in high school, only 36.8% advanced one CELDT level per year. In the five-year span between 2002-03 and 2006-07, the percentage of students who reclassified has increased at all schooling levels. The increase has been greatest for students in

middle school, followed by students in elementary school.

In line with the studies cited above, there was some academic parity between programs. In the 2006/07 school year, students in the SEI programs narrowly outperformed their bilingual program peers in ELA, while students in the bilingual program slightly outperformed their English immersion program counterparts in math (Salazar, 2007).

3.4 Cohort Size and Composition

3.4.1 Cohort 1: The Analysis 1 Group

This cohort was composed of students in four different language program configurations. One group is identified as *English-only* (**EO—**bold coded for ease in identification). This group was in a structured English immersion program for the entire six-year period of the study (2002 – 2007). Another group is identified as *bilingual short duration* (**BBS**). The students in this group were in a transitional bilingual program for one to three years and then transferred to a structured English immersion program for the remainder of the study. Another group is identified as *bilingual long duration* (**BBL**). These students were in a transitional bilingual program for four to five years. This study discovered through previous analysis that the students who were in bilingual programs for one to three years scored very closely to each other the CELDT, so their scores were combined. A similar analysis yielded the same results for the students who were in bilingual programs for four to five years, so their scores were also combined. By including students who received bilingual treatment for different periods of time prior to

switching to structured English immersion programs, it was possible to analyze the results of different participation periods of transitional bilingual education combined with structured English immersion. A final group in the Analysis 1 cohort is labeled *bilingual only group* (**BBO**). These students were in a transitional bilingual education group the entire six-year period of the study. Table 3.1 illustrates the number of students in each group in the Analysis 1 cohort. Each student in the groups progressed from one grade level to the next each year.

Table 3.2 Student Numerical Composition for the Analysis 1 Cohort

	2002 Grade K	2003 Grade 1	2004 Grade 2	2005 Grade 3	2006 Grade 4	2007 Grade 5
EO	875	875	875	875	875	875
BBS	86	86	86	86	86	86
BBL	155	155	155	155	155	155
BBO	80	80	80	80	80	80
Total Students	1,196	1,196	1,196	1,196	1,196	1,196

This chart shows the yearly student population for each sub-cohort in the Analysis 1 group. Each student in these groups started and remained in their respective cohorts the entire 6-year period of the study.

3.4.2 Cohort 2: The Analysis 2 Group

Unlike the Analysis 1 cohort, this cohort consisted of only two groups. One group was labeled *English Immersion* (**EI**). Like the **EO** group in the Analysis 1 cohort, all of these students were in a structured English immersion program. However, to distinguish them from the **EO** group in Analysis 1, they are labeled **EI**. Also like the Analysis 1 group, the other group in the Analysis 2 cohort is identified as *transitional bilingual education only group* (**BB**). They are labeled **BB** to distinguish them from the **BBO** group in Analysis 1. As Table 3.2 indicates, some

student attrition occurred over time in the Analysis 2 group. This is because to make the sample size larger, this cohort included students who weren't in the study for the entire six-year span, which would make it possible to see if the sample size unduly influenced student scores. The number of students in 2007 in Table 3.2 represent the number of students who were in the program for the entire length of the study. It should be noted that the **EO** students from the Analysis 1 cohort were included in the **EI** population of the Analysis 2 cohort, and the **BBO** students in the Analysis 1 cohort were included in the Analysis 2 cohort. Table 3.2 illustrates the number of students by year for the **EI** and **BB** students in the Analysis 2 cohort.

Table 3.3 Student Numerical Composition for the Analysis 2 Cohort

	2002 Grade K	2003 Grade 1	2004 Grade 2	2005 Grade 3	2006 Grade 4	2007 Grade 5
EI	2,094	2,094	1,714	1,495	1,271	973
BB	637	637	462	334	184	82
Total Students	2,731	2,731	2,176	1,829	1,455	1,055

This table identifies the number of students per year for each group in the Analysis 2 cohort.

3.4.3 Cohorts 1 and 2: Student Characteristics

Tables 3.4 and 3.5 provide information regarding socioeconomic status, school transiency, cultural background, and geographic location of the students in the data sample.

Table 3.4 Analysis 1 Cohort: Student Characteristics by Group Type

	EO	**BBS**	**BBL**	**BBO**
Free Lunch	88.2%	93.7%	83.1%*	93.1%
Changed Schools	39.5%	23.2%	38.4%***	6.3%***
Ethnicity				
Hispanic	97.7%	99.7%	99.6%	99.8%
Asian	1.0%	0.0%	0.0%	0.0%
White	1.0%	0.0%	0.0%	0.0%
Other	0.2%	0.3%	0.4%	0.2%
Home Language				
Spanish	97.9%	100.0%	100.0%*	100.0%
Mini-districts that comprise different geographical zones in District X				
1	12.9%	2.6%	8.9%***	0.0%***
2	11.0%	6.0%	6.2%*	11.3%
3	6.3%	0.0%**	9.1%***	0.0%**
4	15.9%	16.3%**	9.1%	28.3%**
5	18.8%	51.7%	20.2%***	50.2%***
6	8.8%	11.3%	6.2%	10.0%
7	18.6%	7.8%**	13.6%***	0.0%***
8	5.8%	0.6%***	13.0%***	0.0%**
R	0.6%	1.3%***	11.6%	0.2%
T	1.3%	2.3%*	2.1%	0.0%
	100%	100%	100%	100%

**p<.05, ***p<.01

Statistical significance is assessed using t-tests comparing each respective group to **EI**

Table 3.5 Analysis 2 Cohort: Student Characteristics by Group Type

	EI	**BB**
Free Lunch	90.1%	94.7%***
Changed Schools	34.6%	24.6%***
Ethnicity		
Hispanic	96.4%	98.7%***
Asian	1.9%	0.5%**
White	1.1%	0.0%***
Other	0.6%	0.8%
Home Language		
Spanish	96.8%	99.5%***
Mini-districts that comprise different geographical zones in District X		
1	12.4%	4.6%***
2	11.0%	8.2%**
3	6.1%	2.8%***
4	19.7%	15.5%**
5	17.1%	43.0%***
6	8.2%	10.0%
7	16.0%	5.3%***
8	7.9%	6.0%*
R	0.4%	3.0%***
T	1.1%	1.6%
	100%	100%

**p<.05, ***p<.01
Statistical significance is assessed using t-tests comparing **BB** to **EI**

Table 3.6 shows information about CELDT scores from the various mini-districts in District X. The first column shows the average CELDT score for all years combined for each mini-district. The second column shows the average CELDT score for each mini-district for the last year of the study.

Table 3.6 Mean CELDT Overall Score by District X Mini Districts

	Mean Overall Score	
District	**All Years**	**2007**
1	459	529
2	464	534
3	461	529
4	444	528
5	442	516
6	459	528
7	452	519
8	459	528
R	459	545
T	479	507

Note: For the mini-districts 1-8 (which comprise 98.5% of the EO/EI students and 98.4 of the BBO/BB students) the Overall mean CELDT scores range from 442 to 459 for the "All Years) category – a difference of 17 points. The Overall mean CELDT scores for 2007 range from 516 to 534 – a difference of 18 points.

Thus, Tables 3.5 – 3.6 provide details regarding the homogeneity of the cohorts. The Analysis 1 and Analysis 2 groups are very similar in composition. From 83% to 95% of the students qualify for the district's Free Lunch program. Twenty-three to forty percent of students in all groups have changed schools at least once (the exception being **BBO**, which was more stable with only 6.3% of the students changing schools). Nearly all students were identified as Hispanic (96% - 100%); and accordingly, the native language of most students (97% - 100%) was Spanish. The **EO/EI** groups were more evenly distributed across District X's eight mini-districts. **BBO** was the least spread out, only showing up in mini-districts 2, 4, 5, and 6. The highest numbers of ELL students across the board was (in order from highest to lowest) in mini-districts 5, 4, 2, and 6 (the same mini-districts where **BBO** students proportionately were found).

Table 3.6 shows the relative CELDT Overall means for "All Years" combined and for 2007. Thirty-five percent of **EO/EI** cluster in the lowest scoring mini-districts (5, 7) and 24% cluster in the highest scoring mini-districts (1, 2). Fifty percent of **BBO** and forty-three percent of **BB** are in the lowest scoring mini-district (5), while 11% (**BBO**) and 8% (**BB**) are in the highest scoring mini-district (2). If schooling in a given mini-district provides an advantage, it is negligible as there is only a 17-point difference between the lowest and highest scores for the "All Years" means, and only an 18-point difference between the lowest and highest scores for the "2007" means (neither enough to move from one CELDT level to another in their respective ranges).

3.5 Instruments and Measures

The measures that were available in the student data set from District X that could help answer the two research questions: (1) "Which language program type is more beneficial for second language acquisition: transitional bilingual education (TBE) or structured English immersion (SEI)?" and (2) "What is the link between oral proficiency and literacy skills?", were the California English Language Development Test (CELDT) scores.

3.5.1 Measure of English Language Proficiency

The CELDT is an assessment based upon the K-12 English Language Development (ELD) Standards adopted by the California State Board of Education in 1999. These California ELD standards were created by a committee of educators

who were asked to use both the national Teachers of English to Speakers of Other Languages (TESOL) standards as a base and link the California ELD standards to the already created California English Language Arts standards (Kuhlman and Nadeau, 1999). While the CELDT is standards based, it is designed to measure English language proficiency, not academic proficiency. Hakuta, Butler and Witt (2000) identified the English Language Development (ELD) Standards and the assessments based on these standards as being a helpful future point of reference for further "policy-relevant" study.

Everyone in this study started in Kindergarten at the same language proficiency level based on a test score from the CELDT, which all students whose first language is not English are required to take upon enrollment in District X. This entry-level evaluation is called the "Initial CELDT". The Initial CELDT score for each student selected for the study was "220" (the lowest possible score), a scale score that indicates *no* proficiency in English. Therefore, every student in the study begins at the same grade level (which implies roughly the same age) and same CELDT level of English proficiency. The chart below shows the respective CELDT scores and corresponding ELD levels based on those scores for 2002 for Kindergarten, the year this study begins.

Table 3.7 ELD Level Placement from Initial CELDT Scores in 2001/02

Kindergarten	ELD LEVEL Based on CELDT Scores				
	Beginning	Early Intermediate	Intermediate	Early Advanced	Advanced
Listening & Speaking CELDT SCORES	220 - 409	410 - 457	458 - 505	506 - 553	554 +

Source: California Department of Education

As Table 3.8 shows, the CELDT scale scores are different for different grade levels.

Table 3.8 CELDT Scale Cut-off Scores to Determine ELD Level

GRADE	ELD LEVEL Based on CELDT Scores Early Intermediate	Intermediate	Early Advanced	Advanced
	Listening/Speaking			
Kindergarten	410	458	506	554
Grade 1	424	471	517	564
Grade 2	454	495	536	577
Grade 3 - 5	438	482	526	569
Grade 6 - 8	438	482	526	569
Grade 9 - 12	438	482	526	569
	Reading			
Grade 2	438	475	511	548
Grade 3 - 5	466	499	533	566
Grade 6 - 8	466	499	533	566
Grade 9 - 12	466	499	533	566
	Writing			
Grade 2	424	469	514	559
Grade 3 - 5	445	488	530	573
Grade 6 - 8	445	488	530	573
Grade 9 - 12	445	488	530	573
	Overall			
Kindergarten	410	458	506	554
Grade 1	424	471	517	564
Grade 2	443	483	524	565
Grade 3 - 5	447	488	529	569
Grade 6 - 8	447	488	529	569
Grade 9 - 12	447	488	529	569

Source: California Department of Education

3.5.2 Test Administration

Every year thereafter the students are required to take the "Annual CELDT" (which is the same as the "Initial CELDT") to monitor progress in English in two domains: oral skills and literacy skills. All students take the Annual CELDT early in the Fall. For this study, from 2002 to 2007, the CELDT oral component is labeled

"Listening/Speaking." As the test domain name indicates, it is a composite aural/oral score. Beginning in the 2007 school-end year, the CELDT records Listening and Speaking as two separate scores. Since students do not begin taking CELDT literacy tests (for Reading and Writing) until second grade, for this study students began being evaluated on their literacy proficiencies in 2004. From these oral and literacy scores, a composite "Overall" score was developed on the CELDT. For the first two years (2002-2003) then, the Overall scores and Listening/Speaking scores are the same as there are no evaluations for reading and writing until grade two. As mentioned, the Reading and Writing evaluations began in 2004 and from then on the composite Overall score reflects progress in all four language domains combined.

Data from the CELDT was used to evaluate the relative effectiveness of English-only SEI programs and transitional bilingual education (TBE, from here on just BB) exit programs. District X provided both individual CELDT scale scores and CELDT proficiency levels. The 5-level sequence of CELDT level proficiency is: "Beginning", "Early Intermediate", "Intermediate", "Early Advanced", and "Advanced". Students can attain scores at each of these proficiency levels for each oral and literacy segment of the CELDT. Thus, student growth in each skill level could be monitored. These data also made it possible to link oral proficiency to literacy proficiency. Both *scale scores* and *proficiency levels* were used to evaluate program effectiveness.

Students in California (see Chapter 2, p. 38) are expected to advance one CELDT level per year, i.e., from Early Intermediate in grade one to Intermediate in

grade two. The goal is for ELL students to learn enough English to be reclassified from ELL to "Reclassified Fluent English Speaker" (RFEP) over four years. To accomplish this, students must score Early Advanced on the Overall portion of the CELDT while simultaneously scoring at least Intermediate on each CELDT domain: Listening/Speaking, Reading, and Writing. In addition, students must meet two other requirements. First, they must receive a classroom grade of "C" or better. Second, they must receive a score on the CST of "Proficient" or higher. In the event that a student attains only two out of these three requirements, he or she is classified "Preparing to Reclassify" PRP, and may then enter mainstream classes. This information is also used to evaluate program effectiveness in this study.

3.6 Research Question 1: Data Analysis

All data analyses for the Analysis 1 and Analysis 2 cohorts were run using Stata 12 software.

3.6.1 Data Analysis for Cohort 1 Means

The goal of these analyses for the Analysis 1 cohort was to assess the difference in improvement from the baseline years (2002 for Overall and Listening/Speaking, and 2004 for Reading and Writing) between the three BB groups — (1) **BBS** (2), **BBL**, and (3) **BBO** — and the **EO** group. To do this, regressions were run that interact each year dummy variables with each group, omitting the year 2002 and the **EO** group to be used as references.

Five types of regressions were run on the means of the Analysis 1 group. The first regression was an Ordinary Least Squares (OLS) with no controls. This is a

simple model and tracks the raw figures exactly since numbers are unadjusted.

Next, an OLS model with controls was run. The controls included: (1) free lunch (i.e., students who qualified for free lunch at school, a proxy SES marker), (2) student home language, (3) student ethnicity, (4) the local (or mini) district within District X that the student was in (a proxy for "best practices"), and (5) the type of school the student attended (i.e., magnet, elementary, span magnet, etc., a proxy for best practices).

Then an individual fixed effects regression was run on the means. Individual fixed effects models control for any time invariant heterogeneity in individual student characteristics. For example, some students may be more "ambitious" or have a more "natural ability" to score high on a test that could bias results. These types of omitted variables could bias OLS coefficients. This model controls for any such heterogeneity mitigating any bias from time invariant heterogeneity.

Then the means were subjected to a mixed effects regression with no controls, which allows for a random intercept for individuals and random slopes for each group.

This was followed by a mixed effects regression with controls.

3.6.2 Data Analysis for Cohort 2 Means

The goal of the means analysis for the second cohort was to measure differences between **EI** and **BB** in average annual score improvement. This differs from the first analysis in that it looks at all years combined and takes the average annual improvement instead of looking at differences for each year individually.

Groups for these analyses were selected by including all students that were in **BB** or **EI** for at least 2002 and 2003 to begin with. It is necessary to have at least two years of data in order to measure annual changes. Students drop out of the sample if they switch to a different program classification or leave the district. For example, if a student is in **EI** from 2002 to 2003 and then switches to **BB** in 2004, they will be dropped in 2004 and for all remaining years. Also, only students that represent continuous enrollment starting in 2002 are included. This means that if a student is in **BB** from 2002-2004, but in **EI** from 2005-2007, only the 2002-2004 years will be included. The years after switching are not included since there may be a complimentary of effect of starting in one group and switching to another (e.g. students that start in **BB** and then switch to **EI** may perform better than students that are in **EI** for all years).

All of the statistical models that were used in the analyses for the Analysis 1 Group are used to measure differences in rate of acquisition. However, in contrast to the earlier analyses the interaction term that identifies the **BB**-**EI** difference in rate of acquisition considers years as one continuous variable as opposed to binary variables for each year. The resulting coefficients on these variables (Group X Year) thus represent the difference in average annual improvement between **EI** and **BB**. Dummy variables for each year are also included in the model, which controls for time trends. For example, there were generally large improvements in earlier years and the **BB** population is much bigger in earlier years. If time trends are not controlled for, the model would erroneously estimate **BB** as much better than **EI**. However, this is only because more students are in **BB** in years where there are

larger improvements. All other variables are treated the same as the analyses for the Analysis 1 Group.

The Analysis 2 analysis adds value over the analysis for the Analysis 1 Group in that it does not require students to be continuously enrolled, which (1) increases sample size and (2) includes students who may have switched out of **BB** to a different program in later years.

3.6.3 Research Question 1: Comparison of CELDT Means

Once the means from the Analysis 1 and Analysis 2 cohorts are analyzed, the results are placed in tables and examined for significance. After the significance of the CELDT mean results are established, graphs and tables of raw CELDT means are generated to compare the relative progress of English acquisition for the groups in the two cohorts. These comparisons were for the annual English skills acquired as recorded on the CELDT Overall, Listening/Speaking, Reading, and Writing from 2002 – 2007.

3.7 Research Question 1: CELDT Proficiency Levels

When students receive a raw CELDT score, the score automatically places them into one of the five levels: Beginning, Early Intermediate, Intermediate, Early Advanced, or Advanced. It is possible for a student to attain different level placements for each of the English skills tested. For example, on the same CELDT exam, a student could attain a level of Early Advanced in Reading, but only a level of Intermediate in Writing, etc. Looking at overall CELDT levels the cohort groups attained from year to year provides another way to evaluate language program

effectiveness. Tables were therefore generated to illustrate yearly CELDT proficiency levels for each category for the Analysis 1 and Analysis 2 cohorts.

3.8 Research Question 1: Scores of Highest Achievers

In addition to observing averaged group levels, it is also useful to compare the highest CELDT levels for each year between groups, thereby indicating which program type facilitates faster growth toward reclassification, a major goal for District X. This is done by creating tables that show the annual number of students in a cohort group that attain a CELDT Overall level of at least Early Advanced; that attain an Overall level of Early Advanced in addition to at least Intermediate on the other English skills; and that in addition to the previous CELDT level attainments, attained the academic scores to qualify for a re-classification as PRP or RFEP.

3.9 Research Question 2: CELDT Oral Proficiency Related to CELDT Literacy Proficiency

Finally, tables are developed to answer the second research question, "What is the relationship between English oral ability and literacy ability?" These tables list the annual CELDT oral levels for each cohort group and simultaneously list the cohort groups' reading and writing levels, making it possible to relate the three skills over time.

This analysis design makes it possible to answer the two main research questions of this study. The means analyses provide the confidence that the CELDT data are mostly driven by the language programs that the students are placed in. The resultant CELDT levels that the students are placed in based on those means

can be used to monitor rates of English acquisition between language programs and to establish relationships between oral and literacy skills.

Chapter 4: Results

The results of the analyses of the data collection makes it possible to answer the two main research questions of this study: (1) "Which language program type is more beneficial for English language acquisition: transitional bilingual education (TBE) or structured English immersion (SEI)?" and (2) "What is the relationship of oral proficiency to literacy skills?" As outlined in Chapter 3, the answer to Research Question 1 involved 4 steps: (1) the collection of raw CELDT scores to establish annual cohort group means; (2) the analysis of the means for significance; (3) the plotting of the means on graphs and tables and their analysis; and (4) the comparison of annual cohort group proficiency levels. To answer Research Question 2, CELDT proficiency levels were used to link oral skills to literacy skills. This chapter discusses in detail the results in that order to answer the two research questions posed for this study.

4.1 Research Question 1: Analysis of the CELDT Means

In order to have confidence that the CELDT means were influenced primarily by program type and not other factors several analyses were conducted for the two cohorts: Analysis 1 and Analysis 2. This section will review the results of those analyses.

From District X data, raw means from the CELDT scores were developed for each cohort group from 2002-2007 (for Overall), from 2002-2006 (for Listening/Speaking), in 2007 (for Listening and Speaking as separate scores), and

from 2004-2007 (for Reading and Writing). Following are analyses of the means for the Analysis 1 and the Analysis 2 cohorts.

Review of Coding Abbreviations of Cohort Groups (from Chapter 3 – 3.3 Cohort Size and Composition):

Analysis 1 Cohort

EO = Students who began in an English Immersion program and may later have "mainstreamed" to English-only instruction.

BBS = Students who were in the bilingual program for 1, 2, or 3 years, then transferred to the SEI program and may later have "mainstreamed".

BBL = Students who were in the bilingual program for 4 or 5 years, then transferred to the SEI program and may later have "mainstreamed".

BBO = Students who were in the bilingual program for the entire period of 6 years.

Analysis 2 Cohort

EI = Students in the Analysis 2 cohort who remained in an English Immersion program the entire study period of 6 years.

BB = Students in the Analysis 2 cohort who were in the bilingual program for the entire study period of 6 years.

4.1.1 The Statistical Models for the Analysis 1 Group

Five statistical analysis models were run on the Analysis 1 means in this order: (1) an ordinary least squares model with no controls, (2) an ordinary least squares model with controls, (3) an individual fixed effects model, (4) a mixed effects model with no controls, and (5) a fixed effects model with controls. Four statistical tables were generated for the CELDT Overall (Table 4.1), Listening/Speaking (Table 4.2), Reading (Table 4.3), and Writing (Table 4.4) means showing the results of the analyses. The controls included (1) free lunch (i.e.,

students who qualified for free lunch at school, a proxy SES marker), (2) student home language, (3) student ethnicity, (4) the local (or mini) district within District X that the student attended school in (a proxy for "best practices"), and (5) the type of school the student attended (i.e., magnet, elementary, span magnet, etc., a proxy for best practices). Tables 4.1 – 4.4 below show the regression effects on the means only. The extended tables showing free lunch, ethnicity, mini-school district, and school type are shown in Appendices 1-4 for the Analysis 1 cohort and Appendices 5-8 for the Analysis 2 cohort .

The goal of these analyses was to assess the difference in improvement from the baseline year (2002 for Overall and Listening/Speaking, and 2004 for Reading and Writing) between the three BB groups — (1) **BBS** (2), **BBL**, and (3) **BBO** — and the **EO** group. To do this, regressions were run that interact each year dummy variables with each group, omitting the year 2002 and the **EO** group to be used as references. The resulting coefficient on each year by group interaction term is the difference in score improvement between 2002 (2004 for reading and writing) and each respective year relative to the same improvement for the **EO** group. A positive coefficient indicates that the respective group improved by more than the **EO** group.

The results for each table are represented in columns showing the coefficients for each analysis. Column 1 presents these results using an ordinary least squares (OLS) regression model without any controls with the exception of year dummy variables. This is a simple model and tracks the "Raw Figures" exactly since numbers are unadjusted. Column 2 presents results using the same OLS model but includes controls. We can see that controls do not change results much in any of

the models. Column 3 presents results using an individual fixed effects regression model. Individual fixed effects models control for any time invariant heterogeneity in individual characteristics. For example, some students may be more "ambitious" or have a more "natural ability" to score high on a test that could bias results. These types of omitted variables could bias OLS coefficients. This model controls for any such heterogeneity mitigating any bias from time invariant heterogeneity. Poverty level, school district, and school type are included as controls in this model as well since these may vary over time. Column 4 presents the results from a mixed effects model (allowing for fixed effects and random effects). This first mixed model includes no fixed effect controls (similar to column 1); however, it allows for a random intercept for individuals and random slopes for each group. Column 5 presents the results for the mixed models with controls.

The individual fixed effects model seems to be the most efficient. However, estimates hardly change across models, indicating that for these analyses, the models do not make much difference. For Table 4.1 (CELDT Overall scores) all of the results are highly significant, indicating that the language program the students were enrolled in made a significant contribution to their CELDT scores. Table 4.2 (CELDT Listening/Speaking scores) shows similar results. All of the coefficients are also highly significant for each program in each year and results vary little between regression analyses.

Table 4.1 Regressions on CELDT Overall Means for Analysis 1 Cohort Differences in Improvement in Overall Scores from 2002 Relative **EO**

VARIABLES	OLS (no controls) 1	OLS (controls) 2	FE (Fixed Effects) 3	Mixed Effects (no controls) 4	Mixed Effects (controls) 5
BBO					
All Years in BB					
Difference 2003 (Grade1)	-48.30***	-48.07***	-47.62***	-48.04***	-47.92***
	(8.105)	(8.137)	(8.116)	(9.391)	(9.413)
Difference 2004 (Grade2)	-40.17***	-40.46***	-40.01***	-40.22***	-40.69***
	(3.921)	(3.989)	(3.944)	(4.391)	(4.488)
Difference 2005 (Grade3)	-29.04***	-29.72***	-28.59***	-28.85***	-29.59***
	(3.754)	(3.783)	(3.754)	(5.686)	(5.639)
Difference 2006 (Grade4)	-23.03***	-24.05***	-22.83***	-22.84***	-23.99***
	(3.918)	(3.899)	(3.879)	(6.279)	(6.231)
Difference 2007 (Grade5)	-17.36***	-18.81***	-17.21***	-17.21**	-18.83***
	(4.763)	(4.849)	(4.782)	(6.847)	(7.058)
BBL (Grades 1-5)					
4 or 5 Years in BB					
Difference 2003 (Grade1)	-53.71***	-53.51***	-52.98***	-53.50***	-53.42***
	(6.179)	(6.195)	(6.180)	(6.976)	(6.954)
Difference 2004 (Grade2)	-35.02***	-34.83***	-34.37***	-35.17***	-35.13***
	(2.761)	(2.761)	(2.758)	(5.524)	(5.511)
Difference 2005 (Grade3)	-23.88***	-23.87***	-22.86***	-22.98***	-23.03***
	(3.124)	(3.137)	(3.125)	(4.772)	(4.765)
Difference 2006 (Grade4)	-19.73***	-20.08***	-18.92***	-18.82***	-19.27***
	(2.782)	(2.814)	(2.800)	(4.232)	(4.285)
Difference 2007 (Grade5)	-15.55***	-15.73***	-14.64***	-14.90***	-15.18***
	(3.121)	(3.148)	(3.133)	(4.547)	(4.580)
BBS (Grades 1-5)					
1,2,or 3 years in BB					
Difference 2003 (Grade 1)	-36.61***	-36.39***	-36.44***	-36.20***	-36.03***
	(7.062)	(7.061)	(7.048)	(12.67)	(12.70)
Difference 2004 (Grade 2)	-20.56***	-20.23***	-20.33***	-20.34***	-20.02***
	(3.838)	(3.862)	(3.799)	(6.691)	(6.703)
Difference 2005 (Grade 3)	-17.46***	-17.44***	-17.26***	-17.39***	-17.19***
	(3.630)	(3.683)	(3.628)	(4.181)	(4.154)
Difference 2006 (Grade 4)	-6.635**	-6.643**	-6.474**	-6.332	-6.204
	(3.238)	(3.304)	(3.251)	(4.245)	(4.209)
Difference 2007 (Grade 5)	-0.0568	0.385	0.257	-0.0223	0.422
	(4.431)	(4.527)	(4.429)	(4.962)	(4.949)
	(1.42e-07)	(14.11)	(6.589)	(0.679)	(15.31)
Observations	7,176	7,176	7,176	7,176	7,176
R-squared	0.895	0.897	0.938		
Number of groups				321	321
Number of id			1,196		

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

The figures in columns 1 – 5 (from left to right) represent CELDT scale score values (in terms of yearly improvement) compared to the **EO** group from the base year 2002, when they enrolled in Kindergarten. For example, in the first row of **BBO**, the figure "-48.30" in Column 1 (OLS (no controls)) means that the **BBO** group improved 48.30 scale points less than the **EO** group improved in the Overall domain from 2002 to 2003. The triple asterisks that follow the figure indicate that the figure is statistically significant.

The increase in scale score of a BB group from year to year as compared to the **EO** group's improvement does not indicate that a BB score increase is lower than, equal to, or higher than an **EO** score. If a BB group started at a low level on a given year, it is possible that the BB group's score *increased more* that another BB group or the **EO** group, but still did not get a higher score.

This table shows that all three categories of BB improved by significantly less from their 2002 score than the EI group in nearly all years. Performance was particularly low in the first couple of years, where the BBO group improved by 48.30 and 40.17 points less than the EI group by 2003, 2004, respectively (all strongly significant). In the later years, performance in the BB group relative to the EI group was not quite as low, with the BBO group only improving by 17 points less than the EI group by 2007. Less time in the BB group seems to have been better for performance. Students with only 2 or 3 years (BBS) in the BB group performed better than students with 4 to 5 years in BB or all 6 in years in the BB group. Results were consistent across all models providing confidence in the estimates.

For Table 4.1 (CELDT Overall scores) most of the results are highly significant, indicating that the language program the students were enrolled in made a significant contribution to their CELDT scores.

Table 4.2 Regressions on CELDT Listening/Speaking Means for Analysis 1 Cohort Differences in Improvement in Oral Scores from 2002 Relative **EO**

VARIABLES	OLS (no controls) 1	OLS (controls) 2	FE (Fixed Effects) 3	Mixed Effects (no controls) 4	Mixed Effects (controls) 5
BBO					
All Years in BB					
Difference 2003 (Grade 1)	-48.30***	-47.87***	-47.67***	-48.13***	-47.85***
	(8.105)	(8.137)	(8.117)	(9.393)	(9.449)
Difference 2004 (Grade 2)	-41.37***	-41.69***	-41.33***	-41.64***	-42.11***
	(5.451)	(5.514)	(5.473)	(5.509)	(5.660)
Difference 2005 (Grade 3)	-28.30***	-28.83***	-27.80***	-28.26***	-28.89***
	(4.487)	(4.486)	(4.476)	(6.824)	(6.717)
Difference 2006 (Grade 4)	-24.13***	-25.42***	-24.08***	-24.27***	-25.65***
	(4.831)	(4.836)	(4.811)	(9.079)	(9.198)
BBL					
4 or 5 Years in BB					
Difference 2003 (Grade 1)	-53.71***	-53.30***	-53.02***	-53.62***	-53.36***
	(6.179)	(6.193)	(6.180)	(6.967)	(6.937)
Difference 2004 (Grade 2)	-39.10***	-38.79***	-38.52***	-39.65***	-39.44***
	(3.632)	(3.630)	(3.629)	(6.704)	(6.658)
Difference 2005 (Grade 3)	-23.08***	-22.81***	-21.97***	-22.56***	-22.39***
	(3.748)	(3.779)	(3.767)	(4.934)	(4.915)
Difference 2006 (Grade 4)	-21.84***	-22.26***	-20.98***	-21.65***	-22.11***
	(3.728)	(3.778)	(3.752)	(4.888)	(4.969)
BBS					
1,2,or 3 years in BB					
Difference 2003 (Grade 1)	-36.61***	-36.42***	-36.23***	-36.38***	-36.23***
	(7.062)	(7.055)	(7.070)	(12.67)	(12.71)
Difference 2004 (Grade 2)	-22.80***	-22.41***	-22.19***	-22.88***	-22.43***
	(4.494)	(4.537)	(4.503)	(6.698)	(6.685)
Difference 2005 (Grade 3)	-20.60***	-20.84***	-20.28***	-20.90***	-20.84***
	(4.086)	(4.115)	(4.113)	(4.170)	(4.099)
Difference 2006 (Grade 4)	-8.864**	-9.136**	-8.397**	-8.931*	-8.931*
	(4.077)	(4.104)	(4.106)	(4.993)	(4.852)
Observations	5,980	5,980	5,980	5,980	5,980
R-squared	0.885	0.887	0.928		
Number of groups				301	301
Number of id			1,196		

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Results for the oral score are very similar to the overall score, with poorer relative performance in the earlier years in the BB groups and better performance with less time in the BB groups. Results are strongly significant.

Table 4.3 Regressions on CELDT Reading Means for Analysis 1 Cohort Differences in Improvement in Reading Scores from 2004 Relative **EO**

VARIABLES	OLS (no controls) 1	OLS (controls) 2	FE (Fixed Effects) 3	Mixed Effects (no controls) 4	Mixed Effects (controls) 5
BBO					
All Years in BB					
Difference 2005 (Grade 3)	10.02**	9.416*	10.35**	10.40**	9.624**
	(5.043)	(5.107)	(5.039)	(4.185)	(4.204)
Difference 2006 (Grade 4)	13.02***	12.58***	13.42***	13.67***	12.80***
	(4.618)	(4.677)	(4.649)	(4.490)	(4.595)
Difference 2007 (Grade 5)	8.979**	7.647*	9.402**	9.455***	7.883**
	(4.362)	(4.398)	(4.393)	(3.597)	(3.728)
BBL					
4 or 5 Years in BB					
Difference 2005 (Grade 3)	3.978	3.606	4.314	5.339	4.694
	(3.656)	(3.650)	(3.673)	(3.829)	(3.618)
Difference 2006 (Grade 4)	6.613**	6.277*	6.957**	8.333**	7.338*
	(3.353)	(3.379)	(3.383)	(3.829)	(3.838)
Difference 2007 (Grade 5)	6.476*	6.092	6.778*	7.571**	6.656*
	(3.750)	(3.789)	(3.744)	(3.698)	(3.690)
BBS					
1,2,or 3 years in BB					
Difference 2005 (Grade 3)	-3.740	-3.579	-3.661	-3.757	-3.490
	(4.829)	(4.894)	(4.847)	(5.293)	(5.276)
Difference 2006 (Grade 4)	7.455*	7.688*	7.617*	7.651*	7.981*
	(4.161)	(4.206)	(4.187)	(4.306)	(4.446)
Difference 2007 (Grade 5)	10.36**	11.18**	10.99**	11.06**	10.88**
	(4.561)	(4.646)	(4.584)	(4.572)	(4.952)
Observations	4,784	4,784	4,784	4,784	4,784
R-squared	0.365	0.394	0.624		
Number of groups				309	309
Number of id			1,196		

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The reading score is somewhat of a different story. We see positive and significant coefficients for the **BBO** groups, indicating better reading performance compared with the **EO** group. Although it appears that these results contrast with

results for other categories, interpretation is slightly different, since the reference year is 2004 and not 2002. Interpretation is different since in 2002, all groups had the same overall score (220) whereas this was not the case for the 2004 reading score. Therefore we may see a relative improvement for the BB groups compared to the **EO** simply because there was less room for improvement in the **EO** group (i.e. a ceiling effect). We don't see significant effects for the other shorter BB groups.

Table 4.4 Regressions on CELDT Writing Means for Analysis 1 Cohort Differences in Improvement in Reading Scores from 2004 Relative **EO**

VARIABLES	OLS (no controls) 1	OLS (controls) 2	FE (Fixed Effects) 3	Mixed Effects (no controls) 4	Mixed Effects (controls) 5
BBO					
All Years in BB					
Difference 2005 (Grade 3)	8.536*	7.793	8.151*	8.655*	8.390*
	(4.856)	(4.949)	(4.883)	(4.490)	(4.508)
Difference 2006 (Grade 4)	21.14***	20.57***	20.81***	21.44***	21.19***
	(5.555)	(5.630)	(5.608)	(4.792)	(4.803)
Difference 2007 (Grade 5)	23.92***	22.58***	24.24***	24.12***	23.33***
	(5.572)	(5.657)	(5.623)	(5.098)	(5.001)
BBL					
4 or 5 Years in BB					
Difference 2005	8.580***	8.027**	8.280***	9.430***	9.259***
	(3.120)	(3.150)	(3.131)	(3.031)	(3.014)
Difference 2006	19.97***	19.50***	19.59***	21.03***	20.80***
	(3.399)	(3.449)	(3.413)	(3.468)	(3.489)
Difference 2007	24.11***	23.59***	23.81***	24.63***	24.38***
	(3.854)	(3.894)	(3.880)	(3.531)	(3.562)
BBS					
1,2,or 3 years in BB					
Difference 2005	11.78**	11.76**	11.57**	11.67*	11.81*
	(4.755)	(4.789)	(4.796)	(6.346)	(6.242)
Difference 2006	20.51***	20.44***	20.10***	20.47***	20.47***
	(5.271)	(5.329)	(5.322)	(7.545)	(7.459)
Difference 2007	26.86***	26.74***	26.95***	27.14***	26.90***
	(5.632)	(5.680)	(5.758)	(6.919)	(6.891)
Observations	4,784	4,784	4,784	4,784	4,784
R-squared	0.309	0.348	0.525		
Number of groups				309	309
Number of id			1,196		

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The results shown on Table 4.4 are somewhat more compatible with Overall and Listening/Speaking results. For 2005 all of the coefficients are less than one for each group. Only one third of those are significant, and all of them are negative. For the other years all of the coefficients are positive for all BB cohorts. **BBO** exhibits the highest score improvement in 2006 relative to **EO** while **BBL** has greater

improvement in 2007. Like the Reading results, all of the coefficients are positive, indicating stronger improvement relative **EO**. Again, we may see a relative improvement for the BB groups compared to the **EI** simply because of EI ceiling effects.

The results shown on Table 4.3 (CELDT Reading scores) are different. None of the coefficients for the **BBS** group are statistically significant for any year. Also, none of the coefficients are statistically significant for the any of the BB groups in 2005. All of the coefficients are significant for 2006 and 2007 for **BBO**, while only one third of the coefficients are significant for **BBL** for those years. The effects sizes are much larger in 2006 and 2007 for all groups, and all group coefficients are positive for those years, showing each BB group increase more each year than did the **EO** group of Reading.

The results shown on Table 4.4 are somewhat more compatible with Overall and Listening/Speaking results. For 2005 all of the coefficients are less than one for each group. Only one third of those are significant, and all of them are negative. For the other years all of the coefficients are positive for all BB cohorts. **BBO** exhibits the highest score improvement in 2006 relative to **EO** while **BBL** has greater improvement in 2007.

4.1.2 The Analysis 2 Cohort

The goal of the means analysis for the Analysis 2 cohort was to measure differences between **EI** and **BB** in average annual score improvement. This differs from the first analysis in that it looks at all years combined and takes the average

annual improvement instead of looking at differences for each year individually. Groups for this analysis are selected by including all students that were in **BB** or **EI** for at least 2002 and 2003. It is necessary to have at least two years of data in order to measure annual changes. Students drop out of the sample if they switch to a different group classification. For example, if a student is in **EI** from 2002 to 2003 and then switches to **BB** in 2004, they will be dropped in 2004 and for all remaining years. Also, only students that represent continuous enrollment starting in 2002 are included. This means that if a student is in **BB** from 2002-2004, but in **EI** from 2005-2007, only the 2002-2004 years will be included. The years after switching are not included since there may be a complimentary of effect of starting in one group and switching to another (e.g. students that start in **BB** and then switch to **EI** may perform better than students that are in **EI** for all years).

All of the same models as the analysis for the Analysis 1 Group are used to measure differences in rate of acquisition. However, in contrast to the earlier analysis the interaction term that identifies the **BB**-**EI** difference in rate of acquisition considers years as one continuous variable as opposed to binary variables for each year. The resulting coefficients on these variables (Group X Year) thus represent the difference in average annual improvement between **EI** and **BB**. Dummy variables for each year are also included in the model, which controls for time trends. For example, there were generally large improvements in earlier years and the **BB** population is much bigger in earlier years. If time trends are not controlled for, the model would erroneously estimate **BB** as much better than **EI**. However, this is only because more students are in **BB** in years where there are

larger improvements. All other variables are treated the same as the analysis for the Analysis 1 cohort.

The Analysis 2 analyses adds value over the analyses for the Analysis 1 Group in that it does not require students to be continuously enrolled, which (1) increases sample size and (2) includes students who may have switched out of **BB** to a different program in later years. The Analysis 2 regressions are shown below.

Table 4.5 Regressions on CELDT Overall Means for Analysis 2 Cohort

VARIABLES	OLS (no controls) 1	OLS (controls) 2	FE (Fixed Effects) 3	Mixed Effects (no controls) 4	Mixed Effects (controls) 5
Average Annual Increase Group X Year (EI relative to BB)	**3.174*** **	**3.259*** **	**3.776*** **	**3.466*** **	**3.548*** **
	(1.025)	(0.991)	(0.700)	(1.128)	(1.087)
Group (EI)	-6,331***	-6,502***		-6,917***	-7,082***
	(2,053)	(1,986)		(2,259)	(2,178)
Year=2003 (Grade 1)	228.4***	228.2***	170.1***	227.7***	227.6***
	(2.727)	(2.764)	(1.148)	(2.776)	(2.791)
Year=2004 (Grade 2)	257.6***	257.4***	140.1***	257.1***	256.9***
	(2.335)	(2.327)	(0.810)	(2.449)	(2.424)
Year=2005 (Grade 3)	261.1***	260.8***	84.37***	260.2***	260.0***
	(2.988)	(2.973)	(0.858)	(3.261)	(3.208)
Year=2006 (Grade 4)	285.2***	284.7***	52.28***	284.2***	283.8***
	(3.867)	(3.852)	(0.931)	(4.264)	(4.194)
Year=2007 (Grade 5)	287.1***	286.5***		286.3***	285.6***
	(4.861)	(4.807)		(5.434)	(5.345)
	(1.872)	(9.736)	(518.4)	(2.107)	(11.65)
Observations	11,977	11,977	11,977	11,977	11,977
R-squared	0.888	0.890	0.929		
Number of groups			2,731	381	381

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

This table shows that the average annual rate of increase for the **EI** group was 3.5 points more than the average annual rate of increase for the **BB** group. [The coefficient on EI should not be interpreted since it measures the difference when year = 0.]

Table 4.6 Regressions on CELDT Listening/Speaking Means for Analysis 2 Cohort

VARIABLES	OLS (no controls) 1	OLS (controls) 2	FE (Fixed Effects) 3	Mixed Effects (no controls) 4	Mixed Effects (controls) 5
Average Annual Increase					
Group X Year (EI relative to BB)	**5.655*** **	**5.575*** **	**5.950*** **	**5.903*** **	**5.853*** **
	(1.328)	(1.348)	(0.797)	(1.400)	(1.403)
Group (EI)	-11,300***	-11,142***		-11,798***	-11,699***
	(2,660)	(2,699)		(2,803)	(2,810)
Year (Relative to 2003)					
Year=2003 (Grade 1)	226.5***	226.4***	154.1***	225.9***	225.8***
	(2.772)	(2.814)	(1.137)	(2.832)	(2.854)
Year=2004 (Grade 2)	284.2***	284.3***	138.6***	283.7***	283.9***
	(2.941)	(2.986)	(1.046)	(3.028)	(3.053)
Year=2005 (Grade 3)	263.5***	263.7***	44.91***	262.8***	262.9***
	(3.916)	(3.998)	(1.095)	(4.127)	(4.157)
Year=2006 (Grade 4)	288.8***	288.7***		288.1***	287.9***
	(5.217)	(5.371)		(5.490)	(5.570)
Observations	10,922	10,922	10,922	10,922	10,922
R-squared	0.878	0.880	0.920		
Number of groups			2,731	370	370

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

This table shows that the average annual rate of increase for the **EI** group was 5.6 points more than the average annual rate of increase for the **BB** group.

Table 4.7 Regressions on CELDT Reading Means for Analysis 2 Cohort

VARIABLES	OLS (no controls) 1	OLS (controls) 2	FE (Fixed Effects) 3	Mixed Effects (no controls) 4	Mixed Effects (controls) 5
Average Annual Increase Group X Year (EI relative to BB)	**-4.195****	**-3.647****	**-3.238*****	**-4.186*****	**-3.852****
	(1.633)	(1.526)	(1.175)	(1.528)	(1.501)
Group (EI)	8,435**	7,334**		8,418***	7,748**
	(3,274)	(3,059)		(3,063)	(3,008)
Year (Relative to 2004)					
Year=2005 (Grade 3)	27.83***	27.54***	-4.729***	28.12***	27.93***
	(1.679)	(1.619)	(0.846)	(1.678)	(1.650)
Year=2006 (Grade 4)	58.56***	58.08***	-0.741	58.93***	58.81***
	(3.009)	(2.777)	(0.892)	(2.859)	(2.779)
Year=2007 (Grade 5)	82.88***	82.03***		84.10***	83.63***
Observations	6,515	6,515	6,515	6,515	6,515
R-squared	0.295	0.324	0.586		
Number of groups			2,250	359	359

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

This table shows that the average annual rate of increase for the **EI** group was 4.2 points less than the average annual rate of increase for the **BB** group. Again, this may be because the **BB** group scores were lower than the **EI** group scores in 2004 (the first writing year).

Table 4.8 Regressions on CELDT Writing Means for Analysis 2 Cohort

VARIABLES	OLS (no controls) 1	OLS (controls) 2	FE (Fixed Effects) 3	Mixed Effects (no controls) 4	Mixed Effects (controls) 5
Average Annual Increase					
Group X Year (EI relative to BB)	-8.293***	-7.834***	-8.574***	-8.833***	-8.467***
	(1.839)	(1.672)	(1.314)	(1.565)	(1.537)
Group (EI)	16,664***	15,741***		17,746***	17,010***
	(3,687)	(3,353)		(3,137)	(3,081)
Year (Relative to 2004)					
Year=2005 (Grade 3)	38.43***	38.37***	8.558***	39.24***	39.11***
	(1.928)	(1.798)	(0.800)	(1.710)	(1.694)
Year=2006 (Grade 4)	65.33***	64.93***	11.03***	66.72***	66.41***
	(3.429)	(3.126)	(0.899)	(2.895)	(2.859)
Year=2007 (Grade 5)	75.95***	75.02***		78.52***	77.75***
	(5.255)	(4.774)		(4.393)	(4.324)
Observations	6,515	6,515	6,515	6,515	6,515
R-squared	0.276	0.307	0.485		
Number of groups			2,250	359	359

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

This table shows that the average annual rate of increase for the **EI** group was 8.3 points less than the average annual rate of increase for the **BB** group.

For Tables 4.5 (Overall CELDT scores), 4.6 (Listening/Speaking CELDT scores), and 4.8 (CELDT Writing) all of the coefficients are statistically significant, and nearly all the coefficients in Table 4.6.3 (CELDT Reading) are statistically significant. Like the tables for the Analysis 1 cohort, there is little variation from column to column indicating that there is little difference in the results of the statistical models being used. This was found for all of the CELDT scores.

4.1.3 Conclusions from the Analysis 1 and the Analysis 2 Cohort Means Analyses

With the exception of CELDT Reading in the Analysis 1 cohort (Table 4.3), the results are statistically significant across the ten analyses performed on the CELDT score means for the Analysis 1 and the Analysis 2 cohorts. This means that we can have confidence that the CELDT scores were influenced by the type of program the ELL students were in. The one exception, Table 4.3, indicates that by that analysis, the effects of the CELDT Reading scores were not influenced by either program. Therefore, the overall results indicate that the type of program students enroll in do make a difference. From here, we can have confidence that the graphs and tables that are generated to compare program effectiveness are meaningful. These comparisons will help answer *Research Question 1*.

4.2 Research Question 1: Graphs and Tables for CELDT Means Across Programs

This section presents tables, from which the graphs are made, of the CELDT means from the Analysis 1 and the Analysis 2 cohorts to establish program effectiveness for ELLs. This program comparison is compartmentalized by discussing CELDT scores in this order: Overall, Listening/Speaking, Reading, and Writing. The Analysis 1 results are compared to the Analysis 2 results and are discussed. Conclusions are then drawn regarding program effectiveness following a discussion of the section results.

4.2.1 Analysis 1 and Analysis 2 CELDT Overall Results

Table 4.9 Annual Overall CELDT Mean Scores for the Analysis 1 Cohort

Year/Grade	EO	BBS		BBL		BBO	
2002 K	220.00	220.00	PG	220.00	PG	220.00	PG
2003 G1	463.75	427.14	-36	410.05	-54	415.45	-49
2004 G2	486.20	465.64	-20	451.18	-35	446.03	-40
2005 G3	487.06	469.59	-17	463.17	-24	458.01	-28
2006 G4	515.51	508.87	-7	495.78	-20	492.48	-24
2007 G5	527.55	527.49	0	512.00	-16	510.19	-18

PG = Point Gap difference between the BB groups and the EO group

This figure shows the annual raw means from which Graph 1 was generated. It also illustrates that each group began at the same English proficiency level. In addition, it records the raw scale score point gap (PG) between each TBE group and the SEI group. The point gap is to the right of each BB group. For example, in 2003 the point gap between **EO** and **BBS** is "36" and the point gap between **EO** and **BBL** is "54". A positive number in the PG column would indicate that the BB group scored higher than the **EO** group for that year. The following key explains the meanings of the abbreviations.

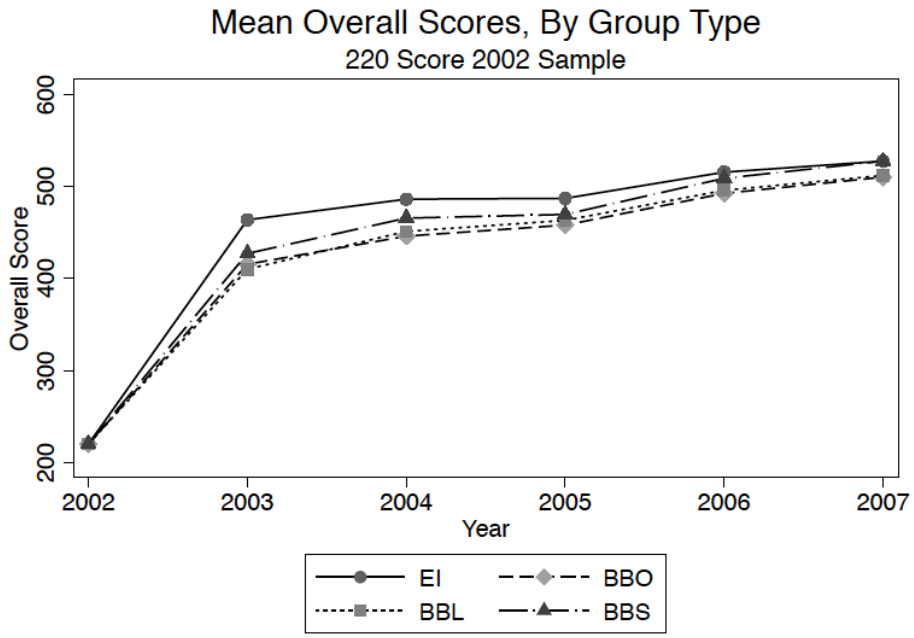
Table 4.10 Annual Overall CELDT Mean Scores for the Analysis 2 Cohort

Year/Grade	EI	BB	PG
2002 K	220	220	0
2003 G1	463.34	409.7567	-54
2004 G2	491.8098	450.7035	-41
2005 G3	494.9706	468.3743	-27
2006 G4	522.1062	496.2337	-26
2007 G5	526.8376	509.4146	-17

PG = Point Gap difference between the BB group and the EI group

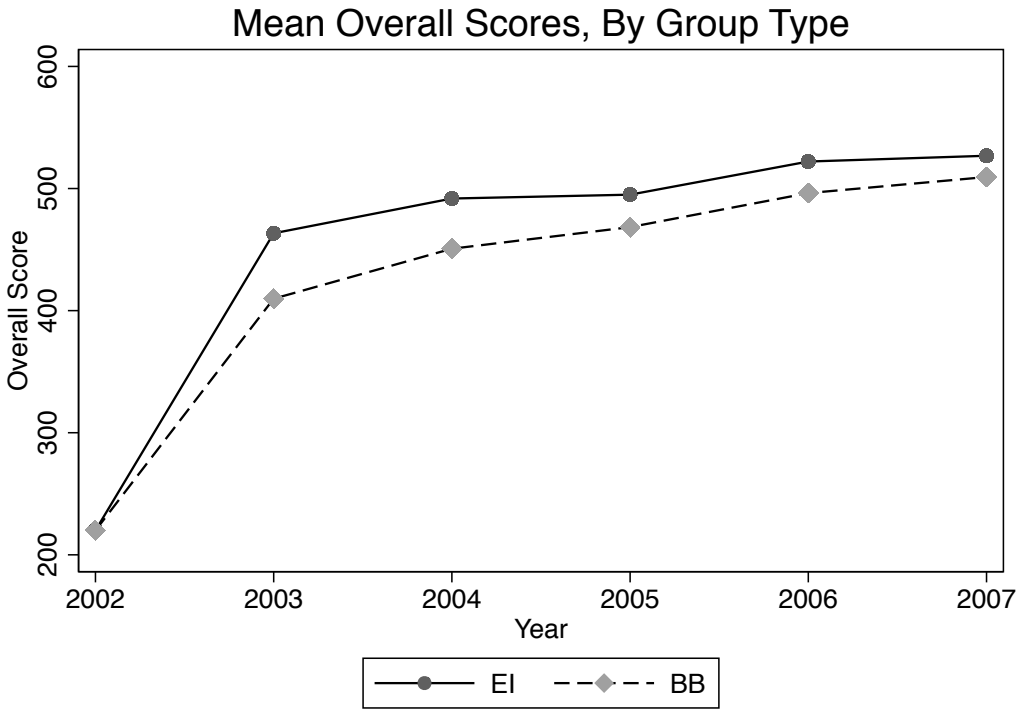
Starting with 2003, this figure shows the annual raw means from which Graph 4.1 was generated. It also shows the yearly scale score point gap between **EI** and **BB**. A negative score indicates the **BB** score is lower than the **EI** score for that year.

Graph 4.1 Annual CELDT Overall Mean Scores for the Analysis 1 Cohort



This graph shows the yearly raw means for the CELDT Overall scale scores for each group in the Analysis 1 cohort. Overall is a composite oral score from 2002 – 2003, and a composite oral and literacy score from 2004 onward.

Graph 4.2 Annual CELDT Overall Mean Scores for the Analysis 2 Cohort



This graph shows the yearly raw means of the CELDT Overall scale scores for both groups in the Analysis 2 cohort.

The data from Table 4.1 show that all BB groups and **EO** evidence a sharp increase in English Overall skills during the first year and then exhibit a very gradual growth in the subsequent years. The **EO** group scores higher than the other groups until 2007, when the **BBS** group matches their score. **BBS** scores are slightly higher than the other BB groups for the entire 5 years, and except for 2003, **BBL** scores higher than **BBO** the entire 5 years. Each year all BB cohorts narrow the distance between their scores and the **EO** group's score. **BBL** and **BBO** score more closely to one another than they do to the **EO** group until 2007, when the gap begins to close more dramatically between all groups. By 2007 the **BBS** cohort's scores merge with the **EO** cohort, and the other two BB cohorts effectively merge.

The data from Table 4.2 show yearly mean trends that are nearly identical to the Analysis 1 cohort, so the conclusions are the same. From the standpoint of CELDT Overall scores then, the results are as follows: (1) Students in Structured English immersion programs learn English at a faster rate initially than students in in bilingual programs; (2) the gap in group mean scores that this initial lead exhibits diminishes each year so that after five years in a respective program, the differences in mean scores are greatly reduced; (3) the longer students are in a bilingual program, the longer it will take to reduce the differences in mean scores compared to students in an Structured English immersion program.

4.2.2 Analysis 1 and Analysis 2 CELDT Listening/Speaking Results

Table 4. 11 Annual Listening/Speaking CELDT Mean Scores for the Analysis 1 Cohort

Year/Grade	EO	BBS		BBL		BBO	
2002 K	220.00	220.00	PG	220.00	PG	220.00	PG
2003 G-1	463.75	427.14	-37	410.05	-54	415.45	-49
2004 G-2	516.32	493.52	-22	477.22	-39	474.95	-41
2005 G-3	495.34	474.74	-20	472.26	-23	467.04	-28
2006 G-4	527.88	519.01	-9	506.03	-22	503.75	-24
2007 G-5	534.67	523.04	-11.63	522.24	-12.43	524.78	-9.89

PG = Point Gap difference between the BB groups and the EO group

This figure shows the annual raw means from which Graph 2 was generated. It also illustrates that each group began at the same English proficiency level. In addition, it records the raw scale score point gap (PG) between each BB group and the **EO** group.

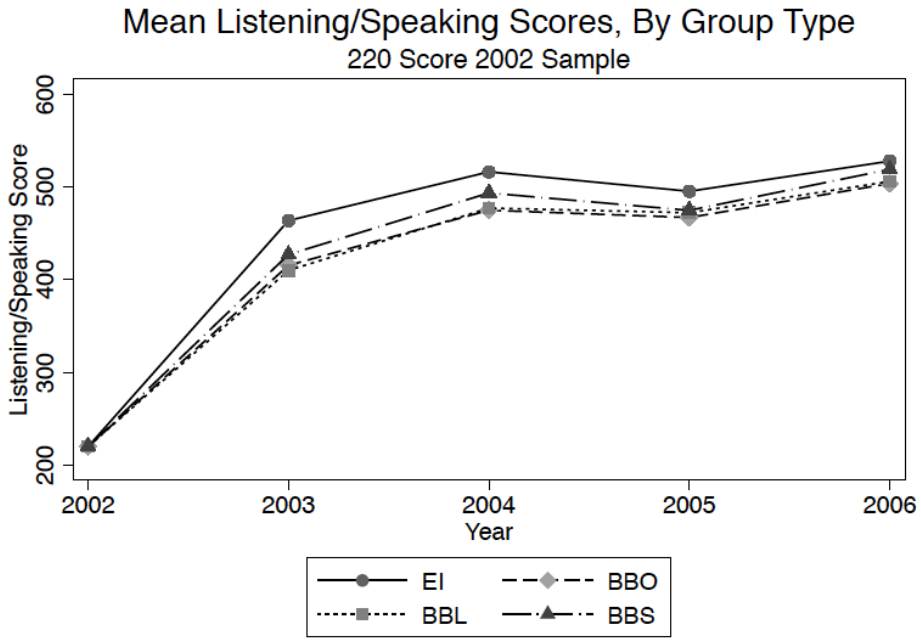
Table 4.12 Annual Listening/Speaking CELDT Mean Scores for the Analysis 2 Cohort

Year/Grade	EI	BB	PG
2002 K	220	220	-53
2003 G-1	463.34	409.7567	-44
2004 G-2	522.8903	478.79	-26
2005 G-3	503.2107	477.4551	-28
2006 G-4	518.25	521.64	+3

PG = Point Gap difference between the BB group and the EI group

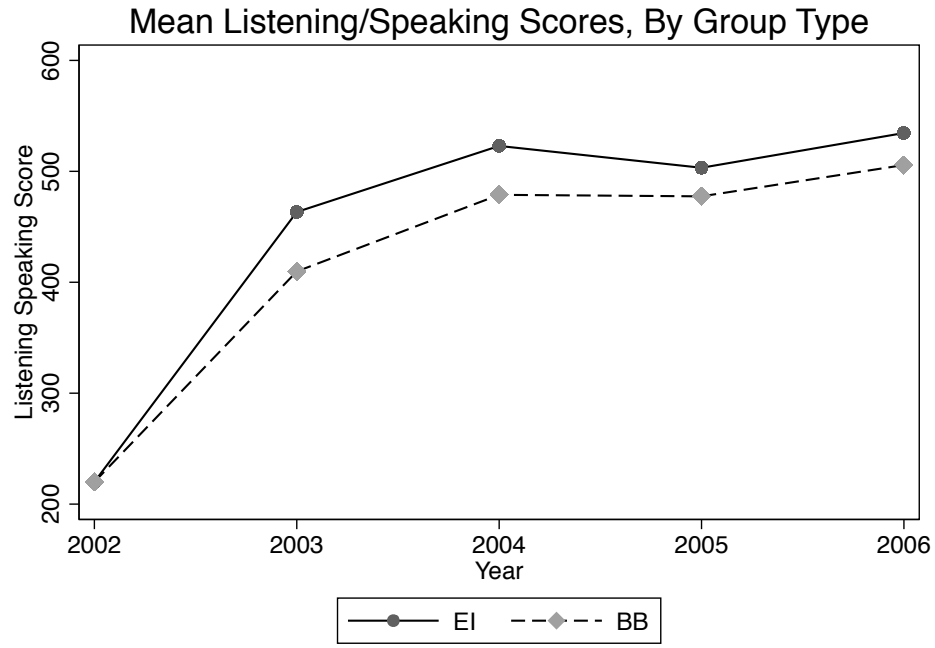
Starting with 2003, this figure shows the annual raw means of the SEI and TBE groups in Analysis 2. It also shows the yearly scale score point gap between **EI** and **BB**.

Graph 4.3 Annual CELDT Oral Mean Scores for the Analysis 1 Cohort



This graph shows the yearly raw means of the CELDT Listening/Speaking scale scores for each sub-cohort in the Analysis 1 cohort.

Graph 4.4 Annual CELDT Oral Mean Scores for the Analysis 2 Cohort



This graph shows the yearly raw means of the CELDT Listening/Speaking scale scores for both groups in the Analysis 2 cohort.

For Listening/Speaking in the Analysis 1 cohort, the **EO** group scores begin higher and remain higher for the entire period. The BB groups virtually merge in 2007 ending very close to the **EO** mean. All group means drop slightly in 2005 before moving up again in 2006; nevertheless, each BB group narrowed the point gap for those years, and narrowed the gap for each of the other years. The Analysis 2 scores were again very similar to the Analysis 1 scores, with the interesting difference in **BBO** scoring higher than **EI** in 2007. In fact, all of the BB groups in Analysis 1 scored higher than **EI** in 2007; however, it should be noted that this coincided with a drop in **EI**'s score from 2006. Nevertheless, by 2007 the BB and EO scores when averaged together were nearly identical.

Again, it is clear that while students in Structured English immersion programs enjoy an initial and sustained advantage in oral skills, the bilingual groups narrow the gap annually and nearly close it after five years. The same trend that more time in bilingual programs resulted in lower scores holds true for oral proficiency, but again, the point gap nearly disappears in five years.

4.2.3 Analysis 1 and Analysis 2 CELDT Reading Results

Table 4.13 Annual Reading CELDT Mean Scores for the Analysis 2 Cohort

Year/Grade	EO	BBS	PG	BBL	PG	BBO	PG
2004 G-2	443.28	432.67	-10	420.95	-22	413.14	-30
2005 G-3	462.86	448.51	-14	444.51	-18	442.74	-20
2006 G-4	491.33	488.17	-3	475.61	-15	474.2	-17
2007 G-5	519.9	519.65	-0	504.05	-16	498.74	-21

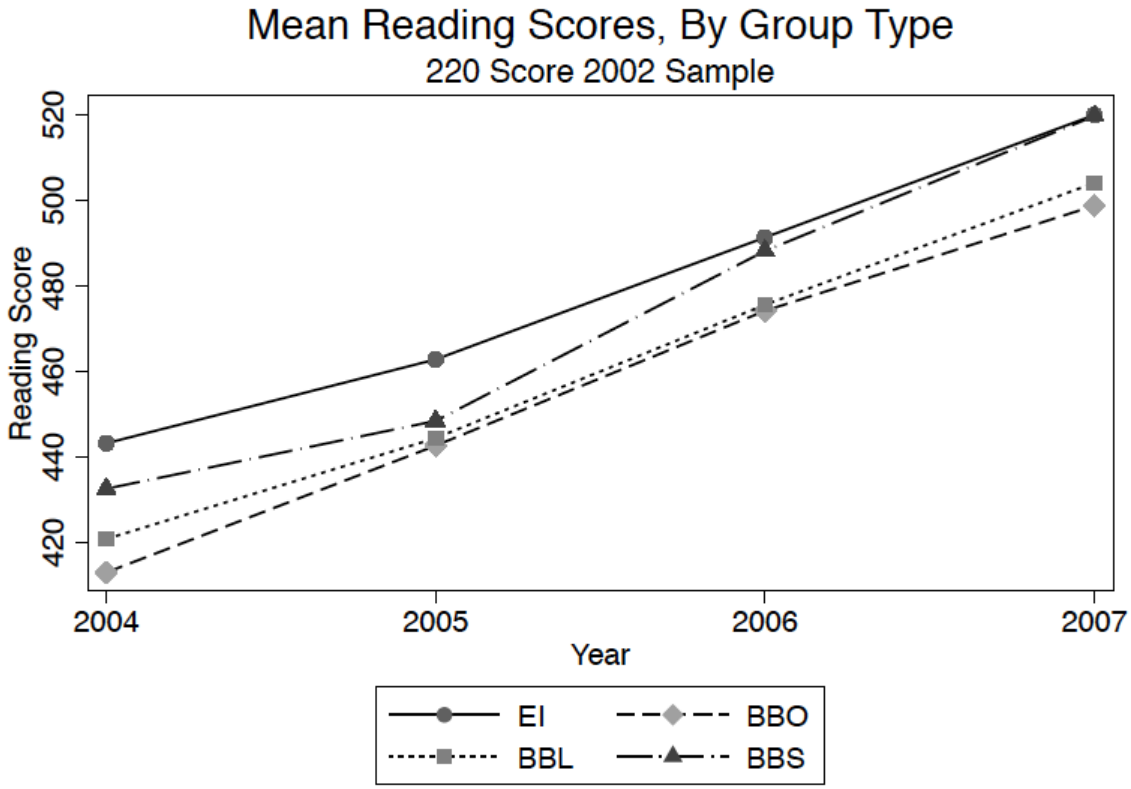
PG = Point Gap difference between the BB groups and the EO group
This figure shows the annual raw means from which Graph 4.5 was generated.
It also illustrates that each group began at a different starting point in 2004.

Table 4.14 Annual CELDT Reading Mean Scores for the Analysis 2 Cohort

Year/Grade	EI	BB	PG
2004 G-2	449.0753	418.158	-31
2005 G-3	471.4856	451.1168	-20
2006 G-4	498.7451	478.75	-20
2007 G-5	519.3505	497.5854	-22

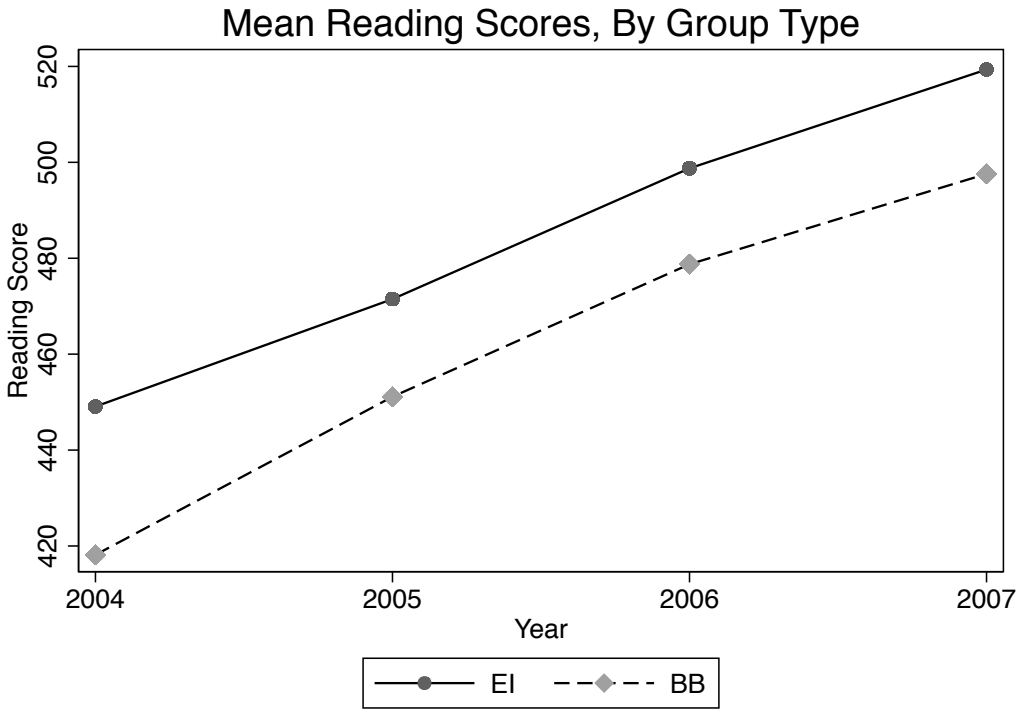
PG = Point Gap difference between the BB group and the EI group
Starting with 2004, this figure shows the annual raw means from which Graph 4.6 was generated. It also shows the yearly scale score point gap between **EI** and **BB**.

Graph 4.5 Annual CELDT Reading Mean Scores for the Analysis 1 Cohort



This graph shows the yearly raw means of the CELDT Reading scale scores for each group in the Analysis 1 cohort.

Graph 4.6 Annual CELDT Reading Mean Scores for the Analysis 1 Cohort



This graph shows the yearly raw means of the CELDT Reading scale scores for both cohorts in the Analysis 2 cohort.

For the Analysis 1 cohort, Table 4.13 shows a 4-year trend with a consistent ranking of highest to lowest scores, which is: **EO**, **BBS**, **BBL,** and **BBO**. With the exception of **BBS** in 2005, every BB group narrows the point gap each year with **EO**. The **BBS** group nearly merges with the **EO** group in 2006, and closes the gap in 2007. The scores of **BBL** and **BBO** are closer to each other than to **BBS** and **EO**.

The Analysis 2 cohort follows a nearly identical trajectory as the Analysis 1 cohort. The same general trends hold for the Reading scores as the Oral scores: (1) the English only groups start out ahead and remain ahead; (2) the means are higher the less bilingual education the students have; and (3) the gaps in mean points continue to close over the five-year study. However, there is a wider point gap

between the bilingual and the Structured English immersion groups for Reading than for Overall or Listening/Speaking. And, the gap between the two cohorts begins to widen in 2007.

4.2.4 Analysis 1 and Analysis 2 CELDT Writing Results

Table 4.15 Annual CELDT Writing Mean Scores for the Analysis 1 Cohort

Year/ Grade	EO	BBS	PG	BBL	PG	BBO	PG
2004 G-2	470.41	444.36	-26	430.83	-39	422.49	-48
2005 G-3	496.22	481.95	-14	465.22	-31	456.84	-39
2006 G-4	516.49	510.95	-5	496.88	-19	489.71	-26
2007 G-5	525.52	526.34	+1	510.05	-16	501.53	-24

PG = Point Gap difference from the BB groups and the EO group

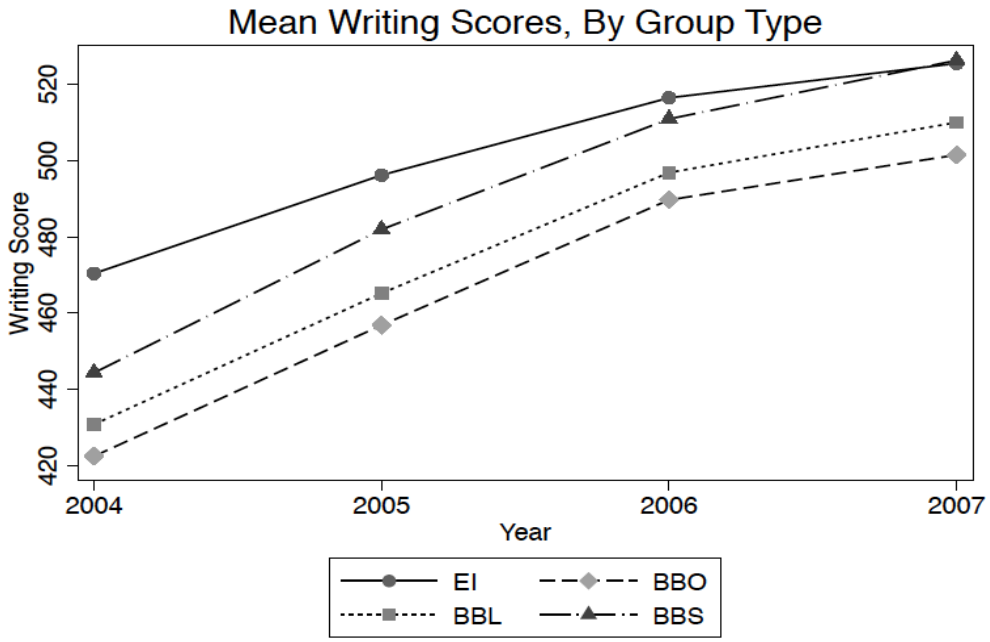
This figure shows the annual raw means from which Graph 4.7 was generated. It also illustrates that each group began at a different starting point in 2004. In addition, it includes the yearly point gap between **EO** and the BB groups.

Table 4.16 Annual CELDT Writing Mean Scores for the Analysis 2 Cohort

Year/Grade	EI	BB	PG
2004 G-2	473.8786	428.5498	-45
2005 G-3	503.5298	469.012	-35
2006 G-4	522.1597	496.3804	-26
2007 G-5	525.0863	501.1341	-24

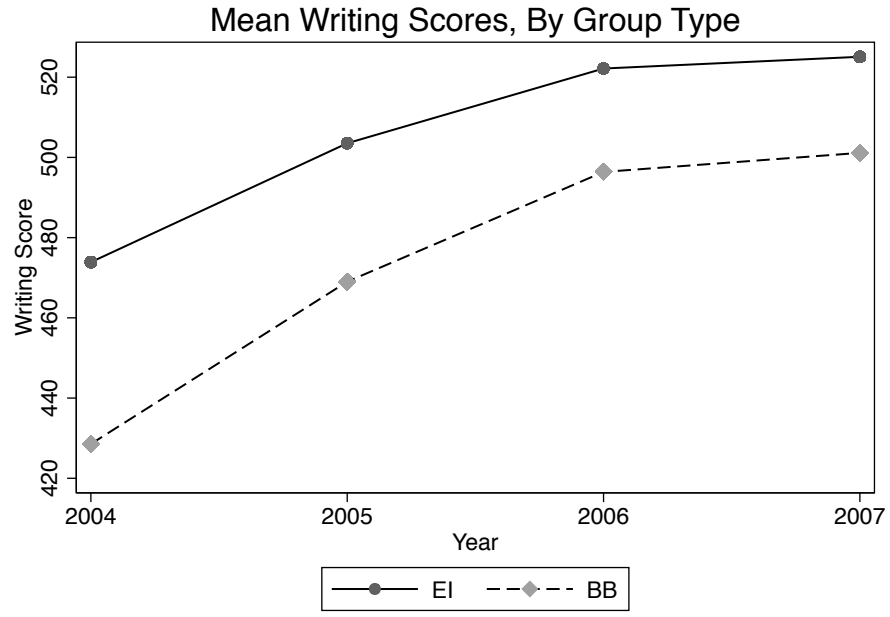
PG = Point Gap difference from the BB group and the EI group
Starting with 2004, this figure shows the annual raw means from which Graph 4.8 was generated. It also shows the yearly scale score point gap between **EI** and **BB**.

Graph 4.7 Annual CELDT Reading Mean Scores for the Analysis 1 Cohort



This graph shows the yearly raw means of the CELDT Writing scale scores for each group in the Analysis 1 Cohort.

Graph 4.8 Annual CELDT Reading Mean Scores for the Analysis 1 Cohort



This graph shows the yearly raw means of the CELDT Writing scale scores for both cohorts in the Analysis 2 cohort.

For the Analysis 1 cohort, writing Table 4.15 shows a consistent trend with the **EO** group scoring higher each year than the BB groups until 2007, when the **BBS** group scores slightly higher than **EO**. Like the previous tables, the next highest scores are followed throughout by **BBL** and **BBO** respectively. Each BB group narrowed the gap with **EO** for all years. The Analysis 2 group follows a similar pattern for each year. However, the scores remained basically the same for **EI** and **BB** in 2006 and 2007.

Taken as a whole, the CELDT tables for each English proficiency measurement illustrate the slight advantage that Structured English immersion programs have over bilingual programs. The results are consistent for all years. Event though the point gap closes from year to year in all English measures, the bilingual group scores more closely to the Structured English immersion group in oral skills than it does in literacy skills.

4.3 Research Question 1: CELDT Proficiency Levels

When students receive a score on a section of the CELDT, that score is converted to a proficiency level. The proficiency levels are Beginning, Early Intermediate, Early Advanced, and Advanced. Since each section of the CELDT is scored separately before assigning the student a composite Overall score, students can be placed into different levels for Overall, Listening/Speaking, Reading, and Writing. This next section, instead of focusing on group means for the students in each language program, focuses on the CELDT levels the students attain from year

to year. The order of analysis proceeds as in the last section, from Overall, to Oral, to Reading, and then to Writing. The key from Chapter 4 is shown below.

Key to CELDT Level Tables

Beg = Beginning

E Int = Early Intermediate

Int = Intermediate

E Adv = Early Advanced (not shown above)

Adv = Advanced (not shown above)

(L) = score in the low range (lower third) of a given level; i.e., Int(L) means that the group scored at the Intermediate level, but in the bottom third of the Intermediate range.

(H) = score in the high range (upper third) of a given level; i.e., Int(H) means that the group scored at the Intermediate level, but in the upper third of the Intermediate range.

These tables show annual CELDT proficiency levels by cohort group. When students take the CELDT, their raw score is converted to a proficiency level, which helps the school district to place them in an appropriate English-language class. Students receive different scores for each separate section of the CELDT (oral, reading, and writing). From there a composite score is generated, which is referred to as the "Overall" score. It is possible for a student to receive a low score on one section of the CELDT and a higher on another section. These tables are therefore useful in comparing the effectiveness of bilingual and Structured English immersion class instruction on English acquisition. This comparison can help answer a main question of this study, "Are bilingual programs better than Structured English

immersion programs in helping ELLs acquire English?" Below is a Key to understanding the CELDT level tables.

4.3.1 Analysis 1 and Analysis 2 CELDT Overall Proficiency Levels

Table 4.17 Annual Overall CELDT levels for the Analysis 1 Cohort

Year/ Grade	2003 G-1	2004 G-2	2005 G-3	2006 G-4	2007 G-5
EO	E Int (L)	Int (L)	Int (L)	Int	Int
BBS	E Int (L)	E Int (H)	E Int	Int	Int
BBL	Beg	E Int	E Int	Int	Int
BBO	Beg	E Int (L)	E Int (L)	Int	Int

This figure shows the CELDT levels that correspond to the raw means listed in Table 4.9.

Table 4.18 Annual Overall CELDT levels for the Analysis 2 Cohort

Year/ Grade	2003 G-1	2004 G-2	2005 G-3	2006 G-4	2007 G-5
EI	E Int	Int (L)	Int	Int (H)	Int
BB	Beg	E Int (L)	E Int	Int	Int

This figure shows the CELDT levels that correspond to the raw means listed in Table 4.10.

In both the Analysis 1 and the Analysis 2 cohorts the EO and BB groups are nearly identical. **BBS** levels are very close to **EO**, and **BBL** is identical to **BB(O)**, all of which follow earlier patterns. The group with less bilingual influence continues to show an advantage (albeit slight) over the other BB groups. Both cohorts show very marginal growth across the five-year study, and both cohorts the BB groups catch up to the EO groups by 2006.

4.3.2 Analysis 1 and Analysis 2 CELDT Oral Proficiency Levels

Table 4.19 Annual Oral CELDT levels for the Analysis 1 Cohort

Year/ Grade	2003 G-1	2004 G-2	2005 G-3	2006 G-4	2007 G-5
EO	E Int	Int	Int	E Adv (L)	E Adv
BBS	E Int (L)	E Int (H)	E Int	Int	E Adv (L)
BBL	Beg	E Int	E Int	Int	E Adv (L)
BBO	Beg	E Int	E Int	Int	E Adv (L)

This figure shows the CELDT levels that correspond to the raw means listed in Table 4.11.

Table 4.20 Annual Oral CELDT levels for the Analysis 1 Cohort

Year/ Grade	2003 G-1	2004 G-2	2005 G-3	2006 G-4	2007 G-5
EI	E Int H	Int	Int	E Adv (L)	Int (H)
BB	Beg	E Int	E Int (H)	Int	E Adv (L)

This figure shows the CELDT levels that correspond to the raw means listed in Table 4.12.

For both cohorts, SEI groups lead the way up through 2006. Then in 2007 all groups, except **EI** (which remained close in the high range of Intermediate) attained the Early Advance level (although all of the BB groups were in the low range of that level). In the first two years of the study **BBS** was slightly above the other BB groups, but all BBs were the same for the last three years.

4.3.3 Analysis 1 and Analysis 2 CELDT Reading Proficiency Levels

Table 4.21 Annual Reading CELDT levels for the Analysis 1 Cohort

Year/ Grade	2004 G-2	2005 G-3	2006 G-4	2007 G-5
EO	E Int (L)	E Int (L)	E Int (H)	Int
BBS	Beg (H)	Beg (H)	E Int	Int
BBL	Beg	Beg	E Int	Int (L)
BBO	Beg	Beg	E Int	E Int (H)

This figure shows the CELDT levels that correspond to the raw means listed in Table 4.13.

Table 4.22 Annual Reading CELDT levels for the Analysis 2 Cohort

Year/ Grade	2004 G-2	2005 G-3	2006 G-4	2007 G-5
EI	E Int	E Int	Int (L)	Int (L)
BB	Beg	Beg	E Int (L)	E Int

This figure shows the CELDT levels that correspond to the raw means listed in Table 4.14.

BBO and **BB** exhibited identical levels for the entire study. Again, in the first two years **BBS** scored slightly higher than the other BB groups. By 2006 all groups achieved Early Intermediate, except **EI**, which although in the Intermediate level, was in the lower range. Only **BBO** and **BB** remained at the Early Intermediate level in 2007. One defining characteristic from the Reading data is the very gradual progress for all groups—it took three years for all groups to advance one level.

4.3.4 Analysis 1 and Analysis 2 CELDT Writing Proficiency Levels

Table 4.23 Annual Writing CELDT levels for the Analysis 1 Cohort

Year/ Grade	2004 G-2	2005 G-3	2006 G-4	2007 G-5
EO	Int (L)	Int	Int	Int
BBS	E Int	E Int (H)	Int	Int
BBL	E Int (L)	E Int	Int (L)	Int
BBO	Beg (H)	E Int	E Int (H)	E Int (H)

This figure shows the CELDT levels that correspond to the raw means listed in Table 4.15.

Table 4.24 Annual Writing CELDT levels for the Analysis 2 Cohort

Year/ Grade	2004 G-2	2005 G-3	2006 G-4	2007 G-5
EI	Int (L)	Int	Int (H)	Int
BB	E Int (L)	E Int	Int (L)	E Int (H)

This figure shows the CELDT levels that correspond to the raw means listed in Table 4.16.

The rate of progress was very slow for the writing component for all groups as well. While the SEI groups began at higher levels, they showed almost no advancement—staying at the Intermediate level for all four years. The BB groups showed more progress, but very gradual. The familiar order from highest to lowest—**EO/EI**, **BBS**, **BBL**, **BB(O)—**prevailed (although BBS and BBL were very close).

4.4 Research Question 1: Percentage of High Achievers

This section consists of tables that show the percentage of ELLs who scored at significant levels above their group means each year on the CELDT according to the language program they participated in. As such, these tables will help answer a main question of this study, "Are bilingual language programs better than Structured English immersion language programs in helping ELLs acquire English?"

These tables show annual CELDT proficiency levels by cohort group. When students take the CELDT, their raw score is converted to a proficiency level, which helps the school district to place them in an appropriate English-language class. Students receive different scores for each separate section of the CELDT (oral, reading, and writing). From there a composite score is generated, which is referred to as the "Overall" score. It is possible for a student to receive a low score on one section of the CELDT and a higher on another section. These tables are therefore useful in comparing the effectiveness of bilingual and Structured English immersion class instruction on English acquisition. This comparison can help answer a main question of this study, "Are bilingual programs better than Structured English

immersion programs in helping ELLs acquire English?" Below is a Key to understanding the CELDT level tables.

When students reach a CELDT Overall proficiency level of Early Advanced, they are at the beginning stages of English proficiency. This is therefore a significant step toward being placed in classes with students who are native-English speakers, or highly fluent English speakers, who are not in need of special language service programs, which is the goal of all transitional language assistance programs.

Students who attain a minimum CELDT Overall proficiency level of Early Advanced in conjunction with at least a CELDT Intermediate level in all of the other English measurements can be considered candidates for reclassification as Reclassified Fluent English Proficient (RFEP). Also to reclassify they need to meet two other criteria: (1) receive a grade equivalent of "C" or better in their English, or English Language Development (ELD) class, and (2) receive a score of "Proficient" or better on the state standardized California Standards Test (CST).

Finally, students who meet the CELDT reclassification criteria, and meet one of the other two criteria mentioned above, are placed in mainstream classes. At this point they receive a new classification as Preparing to Reclassify (PRP).

As these three benchmarks are all significant in a student's journey toward reclassification, they can be useful in helping to determine whether bilingual programs or Structured English immersion programs are more effective in helping students reach that ultimate goal.

4.4.1 Analysis 1 and Analysis 2 Cohorts: Percent Attaining at Least Early Advanced

Table 4.25 Analysis 1 Cohort: Students Who Scored CELDT Overall Early Advanced

EARLY ADV	2003 G-1	2004 G-2	2005 G-3	2006 G-4	2007 G-5	Total
EO	65	76	85	259	358	843
EO %	**7%**	**9%**	**10%**	**30%**	**41%**	
BBS	1	2	0	17	29	49
BBS %	**1%**	**2%**	**0%**	**20%**	**34%**	
BBL	1	2	5	18	33	59
BBL %	**1%**	**1%**	**3%**	**12%**	**21%**	
BBO	3	1	1	8	17	30
BBO %	**4%**	**1%**	**1%**	**10%**	**21%**	

This table identifies the annual percentage of students by cohort group who have scores at least Early Advanced on the Overall portion of the CELDT.

Table 4.26 Analysis 2 Cohort: Students Who Scored CELDT Overall Early Advanced

EARLY ADV	2003 G-1	2004 G-2	2005 G-3	2006 G-4	2007 G-5	Total
EI	171	284	275	470	292	1,492
EI%	**8%**	**17%**	**19%**	**39%**	**36%**	
BB	10	10	23	27	9	79
BB %	**2%**	**2%**	**7%**	**15%**	**13%**	

Table 4.28 shows the number and percentage of students who placed at least Early Advanced on the Overall section of the CELDT each year. These students have met at least the CELDT Overall requirement to be placed in mainstream classes, but may not have met other requirements.

The first year any cohort scored a minimum of Early Advanced was 2003, which would have been an Oral score only since Reading and Writing were not introduced on the CELDT until 2004. While **EO** in the Analysis 1 Group produced higher percentages of students that attained at least Early Advanced than the **EI** Analysis 2 Group for 2007, the Analysis 2 Group produced more for each year

leading up to 2007. The results for **BB(O)** were similar. The EO groups outperformed the BB groups each year. The **BBS** group performed closer to the two EO groups in 2006 and 2007 than the other BB groups, and **BBL** scored similarly to **BBO**. These results were similar to previous analyses.

4.4.2 Analysis 1 and Analysis 2 Cohorts: Percent Meeting CELDT RFEP Criteria

Table 4.27 Analysis 1 Cohort – Students Who Met CELDT RFEP Criteria

Proficient	2003 G-1	2004 G-2	2005 G-3	2006 G-4	2007 G-5	Total
EO	0	58	53	208	305	624
EO%	**0%**	**7%**	**6%**	**24%**	**35%**	
BBS	0	2	0	14	27	43
BBS %	**0%**	**2%**	**0%**	**16%**	**31%**	
BBL	0	0	4	12	31	47
BBL %	**0%**	**0%**	**3%**	**8%**	**20%**	
BBO	0	1	1	4	12	18
BBO %	**0%**	**1%**	**1%**	**5%**	**15%**	

This table indicates the percentage of students who annually attained at least Early Advance on the CELDT Overall, and at least Intermediate in each of the language skills.

Table 4.28 Analysis 2 Cohort – Students Who Met CELDT RFEP Criteria

Proficient	2003 G-1	2004 G-2	2005 G-3	2006 G-4	2007 G-5	Total
EI	0	234	212	412	247	1,105
EI %	**0%**	**14%**	**14%**	**34%**	**30%**	
BB	0	4	19	20	4	47
BB %	**0%**	**1%**	**6%**	**11%**	**6%**	

Table 4.28 indicates the number and percentage of students who achieved at least the CELDT Overall level of Early Advanced and at least a CELDT level of Intermediate for Listening/Speaking, Reading, and Writing, thereby meeting District X's requirement for being English Proficient.

All cohorts except **BBL** began to score in the English proficient range in 2004. The SEI groups scored higher than the BB groups for each year. **BBS** scored higher than **BBL** and **BB(O)**. **BB** scored higher than **BBO** until 2007. Again, the same relative pattern between groups emerges.

4.4.3 Analysis 1 and Analysis 2 Cohorts: Percent Classified PRP

Table 4.29 Analysis 1 Cohort – Students Who Reclassify as PRP

PRP	2003 G-1	2004 G-2	2005 G-3	2006 G-4	2007 G-5	**Total**
EO	0	0	2	17	134	153
EO %	0%	0%	0%	**2%**	**15%**	
BBS	0	0	0	1	11	12
BBS %	0%	0%	0%	**1%**	**13%**	
BBL	0	0	0	0	8	8
BBL %	0%	0%	0%	0%	**5%**	
BBO	0	0	0	2	13	15
BBO %	0%	0%	0%	**3%**	**16%**	

Table 4.29 indicates the percentage of students by language program that were placed in English-only classes because they met two of the three requirements to be reclassified as RFEP. They attained at least Early Advanced on the CELDT Overall (and at least Intermediate in each of the other language skills); or they received a "C" or its equivalent in their school's English language arts class; and/or they received a "Proficient" score on the CST.

Table 4.30 Analysis 2 Cohort – Students Who Reclassify as PRP

PRP	2003 G-1	2004 G-2	2005 G-3	2006 G-4	2007 G-5	Total
EI	0	0	15	60	155	230
EI %	0%	0%	**1%**	**5%**	**19%**	
BB	0	0	0	5	14	19
BB %	0%	0%	**0%**	**3%**	**21%**	

Table 4.30 indicates the number of students that have met two of the three following criteria for reclassification as RFEP: (1) an Overall CELDT level of Early Advanced and at least a CELDT level of Intermediate for Listening/Speaking, Reading, and Writing; and/or (2) the equivalent of a "C" grade in their English language arts class (or its equivalent: a "3" in their ELD class); and/or (3) a score of "Proficient" on the ELA portion of the CST.

As no students in this study reclassified RFEP during their six years in the district, these scores represent the highest performing students overall. They have met the CELDT requirements for reclassification and at least one of the other requirements for reclassification. As this level of proficiency is more stringent, no students meet these requirements until 2005, and then only one percent of students in **EI** (although two students in **EO**—less than one percent also met the requirements). Interestingly, **BB(O)** narrowly outperform **EI/EO** in 2007, and **BBS** is close to **EO** in 2007. Also of interest in these data is this is the only segment of this study that includes academic performance.

4.5 Research Question 1: Overall Conclusions from Cohort Means, Proficiency Levels, and High Achievers

Regarding cohort means comparisons, there are only five instances for all groups in both cohort analyses when a BB group didn't narrow the point gap with an EO group each year, and in each case the widening was extremely small. The **EO/EI** groups when compared to each other, and the **BB(0)** groups when compared to each other, were nearly identical for both analyses for the entire six-year study. Students who had fewer years in BB programs consistently out performed students with more years in BB programs on the CELDT English language acquisition measure. By 2007 the gap between EO means and BB means was substantially reduced for all CELDT English proficiency measures, especially for Oral proficiency and Overall proficiency.

Taken together, the CELDT level comparisons between the Analysis 1 Cohort and the Analysis 2 Cohort show a consistent pattern. They illustrate that initially the **EO/EI** cohorts for both analyses acquire English oral and literacy skills more quickly. The BB groups that catch up to them more quickly are generally the ones that have had fewer years of bilingual education. For the CELDT Overall section, both the Analysis 1 cohort and the Analysis 2 cohort have scores in the Intermediate range from 2006 onward. For Listening/Speaking, **EO** and all BB groups are Early Advanced by 2007. All groups in both cohort analyses are Intermediate in 2007 for Reading except for the **BB(0)** groups, which score Early Intermediate. This same pattern holds for Writing. With the exception of 2005 for Listening/Speaking, both analysis groups showed an annual increase in scores for each year. Of course by

looking at the proficiency levels instead of just the means, the group levels in the comparisons seem closer.

The results of the high achievers for each cohort followed similar patterns of achievement. For most sections SEI groups outperform BB groups for most of the study. Students with fewer years in BB programs advance more quickly than students with more years in BB programs. The advantage of EO groups dissipates on the CELDT measure over time, and in the case of PRP advancement (the only part of the study that includes academic achievement) the BB-only groups in Analysis 1 and Analysis 2 outperform their respective SEI groups.

It is clear from all three analyses that students with no English ability can learn English at a slightly higher rate by being taught in an Structured English immersion medium than if those students are in a Bilingual program for various lengths of time. It is also clear that students who are in bilingual programs for shorter periods of time learn English at a slightly higher rate initially than students who are in bilingual programs for longer periods of time. According to the CELDT Overall scores in all three analyses, the slight advantage that **EO/EI** students have initially begins to disappear in five years. In oral fluency, the BB groups in both cohorts outperformed SEI groups in measures that included some academic requirements by the end of the study. This was especially surprising in that the BB groups did not have the advantage of using data from students that had moved out to an SEI program.

4.6 Research Question 2: CELDT Oral Proficiency Related to CELDT Literacy Proficiency

In a departure from the previous section that compares overall English proficiency across program types, this section seeks to link oral skills to literacy skills, thereby answering the second research question of this study. This oral/literacy feature is also viewed across program types. This Chapter uses the tables from Chapter 4 to observe the relationship of oral skills to literacy performance. The abbreviation key and tables from Chapter 4 are repeated below.

Key to Figures:

L = one level below oral level (for example, if the oral level were Int, then the L would represent E Int)

L(2) = two levels below the oral level (for example, if the oral level were Int, then L(2) would represent Beg)

S = the same level as the oral level (the same general mid-range score for that level; for example if the range for an Intermediate level were 400-500, then the lower range would be 400-433, the mid-range would be 433-466, and the high range would be 467-500)

S(L) = the same level as the oral level, but in the lower scoring range of that level (for example, if the oral level were Int, then this would be in the lower scoring range of Int)

S(H) = the same level as the oral level, but in the higher scoring range of that level (for example, if the oral level were Int, then this would be in the higher range of Int)

L/Sp = Listening/Speaking as a combined score

The tables in this section show the relationship between CELDT oral proficiency and CELDT proficiency in literacy. If it is established that oral skills significantly precede literacy skills, then perhaps school curriculum designers will

attempt to design programs that enhance oral reception and production. These tables can help to answer the second question of this study, "What is the link between oral competency and the attainment of literacy skills?"

The tables below show the CELDT proficiency levels for each cohort group for Listening/Speaking (or Listening and Speaking as separate categories in 2007) on the left side of the table. Again, the levels are Beginning (Beg), Early Intermediate (E Int), Intermediate (Int), Early Advanced (E Adv), and Advanced (Adv). Under the columns for Reading and Writing are letters that explain whether for that year those levels were lower, the same, or higher (see key below) than the Listening/Speaking levels.

4.6.1 Analysis 1 Cohorts

Table 4.31 Analysis 1 Cohort – Relationship of Oral Skills to Literacy Skills for **EO**

Year/ Grade	Listen	Speak	L/Sp	Reading	Writing
2004 G-2			Int	L	S (L)
2005 G-3			Int	L	S
2006 G-4			E Adv	L (2)	L
2007 G-5	Int (H)	E Adv (H)		S(L) / L	S(L) / L

This table shows the relationship of CELDT oral proficiency to CELDT literacy proficiency for **EO** students in the Analysis 1 Cohort.

Table 4.32 Analysis 1 Cohort – Relationship of Oral Skills to Literacy Skills for **BBS**

Year/ Grade	Listen	Speak	L/Speak	Reading	Writing
2004 G-2			E Int (H)	L	S(L)
2005 G-3			E Int	L	S(H)
2006 G-4			E Adv	L (2)	L
2007 G-5	Int (H)	E Adv		S(L) / L	S(L) / L

This table shows the relationship of CELDT oral proficiency to CELDT literacy proficiency for **BBS** students in the Analysis 1 Cohort.

Table 4.33 Analysis 1 Cohort – Relationship of Oral Skills to Literacy Skills for **BBL**

Year/ Grade	Listen	Speak	L/Speak	Reading	Writing
2004 G-2			E Int	L	S (L)
2005 G-3			E Int	L	S
2006 G-4			Int	L	S (L)
2007 G-5	Int (H)	E Adv		S(L) / L	S(L) / L

This table shows the relationship of CELDT oral proficiency to CELDT literacy proficiency for **BBL** students in the Analysis 1 Cohort.

Table 4.34 Analysis 1 Cohort – Relationship of Oral Skills to Literacy Skills for **BBO**

Year/ Grade	Listen	Speak	L/Speak	Reading	Writing
2004 G-2			E Int	L	L
2005 G-3			E Int	L	S
2006 G-4			Int	L	L
2007 G-5	Int (H)	E Adv		L / L(2)	L / L(2)

This table shows the relationship of CELDT oral proficiency to CELDT literacy proficiency for **BBO** students in the Analysis 1 Cohort.

4.6.2 Analysis 2 Cohorts

Table 4.35 Analysis 2 Cohort – Oral to Academic for **EI**

Year/ Grade	Listen	Speak	L/Sp	Reading	Writing
2004 G-2			Int	L	S(L)
2005 G-3			Int	L	S
2006 G-4			E Adv (L)	L	L
2007 G-5	Int (H)	E Adv (L)		S(L)/L	S(L)/L

This table shows the relationship of CELDT oral proficiency to CELDT literacy proficiency for **EI** students in the Analysis 2 Cohort.

Table 4.36 Analysis 2 Cohort – Oral to Academic for **BB**

Year/ Grade	Listen	Speak	L/Sp	Reading	Writing
2004 G-2			E Int	L	S (L)
2005 G-3			E Int (H)	L	S (L)
2006 G-4			Int	L	S(L)
2007 G-5	Int	Int (H)		L/L	L/L

This table shows the relationship of CELDT oral proficiency to CELDT literacy proficiency for **BB** students in the Analysis 2 Cohort.

4.6.3 Conclusions Regarding the Relationship between Oral Language and Literacy

The **EO/EI** groups were identical with the exception of the Analysis 1 cohort being two levels below the oral level in Reading in 2006. There is greater difference between the BB groups. While the oral scores are nearly identical, the **BBO** group from the Analysis 1 cohort scores slightly higher than the **BB** cohort in the Analysis 2 cohort in 2007, when the Analysis 1 cohort is in the higher range of the Intermediate level for Listening, and one level higher (E Adv vs. Int(H)) for Speaking. On the other hand, the **BB** group from the Analysis cohort scored higher in the Writing domain most years (especially in 2007 when the Analysis 1 cohort was 2 levels lower in Writing than in the Speaking domain).

In both analyses for all years no group attained a higher level in Reading or Writing than was attained in Listening/Speaking or in Listening and Speaking separately in 2007. There were times when a group attained the same CELDT level as the oral domain, but when that occurred, the oral score was in the higher scoring range of that level (with one exception: the **BBS** score for Writing in 2005, which was at the same level as the oral level, but in the higher range). On only three occasions was a literacy score two whole levels lower than an oral score (**EO** in the Analysis 1 cohort for Reading in 2006, and **BBO** in the Analysis 1 cohort for Reading and Writing compared to Speaking in 2007). With the exception of 2007, if a group scored at the same level academically as their corresponding oral level (or at the same level but slightly higher), it was in the Writing domain. The table below presents the preceding results numerically.

Table 4.37 The Number of Times the CELDT Literacy Level was Lower than, equal to, or higher than the CELDT Oral Level from 2004-2007

Group	Oral > Literacy		Oral = Literacy		Oral < Literacy	
	Reading	Writing	Reading	Writing	Reading	Writing
Anaylsis 1						
EO	5	4	0	1	0	0
BBS	5	4	0	0	0	1
BBL	5	4	0	1	0	0
BBO	5	4	0	1	0	0
Analysis 2						
EI	5	4	0	1	0	0
BB	5	5	0	0	0	0
TOTAL	30/30	25/30	0/30	4/30	0/30	1/30
%	100%	83%	0%	13%	0%	3%

This figure shows the relationship of oral proficiency all years combined. The first column (Number < Oral) shows the number of times from 2004-2007 that a

literacy score (Reading and Writing) was at least one level lower than the oral. In this case **EO** Reading was lower than Oral 5 times (over the four year period), and **EI** Writing was lower than Oral 4 times (over the four year period). Five possibilities are included during this time frame because in 2007 Listening and Speaking are separate scores, making it possible to compare the Listening level and the Speaking level to the literacy levels.

The trends are clear. With few exceptions, students scored lower on literacy skills than on oral skills for each assessment period. Every year each group scored lower on Reading than on Oral assessments. All students also scored lower on Writing 80% the time. There were four groups (**EO**, **BBL**, and **BBO** in Analysis 1, and **EI** in Analysis 2) that scored the same in Writing as on Oral on one occasion each (13%). Only one group, as indicated above (**BBS** in the Analysis 1 Group), scored higher in Writing than Oral in one year; however, it was within the same CELDT level as the Oral, but in a higher range. These results held true whether students were in EO classes or BB classes.

Chapter 5: Discussion

5.1 Studies that Compared Literacy Progress Across Program Types

As mentioned in Chapter 1 and Chapter 2 most studies that have compared bilingual programs to Structured English immersion programs have looked at program effects on academic performance. Researchers who have reviewed large numbers of studies have reported varying results. Some of these researchers (Willig, 1985; Greene, 1998; Thomas & Collier, 2002; Slavin & Cheung, 2003; Rolstad et al., 2005a; Rolstad et al., 2005b; Goldberg, 2008) contend that bilingual programs, especially dual language programs, are superior to Structured English immersion programs. Other researchers (Baker & de Kanter, 1981; Rossell & Baker, 1996) declare that bilingual programs are equal to SEI programs, or that there is not much difference between them. Still other researchers (Rossell & Kuder, 2005) assert that Structured English immersion (or mainly Structured English immersion) programs are more beneficial, or more beneficial most of the time, in helping ELLs learn English. More directly related to this study, the Parrish et al. (2006) five-year study commissioned by the state of California in 2005 used the vast student database of District X to measure the effects of Proposition 227 on ELL student achievement. Their results showed no difference in student performance in English second lnaguage acquisition based on the type of language program they were in. However, it should be noted that they did not control for initial student proficiency.

The "Post Proposition Studies" in Chapter 2 (see section 2.3) also evidence mixed academic performance across program type. Recall that these studies looked at test results in states that had dismantled bilingual education. Gordon and Hoxby

(2002) found little difference in California achievement. Rossell (2005) showed a slight improvement in English reading and math. Lindholm-Leary and Genesee (2010) noticed no improvement in test scores. Rumberger and Tran (2010) stated that bilingual programs helped lower the achievement gap. Losen (2010) cites poorer performance in Structured English immersion states.

One of the major arguments for or against these programs is centered in the way the data is analyzed by the researchers (see Rolstad et al., 2005a; Rolstad et al., 2005b; and Rossell and Kunder, 2005 for detailed discussion on these types of data analyses). Nevertheless, one is often struck that at times the results favor a bilingual group or an Structured English immersion (or English-mostly) group. Also, when a study favors one group over another, it is often by a small margin.

Only two elements of this study relate to academic achievement. They are the California state classifications: Preparing to Reclassify (PRP) and Reclassified Fluent English Proficient (RFEP). In Chapter 3 and Chapter 4 criteria were introduced for students to reclassify as RFEP. Students who reclassify as RFEP must meet specific CELDT, school, and CST benchmarks. The CELDT criteria for reclassification is attaining a CELDT Overall level of at least Early Advanced while at the same time attaining no lower level than Intermediate for all other CELDT sections. Students also have to receive a passing grade in their English (or ELD) class equivalent to a "C" or better, and they have to attain a score of "Proficient" or above on the ELA portion of the CST.

Before reaching the RFEP benchmark, students can achieve the label Preparing to Reclassify (PRP). This step requires meeting any two of the RFEP

criteria. Therefore, when students receive a PRP label, some academic proficiency has been established, either in the school classroom or on the CST (or both in the case of a RFEP classification).

Implied in District X's goal for yearly CELDT advancement is the expectation of reclassification within four to six years of starting a program. If the students in the two cohorts had advanced from one CELDT level to the next each year, all of them would have reclassified by 2006 or 2007. Instead, no students from either the Analysis 1 cohort or the Analysis 2 cohort reclassified by 2007. In fact, the highest achieving cohort in either analysis group remained at the Intermediate level for four years, from 2004 – 2007. The following table shows the District X reclassification rate for the years 2002-06.

Table 5.1 District X Elementary Students Who Reclassified as RFEP from 2002-03 to 2005-06

	Study Samples	
YEAR/ Grade	Analysis 1 / Analysis 2	District X
2002-03 G1	0	2.4%
2003-04 G2	0	4.5%
2004-05 G3	0	6.5%
2005-06 G4	0	8.6%

This table compares the percentage of students who advanced to Fluent English Proficient over a four-year period (adapted from Salazar, 2007)

The percentage of RFEP students increased over time in the district. These numbers for the district reflect the total number of students who reclassify each year from the total elementary ELL population, not from a particular cohort. Salazar mentions that these achievements were basically the same whether the students

were enrolled in bilingual programs or English Immersion programs. (Salazar, 2007). It should be noted that in the comparison in Table 5.1 the District X students may have started their particular program at any CELDT proficiency level.

While no students attained the RFEP reclassification in this study, many were eventually classified as Preparing to Reclassify (PRP). Table 4.29 shows that students in the **BBO** program of the Analysis 1 cohort had a higher percentage of students reclassify as PRP than any other bilingual or SEI group in 2006 and 2007. Likewise, Table 4.30 shows a higher percentage of **BB** reclassified as PRP than **EI** in 2007. Therefore, both cohorts report slightly higher academic results at or near the end of the study for the bilingual groups.

5.2 Studies That Used the CELDT to Measure Academic Proficiency

Also mentioned in Chapter 2, Katz, Low, Stack, & Tsang, (2004) were interested in linking student language proficiency on the CELDT to academic proficiency. Using data from the San Francisco Unified School District from 2000 to 2002, they matched student CELDT scores to their proficiency levels on the Stanford-9 Test (SAT-9), California's annual standardized state test for all students from 1997 to 2002. Their study included Asian- and Spanish-speaking ELLs, and also included RFEP students. When Katz et al. compared CELDT Overall scores to SAT-9 math scores, they noticed a moderate, but diminishing correlation from grades 2 through 5. The correlations were much weaker for middle and high school students. They observed a stronger, and non-linear, correlation between CELDT Overall scores and SAT-9 reading—again, much stronger for elementary grades than for middle and high school.

In an attempt to identify a stronger correlation between the CELDT and SAT-9 reading scores Katz et al. used the scores of high achievers. The first group consisted of students who met CELDT RFEP requirements. Elementary students in this category did much better compared to the national norm group than middle school or high school ELLs. Katz et al. next looked two sets of even higher scorers: students who had CELDT Overall scores of at least Early Advanced and at least Early Advanced on all other CELDT sections, and students who scored Advanced Overall and at least Early Advanced in on all other CELDT sections. No data was available for elementary, but middle and high school students performed quite poorly on the SAT-9 in spite of strong CELDT scores.

Katz el al. then compared CELDT Reading-only scores with SAT-9 Reading scores. Two sets of scores were used: those who scored Early Advanced in CELDT Reading, and those who scored Advanced in CELDT Reading. Elementary students with these CELDT scores performed much better than middle or high school students, again verifying that there is a greater correlation between CELDT scores and academic performance for the early grades than later grades.

Hakuta (2011) used CELDT data from Sanger Unified School District in California to show a correlation between CELDT proficiency and academic proficiency on the CST. His student database also included RFEP students with ELLs. Plotting graph data from 2003 – 2009 for third, fifth, and seventh grade students, he illustrates that once students attain CEDLT proficiency, they also attain academic proficiency in CST Reading and Math.

Again, the principal focus of the Analysis 1 and Analysis 2 study was not to measure academic proficiency. However, since other CELDT comparison studies find a moderate to strong correlation for the elementary grades to standardized scores in California, and both language programs in this study produced a similar percentage of PRP students (and neither produced RFEP students), it is appropriate to state that this study shows a similar pattern to Katz et al. (2004), and Hakuta (2011). In other words, when students perform with high CELDT Overall proficiency in the early grades, they are more likely to perform well academically.

5.3 Studies that Use CELDT Data to Only Measure English Language Proficiency

The students in this study were competitive with the district's students who met District X 's English proficiency standard. While the district average was 37.2% (for elementary schools) in 2006, the Analysis 1 Group was 35% for **EO**, 31% for **BBS**, and 30% for **EI** Analysis 2 (the state average was 33.3%) in 2007. Those with longer time in the BB programs scored at 15%, 16%, and 20% in 2007. These numbers stand up very well for cohorts who started out at rock bottom English proficiency compared to students who could have started District X at any level.

Jepsen's (2009) study mirrors this study more than any other. Like this study, Jepsen (a) compared bilingual to Structured English immersion programs with a focus on English proficiency as measured on CELDT scores, not academic proficiency as measured by CST scores; (b) used data from elementary grades from a similar time period, (c) evaluated only Spanish-speaking students from primarily

low SES backgrounds; controlled for potential differences in the nature of bilingual programs between schools; and (e) comes to similar conclusions.

Unlike this study, Jepsen (a) has a larger corpus; (b) conducts a cross sectional study, not a longitudinal study; (c) included students who receive special education services; (c) included students in dual language programs and RFEP students; and (d) had access to some non-student (parental background, teacher qualifications, etc.) information unavailable to this study.

Jepsen's (2009) primary statistical interest in comparing bilingual programs to Structured English immersion programs was in mitigating the effects of non-randomness in students being placed in bilingual programs. He employs three statistical tools to accomplish this and to compare bilingual programs with Structured English immersion programs. His main findings are, "In sum, the results are generally consistent across all three methods (OLS, propensity score, and IV). Compared to other EL programs, bilingual education has a sizable, negative association with English listening/speaking proficiency for EL students in first and second grade. For EL students in grades three through five, bilingual education has a small and often insignificant association with English proficiency." (Jepsen, 2009, p. 21) The effects of bilingual programs are uneven on English reading, but overall there is not a strong association with reading scores. English writing proficiency is moderately negatively associated with bilingual education but less strongly over time.

Compared to Jepsen's (2009) study, the greatest gap between the transitional bilingual oral scores and the **EO** oral scores also occurs within the first two years in

this study. However, in this study, the gap doesn't close until the fifth year of the study, when **BBO** in the Analysis 1 group attains the same CELDT level as **EO,** and **BB** in the Analysis 2 group actually surpasses **EI**. Jepsen (2009) reports a lack of strong correlation between the bilingual program and reading. The Analysis 1 group in this study also shows weak correlation between CELDT Reading scores and program type. The BB groups never fully catch up with EO groups for reading or writing scores, but they are quite close by 2007.

Jepsen's (2009) use of cross-sectional data reveals that students are in bilingual programs in greater numbers in earlier years. This is no doubt due to the nature of transitional bilingual programs. His analysis indicates however that the effects of bilingual education are not driven by the attrition of ELL students.

It should be noted that because Jepsen's (2009) comparison is cross-sectional, we cannot be sure how many years of bilingual education the students had experienced in each group and therefore, unlike this study, we are not able to measure changes over time. His study also doesn't control for starting levels of English proficiency when students begin school. Nevertheless, the results of Jepsen's (2009) study are quite similar to this study in showing that in the later years of elementary schooling students in bilingual programs and Structured English immersion programs perform similarly.

5.4 Research Question 2: Studies that Focus on Oral Skills

Chapter 2 introduces several studies have attempted to use oral proficiency to predict academic success. Katz et al. (2004) reviewed several studies that compared various oral language instruments, such as the BINL, LAB, LAS, and BSM. Researchers found very low correlations between these instruments and their use in placing ELL students in program levels. Furthermore, these oral instruments were not able to predict success on standardized achievement tests. More ironically, some students who scored low on the oral measures scored quite high on the academic tests, and visa versa.

Later, Saunders & O'Brian (2006) pointed out only one study (Hakuta et al., 2000) since the mid-1980s had attempted to link oral proficiency with academic proficiency. Using their measurements of academic oral proficiency Hakuta et al. (2000) established that it took from 3 to 5 years (or possibly longer because of their data constraints) for students to attain academic oral fluency. In the Analysis 1 and Analysis 2 studies students begin to reach aural/oral proficiency when they score Early Advanced on the CELDT. The Analysis 1 and Analysis 2 groups showed similar progress to the Hakuta study. The group oral means for this study reached this level for **EO** and **EI** after 4 years of schooling (in 2005). By 2007 everyone in the Analysis 1 Group attained Early Advanced for Listening and Speaking. In the Analysis 2 Group, **EI** scored in the high Intermediate range (close to proficiency—this was after being Early Advanced the previous year) for both Listening and Speaking, while **BB** scored Early Advanced for Listening and Speaking. So after five years, most of both cohorts were at the beginning areas of aural/oral fluency.

In Hakuta et al.'s (2000) study, academic proficiency lagged behind oral proficiency. While oral proficiency took from between 3 – 5 years, academic proficiency took from between 4 – 7 years. Even though the results of this study are not directly comparable to Hakuta's study because this study didn't look at student state academic scores, it's interesting to note that oral skill levels were consistently higher than literacy skills regardless of program type. This is also manifested in Table 5.2 below.

Table 5.2 CELDT levels for DISCTRICT X, L.A. County, California

District X	**2004**	**2005**	**2006**	**2007**
Lis/Speak	E Adv (L)	Int	E Adv	E Adv
Reading	E Int	E Int	Int (L)	Int
Writing	Int	Int	Int	Int
LA COUNTY				
Lis/Speak	E Adv (L)	Int	E Adv	E Adv
Reading	E Int	E Int	Int (L)	Int
Writing	Int	Int	Int	Int
CALIFORNIA				
Lis/Speak	Int	Int	E Adv (L)	E Adv (L)
Reading	E Int	E Int	Int (L)	Int
Writing	E Int (H)	Int	Int	Int

This table shows local, regional, and state CELDT levels. It captures the relationship between oral English skills and English skills in literacy. (H) and (L) indicates the (Higher) or (Lower) part of the level. E Adv = Early Advanced; Int = Intermediate; and E Int = Early Intermediate.

The same trends occur across the state. In Table 5.2 the Reading and Writing levels are nearly all at least one CELDT level below the oral levels. The exceptions are all in 2005, when the Writing levels for all three groups are the same as the oral

levels. As in the Analysis 1 and Analysis 2 studies, there are no years when the literacy levels are higher than the oral levels.

Genesee et al. (2006) synthesized several studies that recorded oral results from ELLs in various language programs. To make the studies comparable to each other they devised a 5-point proficiency scale with "1" as Beginning and "5" as near-native speaker proficiency. As this is comparable to the CELDT divisions, they may be compared to this study. Only one study, Hakuta et al. (2000), was somewhat comparable to the Analysis 1 and Analysis 2 studies for **EO** and **EI**. Hakuta's (2000) cross-sectional study included 1,872 students in a Structured English immersion program. The figure below compares Hakuta's student levels to the Analysis 1 and Analysis 2 EO levels.

Table 5.3 Hakuta et al. Oral levels compared to the Analysis 1 and Analysis 2 Group CELDT Oral Levels

	Grade Level					
Group	K	1	2	3	4	5
Hakuta		Beg	Int	E Adv (L)	E Adv	E Adv (H)
Analysis 1	Beg	E Int	Int	Int	E Adv(L)	E Adv
Analysis 2	Beg	E Int (H)	E Int (H)	E Int (H)	E Int (H)	E Adv

Adapted from Genesee et al. 2006, pg. 24

This table compares the **EO** and **EI** longitudinal Analysis 1 and Analysis 2 comparisons to Hakuta et al.'s (2000) EO cross-sectional study. All students entered Kindergarten at the Beginning level.

The sample from Hakuta's group differs somewhat from those of the Analysis 1 and Analysis 2 groups. It is not only a cross-sectional study, but his students were

sampled from one of the most successful school districts in California, which had a comparatively low percentage of students considered to be low SES, and that district's annual reclassification rate is nearly four times that of the state. His study is also trying to profile academic oral English, not oral proficiency in general. However, like the Analysis 1 and Analysis 2 groups, his students began Kindergarten at the Beginning level (but we don't know how low the Beginning level was).

Hakuta's group progresses more slowly at first, then moves ahead very gradually starting in grade 2. The Analysis 1 cohort shows faster initial progress, but starting in Grade 2 slows down. Both are at Early Advanced by Grade 4 and remain at that level (Hakuta's group scoring higher within the same level). The parity of Hakuta et al.'s group to this study may reflect similar progress as the CELDT Early Advanced level indicates the state's evaluation that students at this level can succeed in mainstream classrooms.

Table 5.4 Oral Proficiency for Dual Immersion and Two Way Immersion programs compared to the Analysis 1 and Analysis 2 cohorts.

	K	1	2	3	4	5
DBE	Beg		Int			
TWI				E Adv		E Adv
TWI 50/50		E Int	Int	Int	Int	E Adv
TWI 90/10	E Int (H)	Int	Int	Int	E Adv	E Adv (H)
TWI 90/10	E Int	Int (L)	Int	Int (H)	Int (H)	E Adv
Analysis 1	Beg (L)	Beg	E Int	E Int	Int	E Adv (L)
Analysis 2	Beg (L)	Beg	E Int	E Int (H)	Int	E Adv (L)

Adapted from Genesee et al. 2006, pg. 24
This figure compares Listening/Speaking proficiencies between dual language programs and the **BB(O)** groups for this study.

No longitudinal oral data were available for early or late exit bilingual programs in Genesee et al.'s review. However, for rough comparisons, data from dual language programs can be compared to the bilingual programs from this study. The bilingual programs for the Analysis 1 and Analysis 2 groups were exit programs. However, all of the students in the Analysis 1 Group were in the bilingual program for the full 6 years of the study. While the Analysis 2 Group includes students that were in the bilingual program for various lengths of time, their scores are nearly identical to the Analysis 1 Group. Even though there are no data for the amount of time each day that content classes were taught in Spanish, a reasonable assumption is that in the earlier grades only Spanish was used to teach content course material and in the later grades some of the content courses were taught in Spanish.

In the first study in Table 5.4, Medina and Escamilla (1992) conducted a longitudinal study of students the DBE (Developmental Bilingual Education) program. These students were taught a minimum of 40% of the day in their primary

language. The first TWI (Two-Way Immersion) data set in Table 5.4 (Howard et al., 2003) was a 3-year longitudinal study of 11 different configurations of TWI programs (90/10; 50/50; or a modification of either). The next data set is from a 50/50 TWI quasi-longitudinal study (Thomas & Collier, 2002). The final two TWI 90/10 studies are cross-sectional from Lindholm-Leary (2001). Most dual language programs try to have a balanced L1 and L2 instructional division by Grades 4 or 5.

In Table 5.8.3, dual language programs reach parity with each other by Grade 2. The Analysis 1 and Analysis 2 groups are nearly the same as the dual language programs by grade 4, and are at the same relative levels by Grade 5 (albeit in the lower range of those levels). If the Analysis 1 and Analysis 2 groups "enjoyed an advantage" of more English being used in the classroom, it didn't manifest itself in a more rapid rate of English acquisition.

More comparable to this study, MacSwan & Pray (2005a) conducted a study in Arizona among Hispanic students in bilingual exit programs. They describe the bilingual programs as being "well-designed", with 65% of the participants as low SES, and a high percentage of the teachers in the program being highly trained in bilingual education techniques. They report that students who started Kindergarten at a level of "1" on the Bilingual Syntax Measure (BMS) took an average of 3.77 (with a standard deviation of 1.21) years to attain fluency, the range being 1 to 6 years. Oral fluency rates occur in a similar time course for the Analysis 1 and Analysis 2 cohorts.

5.5 Research Questions 1 and 2: Overall Conclusions

5.5.1 Research Question 1

As discussed in the Literature Review, studies comparing bilingual programs to Structured English immersion programs have produced mixed results for a number of reasons. In studies where one program may favor another, the margin of superiority is often not very great. This study, with its limitations, is no exception. While the findings tend to favor structured English programs initially, students in the bilingual programs were not far behind from year to year, and were very close to each other by the end of the study. These findings were consistent across the different types of comparisons between cohort groups.

Looking at the *means from the CELDT scores* was the first level of analysis in comparing programs. Comparing the different group means made it possible to make an interpretation of language program effectiveness. It quickly became clear SEI programs were receiving higher scores than their BB counterparts in both cohorts. It was also equally clear that the BB groups narrowed the difference in mean scores from year to year and in some cases caught up and surpassed SEI groups.

However, by using *CELDT proficiency levels* to compare the means, the disparity in performance seemed lessened. The SEI groups still led the way in scoring, but the BB groups appeared more competitive. Nevertheless, the BB groups trailed in literacy skills more than in oral skills throughout the program evaluation. This was somewhat surprising as most of the literature indicates that the one advantage that bilingual education consistently has over SEI is that literary skills

learned in the first language puts younger learners at an advantage in their second language literary skills (Thomas & Collier, 2002; Slavin & Chung, 2003; Roslad et al., 2005a,b; Genesee et al., 2006; Gandara & Hopkins, 2010; Hakuta, 2011).

One unique dimension of the Analysis 1 study was being able to look at the effects of *short-term vs. long-term bilingual treatment*. The length of time that students were in a bilingual program was clearly connected to how well the students performed. Those with 1-3 years in the program consistently performed better than those who were in the program for 4-6 years. Nevertheless, their Overall CELDT levels were the same as SEI CELDT levels from 2006.

Another component of the Analysis 1 Cohort was the inclusion of data from students who transferred to the mainstream program. Both SEI students and BB students who transferred to mainstream classes were exposed to a greater number of fluent English models. Even compared with students who had these added advantages, the BB-only students continued to close the gap with EO students each year and match their CELDT Overall and Oral levels after four years.

The Analysis 2 Cohort was designed to see how well each temporary cohort, **EI** and **BB**, would do *without the added possible benefit of a mainstream environment* that was available to the Analysis 1 group. The SEI groups and BB groups in both cohorts performed very closely to each other across programs independent of the mainstream influence. Seventy-five percent of the time the BB groups narrowed the scale-point gap by far more than one-half with their respective EI group by 2007. Like their counterparts, the BB group in the Analysis 2 cohort reduced the means scale point gap with the EI cohort by nearly the same number each year.

Looking at the CELDT scores to determine the percentage of students who, by the district's standard, *attained English Proficiency* was also useful. The same trends for EOs vs. BBs held for that measure. Those with little or no BB programming scored better each year than those with only BB programming. However, when it was time to mainstream students the BB groups outperformed their EO counterparts in both studies. By this important measure the BB groups do as well as or better than the EO groups by 2007.

5.5.2 Research Question 2

Many researchers have highlighted the need to deepen our understanding of the connection between oral proficiency and literacy skills (Hakuta, 2000; Peregoy & Boyle, 2001; Katz et al, 2004; Roberts & Neal, 2004; August & Shanahan, 2006; Saunders & O'Brien, 2006). Students in this study consistently did not perform above their aural/oral levels on literacy measures. The level of CELDT Reading and Writing scores were typically one level below Listening/Speaking proficiencies. This implies a relationship between aural/oral skills and literacy development. When literacy skills were closer to aural/oral skills, it was always in the Writing domain. This may be due to students being able to draw upon and utilize a range of competencies they are more confident about, and being able to prewrite and draft before producing a finished product. When students are exposed to reading passages, if their vocabulary, grammatical, and social communicative competency does not match the text level, the inner resources they have to draw on are more limited. Since this study was concerned only with language proficiency and not

academic proficiency, these connections were especially interesting. Other studies that attempt to correlate oral skills to academic skills still need to more explicitly define oral academic competency in order to make predictions regarding textual competency in academic areas.

5.6 Contributions of This Study

5.6.1 Research Question 1

This study was able to contribute to the research on bilingual and structured English immersion programs that serve English language learners. It was able to include controls for many variables that had not been controlled for in earlier studies, including: (1) the number of subjects; (2) the homogeneity of the subjects (including age, home language, ethnicity, geographic location, SES status, and types of schools attended); (3) the randomness of program inclusion; (4) the prior effects of language level and schooling; (5) the length of the study; (6) the use of standardized tests for all subjects; and (7) the use of the same test for all subjects administered at the same time of year. As such, the results are considered reliable within the limitations of the study (discussed below).

This research adds to other studies that students can learn a second language effectively while participating in bilingual and/or SEI programs. Much of the literature reviewed in Chapter 2 indicated various results when bilingual and Structured English immersion program types were compared to each other. Some researchers found bilingual programs to be superior (Willig, 1985; Greene, 1998; Thomas & Collier, 2002; Slavin & Cheung, 2003; Rolstad et al., 2005a; Rolstad et al.,

2005b; Goldberg, 2008), other researchers found the programs to be roughly equal to each other (Baker & de Kanter, 1981; Rossell & Baker, 1996), and a third group maintains that structured English immersion programs are superior (Rossell & Kuder, 2005). These studies often show superior results by a small margin, or that advantages of one program over another dissipate over time (Thomas & Collier, 2002). This study, while not focusing on the academic performance of ELLs, also shows both programs producing similar results over time. Still, longer-term effects on these students cannot be predicted, but Thomas and Collier's (2002) work would suggest that the students in the bilingual programs would ultimately fair better.

This study joins others in showing unequivocally that bilingual programs are not harmful to students, and that students in bilingual programs develop English skills at rates comparable to or better than students in structured English programs over time. Even researchers who favor English only instruction over bilingual instruction maintain that bilingual programs do not prevent students from learning English (Krashen, 2005, Rossell & Kuder, 2005, Hakuta, 2011). It therefore supports the position of those who are against dismantling bilingual programs on these bases.

No students in either program of this study were reclassified from ELL to RFEP during the entire six-year period of data collection. This is clear evidence that students need more than a year or two of English language instruction, no matter what program they are in, before they are ready to be mainstreamed. This study supports Cummins' (2000a) and Hakuta et al.'s (2000) position that students may need between 7 to 10 years of schooling before they are able to perform at a level comparable to their mono-lingual peers. Accordingly, Federal, state, and district

benchmarks that require students to advance at this rate are, in the words of Hakuta et al. (2000), "wildly unrealistic." While these goals are within the reach of some students, most students need more time.

This was also the only longitudinal study that compared the development of English language proficiency (and not academic performance) across program types, using both oral and literacy measures. The results were very consistent across all measures, and across time, indicating the reliability of the instrument and program influence. The only study similar to it was Jepson (2009), which was a cross sectional study, which included some variables not included in this study. Nevertheless, his conclusions were quite similar to this study's conclusions.

5.6.2 Research Question 2

This phase of the study also contributed to the body of research regarding oral studies. It demonstrated that student oral development precedes literacy development. It did so without "comparing apples and oranges." In other words, oral language proficiency was compared with literacy language proficiency. The CELDT is an indicator of how native-like a students language is orally, and in reading and writing. It is therefore a multidimensional measure of English language proficiency, unlike many instruments that use reading to identify English language growth. It was therefore unnecessary to be concerned with the definition of "academic oral language" when linking oral skills with literacy skills. As Katz et al. (2004) showed, oral language fluency is not necessarily a good predictor of student performance on

academic tests. The reading and writing portions of the CELDT reinforce the oral parts.

Other researchers (Cummins, 2000a, and Hakuta et al., 2000) have shown that oral fluency precedes academic fluency, and this study shows that literacy skills that indicate language fluency follow oral skills of the same measure.

The implication may be that greater emphasis should be placed on oral development, both communicative and academic, to ensure more rapid development in literacy domains. More research needs to be conducted to show the link between the type of oral skills needed for students to perform at high academic levels.

5.7 Limitations of This Study and Directions for Future Research

This study was limited in several ways. It would be beneficial for future studies of this kind to control more fully for the amount of English and the amount of primary language used in the classroom, and the amount of English and the primary language used outside of the classroom in defined contexts.

The data in this study was also not able to account directly for "best practices" in the classroom. A combination of direct classroom observations and student/teacher interviews would have enhanced this project. Katz et al. (2004) were able to illustrate the importance of complimenting quantitative studies with qualitative studies. One example was the type of and amount of interactions that reclassified students have in mainstreamed classrooms. That important variable

(and many others) cannot be captured in only looking at test scores. This would be a welcome addition to future research of this type.

It was also difficult to control for transience. Students in District X often move between schools (either within or outside of the local mini district), making monitoring for consistency in quality of education a great challenge. While this study was able to follow the test scores of students who remained in the district, it was not able to directly observe the type of teacher quality or school quality once students moved to other schools.

Another limitation of this comparison was the inability to monitor which classrooms may have had more than one proficiency level in the same class—this has been considered as one possible reason for the slow progress of ELL students in District X (Salazar, 2007).

Program compliance was also not possible to control for in the database for this project. While District X provided teachers in both programs with a pedagogical road map for instruction, the extent to how teachers as a whole adhered to program requirement could not be determined.

The current analyses only compared students who began at Kindergarten at zero English proficiency. Would the results have been different if students who started at different grade levels of no English proficiency have been different? At the district level, more students tend to reclassify in middle school than in elementary or high school when the entry-level proficiency is not considered. Alternatively, what would the results have been if the study had controlled for higher levels of proficiency upon enrollment at different grade levels?

Also, this study did not have access to state standardized test (CST) scores that monolingual English students take. This type of comparison would yield information of interest to school administrators and policy makers.

The cross sectional data currently available through District X (Salazar, 2007) indicates that students perform comparably regardless of language program. More sophisticated types of linguistic feedback for the oral evaluation (such as the BSM) than is provided by the CELDT could be developed to link up a more sophisticated evaluation of the reading and writing aspects of the CELDT. Unlike the annual state exams, the CELDT does not provide feedback on sub-skill segments of the test (for example, the CELDT provides a reading score, but does not provide separate scores for the different parts of the reading test).

Currently, District X reports having very few students who are in no language program once they are identified as LEP. It would be interesting to compare the effects of "submersion" (students just being placed in regular mainstream classes without the benefit of special sheltered instruction) to students in English Immersion or a bilingual program.

At the time of this study, District X did not have many students in dual language immersion programs, but the number of programs is growing slowly (currently available in Spanish, Korean, and Cantonese). This study would benefit by including results from the dual language programs, as much of the literature that favors bilingual programs indicates that dual language programs are better than transitional bilingual programs (see Chapter 2 under Meta-Analyses).

5.8 Implications of This Research for Policy

Independent of the program that ELLs are in, they do not perform well by federal, state, or district standards. In this study there were no students who were reclassified after six years of being in their respective programs. Students who entered District X at higher levels than the students in this study also took much longer to become proficient in English or to reclassify by the same standards. While this level of attainment may be in reach for a small number of students, it is clearly unrealistic to expect the majority of students to advance within the time frame these institutions have set for them. In this light, policy makers and educators need to adjust their expectations for ELL progress.

Furthermore, policy makers, administrators, and parents need not worry about placing students in bilingual programs. The initial early lead the Structured English immersion programs provide is both slight and temporary. This study and others cited above should slow down the momentum of political movements that seek to minimize or do away with bilingual programs altogether. Proponents of these programs are clearly out of touch with current research and are seemingly unaware of the benefits of developing skills in two or more languages, which can facilitate improvement in their social, academic, and financial future, not to mention mental health (Gonzalez, 2000; Trejo, 2003; Bleakley & Chin, 2004; Bialystok, 2008, 2010; Bialystok & Craik, 2010).

For students whose first language is not English, but who live in a community where the school district is not able to support the students' learning with a bilingual program in the students' first language, the focus will need to be on the

best allocation of available resources. District X is a vast district whose students speak over 80 different languages, but currently there are only bilingual programs in Korean/English, Chinese/English, and Spanish/English. August and Hakuta (1997) and Genesee (1999, 2006) have suggested that there is no one best model that will serve all students, and emphasize the importance of designing services for English learners that consider the community context, the needs of students who will be served, and the resources that are available for implementing the program. A manual issued by the New York City Board of Education (2010, p. 20) on the progress of English learners in New York City Schools indicated that consistency of programmatic approach appeared to be a more important determinant of progress than the specific educational philosophy and methods of the bilingual/ESL programs. Nevertheless, Hakuta, Butler, and Witt (2000) (see also Hakuta, 2011) call for research that can more clearly identify salient features of program types that may be more advantageous when resources are present. In this vein, Tong et al. (2008) and colleagues demonstrated that intervention techniques can facilitate language growth independent of program type (English immersion or bilingual).

This is an invitation for policy makers and educators to get in touch with and stay abreast of current research. While this study strongly inidcates there is no reason to cut bilingual programs, there are reasons to keep them. Thomas and Collier (2002) found that the single strongest indicator of long-term student academic gains was the length of time that students received instruction in their native language. They declared that that variable overshadowed even socioeconomic status as a predictor of success.

Chapter 6: Conclusions

Introduction

With the large influx of foreign students pouring into the country over the past several years, together with the large number of native born students whose first language is not English, the need is as great today as ever in finding methods to teach these students English. This population faces the twin challenges of needing to learn English and also learn the subject matter in this new language. Schools want to know the best way to accomplish this. What is the best approach to teaching ELL students English and what is the best approach to helping them gain access to the core curriculum in U.S. schools?

While many researchers over the years have sought to answer these questions, for some, the questions are still unsettled. Since, in this country, English language proficiency is likely foundational to academic success, perhaps more research should focus on which language programs better help students become fluent in English. As this is the focus of this study, it differs from most other program comparisons.

Since oral competency precedes reading or writing competency in natural first language development, this study also seeks to find a relationship between oral second language development and reading and writing skills. It is hoped that this can be useful in better understanding second language acquisition and later academic achievement in the second language.

To the first question: What is the best approach to teaching foreign students English? One approach to teaching these students has been the use of bilingual

programs—programs that teach students academic subject matter in their native language while providing English instruction. Another approach is to use only English to teach them language skills and academic skills. While there are proponents of both philosophies, currently there is no consensus among researchers regarding the superiority of one approach over the other. Therefore, there is an ongoing need to investigate the advantages that one program may have over the other.

Research Question 1

This study sought to answer two questions: (1) "Which language program type is more beneficial for second language acquisition: transitional bilingual education (TBE) or structured English immersion (SEI)?", and (2) "What is the relationship between oral language proficiency and literacy skills?"

To answer the first question it was necessary to find data from students that had spent time in these language programs so they could be compared. District X, a large urban school district in California, was chosen for its large pool of students in each program, and its extensive student database that spanned several years.

For the comparison to be legitimate, homogeneity between the groups would be important. Therefore, of the 2,731 students from whom the data were collect, 99% were Latinos, whose first language was Spanish. And, ninety-eight percent of theese students were from low socioeconomic backgrounds. To control for age and prior schooling effects, all students in the study enrolled in Kindergarten in the 2001-02 school year with no English proficiency.

All students were in either a structured English immersion program (SEI) program, a transitional bilingual program (TBE) program, or a program that was a combination of the two. By including students that switched from one program to another, it was hoped that the relative time in each program would shed light on the English acquisition process compared to only being in one program or another. To control for randomness, data for individual students were tracked longitudinally.

The instrument used to measure student English proficiency annually (the dependent variable) was the California English Language Development Test (CELDT). The CELDT is used to measure oral, reading, and writing proficiency in English, as opposed to academic proficiency in school curricular content areas. We can answer the question about language program efficacy by comparing the CELDT scores from the bilingual or structured English immersion programs the students are in. We can answer the research second question of this study by comparing oral CELDT proficiency levels to CELDT reading and writing proficiency levels.

To answer the first research question, how were these results obtained, and what do they mean? To be able to compare the two programs, after CELDT data from the school district were obtained, students were divided into two cohorts, which were in turn divided into different groups. The first cohort, named Analysis 1, was composed of four groups: a group that was only in SEI the full time (named EO); a group that was in bilingual education 1-3 years before switching to SEI (named BBS), a group that had 4-5 years of bilingual education before switching to SEI program (named BBL), and a group that was only in TBE the full time (named BBO). The second cohort, named Analysis 2, was only composed of two groups: a group

that was only in SEI the full time (named EI), and a group that was only in TBE the full time (named BB).

Why were two cohorts created? The scores of students in Analysis 1 were able to reflect the data of students from the SEI programs who moved out of SEI into mainstream programs, or who were in a bilingual program and later switched to an SEI program or mainstream program. Thus, their scores would reflect an advantage for the SEI groups over the BBO group, which did not leave the bilingual program. For this cohort, it would be expected that BBO would likely score lower than the other groups.

Table 6.1: Student Numerical Composition for the Analysis 1 Cohort

	2002 K	2003 G-1	2004 G-2	2005 G-3	2006 G-4	2007 G-5
EO	875	875	875	875	875	875
BBS	86	86	86	86	86	86
BBL	155	155	155	155	155	155
BBO	80	80	80	80	80	80
Total	1,196	1,196	1,196	1,196	1,196	1,196

Table 6.2 Student Numerical Composition for the Analysis 2 Cohort

	2002 K	2003 G-1	2004 G-2	2005 G-3	2006 G-4	2007 G-5
EI	2,094	2,094	1,714	1,495	1,271	973
BB	637	637	462	334	184	82
Total	2,731	2,731	2,176	1,829	1,455	1,055

The Analysis 2 cohort was designed to hold constant program participation while increasing the number of participating students. So, if there were an advantage of a cohort being larger, or composed of brighter or more talented

students participating earlier in the program, this would be reflected in the data. If this were the case, then the Analysis 2 group would show greater progress against the Analysis 1 group in the early years of the data comparison.

After dividing the students into cohort groups, the CELDT means were subject to several regression analyses to ensure the scores were driven by the program the students were placed in (see Chapters 3 and 4). Then three types of student performance on the CELDT were compared across program type. First, the raw CEDLT scale score means were compared. Then cohort groups were compared according to CELDT proficiency level. Finally, the highest achieving students were compared across program type. These three comparisons would help determine if one program were superior to the other (with the caveat for the bilingual-only groups).

Looking at the raw means from the CELDT scores in Tables 6.2 and 6.3 is revealing

Table 6.3 Analysis 1 Cohort: Overall CELDT Means

Year/ Grade	EO	BBS		BBL		BBO	
2002 K	220.00	220.00	PG	220.00	PG	220.00	PG
2003 G-1	463.75	427.14	-36	410.05	-54	415.45	-49
2004 G-2	486.20	465.64	-20	451.18	-35	446.03	-40
2005 G-3	487.06	469.59	-17	463.17	-24	458.01	-28
2006 G-4	515.51	508.87	-7	495.78	-20	492.48	-24
2007 G-6	527.55	527.49	0	512.00	-16	510.19	-18

PG = Point Gap

Table 6.4 Analysis 2 Cohort: Overall CELDT Means

Year / Grade	EI	BB	PG
2002 K	463.34	409.76	-53
2003 G-1	491.81	450.70	-41
2004 G-2	494.47	468.37	-26
2005 G-3	520.25	495.20	-25
2006 G-4	522.00	501.53	-20

PG = Point Gap

One of the first impressions when the two cohorts are compared is how similar the SEI groups are to each other. Any potential advantages one group would have had over the other were not manifested. The potential higher scores available to cohort 1 were of no advantage, and the larger numbers of potentially more capable students in cohort 2 didn't result in higher scores over all. After the first year (2003) where the greatest difference would be expected, the scores are identical. Every year after that the two cohorts are only an average of five points apart—not a meaningful difference. So it appears that being in an SEI program in this study yielded somewhat predictable results.

The results for the bilingual groups that switched to SEI are a little less straightforward to interpret. To begin with, the students in the BBS group were in a bilingual program for one to three years. That is, some of the students in BBS switched to SEI in 2003, some in 2004, and some in 2005. When their CELDT scores were analyzed they were close enough together to be considered one group for the entire study. One question that could be asked at this point is, "If being in an SEI group is an advantage, why did the different groups in BBS not have significantly

different scores from each other over the years?" The part of BBS that was in the bilingual program for only one year should have had significantly higher means that the part that was in the bilingual program for two years or three years, etc. Instead, they were very closely grouped together.

So for BBS, the 2003 results would be only results attributable to a bilingual influence, as it would be for the BBL, BBO, and BB (in cohort 2) groups. Their score of 427 in 2003, which is 17 points higher than BBL, BBO, and BB may indicate that these are a more talented group. Therefore, it is questionable that their merger (or even surpassing in the case of EO) with the SEI groups by 2007 can be attributed to SEI influence. If so, it would be difficult to explain how they surpassed EI in cohort 2 without somehow attributing it to a latent bilingual influence.

The BBL group also requires a closer look. Some students of this cohort were in the bilingual program for four years (until 2004), while others were in the program for five years (until 2006). Therefore, results indicating a pure bilingual influence on this group continue thorough 2006 (since the 2006 scores represent the influences through 2005—the scores reflect previous years treatment).

With this information, the comparison between the three bilingual groups is quite striking. BBL and BB are nearly identical for each year and there is seldom a point gap difference greater than five between them and BB. It's fairly safe to say that they are all virtually the same. The extra year of SEI that BBL received in 2006 accounted for an 11-point lead over BB and a 2-point lead over BBO—hardly an influence!

By 2007 the 5-year influences of the programs have run their course. While the SEI groups maintain an overall lead of between 16 – 20 points over the three bilingual groups, this lead is not very substantial, as the next section illustrates with graphs of CELDT proficiency levels. Viewing the Listening/Speaking, Reading, and Writing CELDT means tables yields similar conclusions. Using the CELDT means over time to compare the two groups helps answer research question 1: "Which language program type is more beneficial for second language acquisition: transitional bilingual education (TBE) or structured English immersion (SEI)?" The answer: The SEI groups begin with and maintain a slight lead over the bilingual groups that diminishes over time. This lead does not appear to be significant. If SEI were an overwhelmingly superior program, we would expect the point gap to be greater and not narrow annually.

The closeness between the bilingual and SEI programs is more evident when looking at CELDT proficiency levels.

Table 6.5 Annual Overall CELDT levels for the Analysis 1 Cohort

Grade /Year	2003 G-1	2004 G-2	2005 G-3	2006 G-4	2007 G-5
EO	E Int (L)	Int (L)	Int (L)	Int	Int
BBS	E Int (L)	E Int (H)	E Int	Int	Int
BBL	Beg	E Int	E Int	Int	Int
BBO	Beg	E Int (L)	E Int (L)	Int	Int

Beg = Beginning; E Int = Early Intermediate; Int = Intermediate
(L) = In the low range of a level; (H) = In the high range of a level

Table 6.6 Annual CELDT Overall Levels for the Analysis 2 Cohort

Grade/ Year	2003 G-1	2004 G-2	2005 G-3	2006 G-4	2007 G-5
EI	E Int	Int (L)	Int	Int (H)	Int
BB	Beg	E Int (L)	E Int	Int	Int

By 2006 all groups in both cohorts are at the Intermediate level. While the initial SEI lead is again apparent from looking at both tables, within five years each group is scoring in the same level, and remains so in 2007. Similar findings occur when the other proficiency level tables are consulted.

As the data emerge (Tables 6.6 and 6.7) for students who transitioned into mainstream classes in 2006 and 2007, reflecting the highest scoring students in all groups, the bilingual groups in both cohorts outperform their SEI counterparts in 2007.

Table 6.7 Analysis 1 Cohort – Students Who Reclassify as PRP

PRP	2003 G-1	2004 G-2	2005 G-3	2006 G-4	2007 G-5	Total
EO	0	0	2	17	134	153
EO %	0%	0%	0%	**2%**	**15%**	
BBS	0	0	0	1	11	12
BBS %	0%	0%	0%	**1%**	**13%**	
BBL	0	0	0	0	8	8
BBL %	0%	0%	0%	0%	**5%**	
BBO	0	0	0	2	13	15
BBO %	0%	0%	0%	**3%**	**16%**	

Table 6.8 Analysis 2 Cohort – Students Who Reclassify as PRP

PRP	2003 G-1	2004 G-2	2005 G-3	2006 G-4	2007 G-5	Total
EI	0	0	15	60	155	230
EI %	0%	0%	**1%**	**5%**	**19%**	
BB	0	0	0	5	14	19
BB %	0%	0%	**0%**	**3%**	**21%**	

It should be noted that even with the disadvantages of the bilingual-only groups in cohort 1 (i.e., not being able to use later SEI or mainstream data) and cohort 2 (not being able to take advantage of earlier higher scores), the bilingual groups did quite well. They did, as expected perform lower than the SEI groups.

Nevertheless, part of what may be happening in these programs is what Thomas and Collier (2002, 2004) observed in their longitudinal study. They noticed in the early years of schooling, students in Structured English immersion instruction out performed students in bilingual programs. Then, in the later years of secondary schooling they noticed a reversal in that trend. Unfortunately, this study did not have access to secondary school data for these students. The answer again to the first research question of this study can only be qualified. For the elementary school years, when students begin school with no language proficiency in English, students learn English more quickly initially and maintain that lead for four years across most measures. However, because the differences in performance between the two groups narrows each year, and in some important measures the bilingual groups surpass their SEI peers by the end of elementary school, it appears the bilingual program may ultimately be superior.

Jepsen's (2009) study was quite similar to this study. He also was only interested in second language proficiency, not how second language students performed academically on standardized tests. As he conducted a cross-sectional study of Spanish-speaking second language learners throughout California (some 500,000) students, he also found that initially SEI students learned English more rapidly, but by the third grade, differences began to disappear. Unfortunately, his study, like this one, only continued to grade 5.

There are some salient findings from this study that should be noted. Federal and California state legislation (see Chapter 1) expects students to advance one CELDT level for each year of schooling. The students in this study, and indeed throughout California (see Chapter 2) do not acquire English at that speed. While a few can progress that quickly, this study and many others (Hakuta, Butler & Witt, 2000; MacSwan & Pray, 2005) show that students progress at a much slower rate.

By the same token, this study (along with many others) should also be a red flag for states that have legislated against bilingual education. The philosophy behind the movements in California (Proposition 227 in 1998), Arizona (Proposition 203 in 2000), and Massachusetts (Question 2, in 2002) is that bilingual education does more harm than good, and that students after a year of learning English should be ready for regular classroom instruction. Clearly, no students in this study appeared to suffer from bilingual instructional treatment; on the contrary, they consistently showed improvement. Also, no students in this study were able to reclassify (become ready to be placed in regular English classes) for the entire six year period of the study. In fact very few students in District X performed that well.

Furthermore, researchers who have reviewed test scores in states that have curbed or forbidden bilingual education report that either (1) there was a *slight improvement* in test scores (Rossell, 2005), or (2) that test scores were *lower* (Lindholm-Leary & Genesee, 2010; Rumberger and Tran, 2010; Dan Losen, 2010). Apparently the wrong lobbyists are appearing in legislative assembly halls.

The value of bilingual education is not just that it doesn't interfere with second language acquisition, there are other benefits to be considered. Anecdotally, many Americans have heard stories of English speaking students who have studied foreign languages in traditional ways for a few years, only to find out that they cannot communicate in that language when visiting its host country. Bilingual programs seek to prepare students to communicate in and perform academically in a new language from the beginning. Many researchers have demonstrated that being bilingual can facilitate improvement in one's social, academic, and financial future, not to mention mental health (Gonzalez, 2000; Trejo, 2003; Bleakley & Chin, 2004; Bialystok, 2005, 2010; Bialystok & Craik, 2010). In this global economy it's difficult to imagine not putting bilingual and bicultural education on the nation's curricular front burner.

If we return to the first research question after these thoughts, we would be tempted just to say that bilingual programs are superior.

Research Question 2

The second research question of this study, "What is the relationship between oral language proficiency and literacy skills?" was somewhat easier to

answer. The CELDT proficiency level tables (Chapter 5) were able to show a clear relationship between oral acquisition and reading and writing acquisition. The CELDT utilizes reading and writing skills principally for the purpose of establishing English fluency, not to identify academic proficiency. For all students, bilingual or SEI, reading scores were one level below the oral scores for each year, and writing scores were usually one level below the oral scores for each year. This relationship is interesting in that it suggests that oral language is foundational to literacy, which has long been suspected (Gottlieb, 2006).

It follows a pattern suggested by Cummins (2000a), Hakuta et al. (2000), and Hakuta (2011) that oral academic proficiency precedes academic proficiency. It is however, very difficult to establish predictive power base on oral proficiency on how students will perform academically (Katz, Low, Stack, & Tsang, 2004). Nevertheless, Katz et al. (2004) and Hakuta (2011) have noticed moderate predictive ability based on overall CELDT scores for student performance on the California Standards Test (CST) in the early elementary years. Based on these examples, if CST data for this study's students were available we might predict that by 2006 the SEI and bilingual groups were scoring at a comparable level by 2006.

The fact that CELDT Overall proficiency is much less predictive for secondary academic outcomes lends weight to Cummins' (2000a) suggestion that social communicative skills follow a somewhat different developmental trajectory than academic proficiency. As the cognitive load increases in secondary years, there is weak correlation between very high language proficiency on the CELDT and state standardized exams (Katz et al., 2004). This invites further research into how oral

language skills can be developed in academic settings. If the CELDT is reflective of language proficiency, and oral language skills precede reading and writing skills, then perhaps emphasis needs to be placed in developing oral academic skills from the outset of introducing higher cognitive concepts. Working with elementary students, Tong, Lara-Alecio, Irby, Mathes, & Kwok, (2008) have demonstrated that intervention with elementary students (in both bilingual and SEI settings) can produce more rapid acquisition of academic English. Perhaps this is a "missing link" in secondary (or beyond) second language pedagogy.

Program Effectiveness vs. a Specific Bilingual Program or English Program

Several years ago August & Hakuta (1997) and Genesee (1999) suggested that in the case where bilingual services are not available, there is no one best model that will serve all students, and emphasize the importance of designing services for English learners that consider the community context, the needs of students who will be served, and the resources that are available for implementing the program.

A report issued by the New York City Board of Education (2000) on the progress of English learners in New York City Schools indicated that consistency of programmatic approach appeared to be a more important determinant of progress than the specific educational philosophy and methods of the bilingual/ESL programs. This advice seems prudent enough when resources need to be considered in program development.

Nevertheless, Hakuta, Butler, and Witt (2000) call for research that can more clearly identify salient features of program types that may be more advantageous when resources are present. They specifically highlight the need for longitudinal

research that can link oral and academic performance across program types so researchers can speak more authoritatively on program effectiveness. More recently, while expressing his confidence in the superiority of bilingual programs, Hakuta (2011) has reiterated the need to move past program types stating they are a distraction because of the current anti-bilingual sentiment in the country.

Many characteristics of effective school/communities have been highlighted in research literature (Williams et al. 2005, Williams et al. 2007). Some of the leading predictors of student achievement are schools that utilize assessment data, have needed resources, focus on standards-based instruction, and prioritize student achievement. Parent involvement, teacher collaboration, and high expectations for student performance are also necessary. These need to be in place regardless of language program type.

Contributions of This Study

The results of this study add to the research literature that students can learn a second language effectively while participating in bilingual and/or SEI programs. This study was able to include controls for many variables that had not been controlled for in earlier studies, including: (1) the number of subjects; (2) the homogeneity of the subjects (including age, home language, ethnicity, geographic location, SES status, the same school district, and types of schools); (3) the randomness of program inclusion; (4) the prior effects of language level and schooling; (5) the length of the study; (6) the use of standardized tests for all

subjects; and (7) the use of the same test for all subjects administered at the same time of year.

It is also clear from these results that bilingual education is not harmful to student English acquisition, as promoters of anti-bilingual education would have us believe. The initial performance gap between the two groups is overcome in time, and while there were no Spanish performance data available for this study, other studies have shown that bilingual students have the additional advantage of becoming more proficient in two languages (see Gándara & Hopkins, 2010; Lindholm-Leary & Genesee, 2010).

This study also clearly indicates that "one or two years of English language training" is not sufficient to prepare most students for mainstream classes. No students in this study from either program were reclassified over the six-year period of this study. While a few students were reclassified as PRP early in the study, no more than 21% from either cohort reached that level by 2007 in either cohort. Federal, state, and local benchmarks should be lowered accordingly.

This was also the only longitudinal study that compared the development of English language proficiency (and not academic performance) across program types, using both oral and literacy measures. The results were very consistent across all measures, and across time, indicating the reliability of the instrument and program influence. The only study similar to it was Jepsen (2009), which was a cross sectional study, which included some variables not included in this study. Nevertheless, his conclusions were quite similar to this study's conclusions.

This phase of the study also contributed to the body of research regarding oral studies. It demonstrated that student oral development precedes literacy development. It did so without comparing academic achievement to language proficiency. In other words, oral language proficiency was compared with literacy language proficiency. The CELDT is an indicator of how native-like a student's language is orally, and in reading and writing. It is therefore a multidimensional measure of English language proficiency, unlike many instruments that use reading to identify English language growth. It was therefore unnecessary to be concerned with the definition of "academic oral language" when linking oral skills with literacy skills. As Katz et al. (2004) showed, oral language fluency is not necessarily a good predictor of student performance on academic tests. The reading and writing portions of the CELDT reinforce the oral parts.

Other researchers (Cummins, 2000a, and Hakuta et al., 2000) have shown that oral fluency precedes academic fluency, and this study shows that literacy skills that indicate language fluency follow oral skills of the same measure.

The implication may be that greater emphasis should be placed on oral development, both communicative and academic, to ensure more rapid development in literacy domains. More research needs to be conducted to show the link between the type of oral skills needed for students to perform at high academic levels.

Some Limitations of This Study and Directions for Future Research

This study had some limitations. Future studies of this kind would benefit from more rigorous control for the amount of English and the primary language used in the classroom, and the amount of English used outside of the classroom. Katz. et al. (2004) pointed out several instances of how qualitative studies enhance test data. Looking only at test scores does not allow the researcher to get a full picture of many of the variables. As such, the data in this study was also not able to account directly for "best practices" in the classroom. A combination of direct classroom observations and student/teacher interviews would have enhanced this project. In this light, it was also difficult to control completely for transience. Students in District X often move between schools (either within or outside of the local mini district), making monitoring for consistency in quality of education a challenge.

Another shortfall of this comparison was the inability to monitor which classrooms may have had more than one proficiency level in the same class – this has been offered as one possible reason for the slow progress of ELL students in District X (Salazar, 2007).

Program compliance was also not possible to control for in the database for this project. While District X provided teachers in both programs with a pedagogical road map for instruction, the extent to how teachers as a whole adhered to program requirement could not be determined.

There are ways the state could make the CELDT more useful to researchers and teachers. More compartmentalized types of feedback for the oral, reading, and

writing tests than is currently provided by the CELDT could be developed to link up a more nuanced exam. Unlike the annual state exams, the CELDT does not provide feedback on sub-skill segments of the test (for example, the CELDT provides a reading score, but does not provide separate scores for the different parts of the reading test).

The current analyses only compared students who began at Kindergarten at zero English proficiency. Would the results have been different if students who started at different grade levels of no English proficiency have been different? At the district level, more students tend to reclassify in middle school than in elementary or high school when the entry-level proficiency is not considered. Alternatively, what would the results have been if the study had controlled for higher levels of proficiency upon enrollment at different grade levels?

Also, this study did not have access to state standardized tests (CSTs), which all students take annually. Comparing CELDT scores with CST performance would have made possible a comparison between this study and Katz et al.'s (2004) and Haktua's (2011) observations.

Currently, District X has very few students that are in no language program once they are identified as LEP. It would be interesting to compare the effects of "submersion" (students just being placed in regular mainstream classes without the benefit of special sheltered instruction) to students in English Immersion or a bilingual program.

At the time of this study, District X did not have many students in dual language immersion programs, but the number of programs is growing slowly

(currently available in Spanish, Korean, and Cantonese). This study would have benefited by including results from the dual language programs had that been possible. Hopefully, greater numbers of students in the near future will help this occur.

Implications of This Research for Policy

Independent of the program that ELLs are in, they do not perform well by federal, state, or district standards. California state standards (California Department of Education, 2013) require that students advance one CELDT level each year. In this study there were no students who were reclassified after six years of being in their respective programs. Students who entered District X at higher levels than the students in this study also took much longer to become proficient in English or to reclassify by the same standards. While this level of attainment may be in reach for a small number of students, it is clearly unrealistic to expect the majority of students to advance within the time frame these institutions have set for them. In this light, policy makers and educators need to adjust their expectations for ELL progress.

Furthermore, policy makers, administrators, and parents need not worry about placing students in bilingual programs. The initial early lead the Structured English immersion programs provide is both slight and temporary. This study and others cited above should slow down the momentum of political movements that seek to minimize or do away with bilingual programs altogether. Proponents of

these programs are clearly out of touch with current research and are seemingly unaware of the benefits of developing skills in two or more languages.

For students whose first language is not English, but who live in a community where the school district is not able to support the students' learning with a bilingual program in the students' first language, the focus must utilizing available resources. District X is a vast district whose students speak over 80 different languages, but currently there are only bilingual programs in Korean, Chinese, and Spanish. August and Hakuta (1997) and Genesee (1999, 2006) have suggested that there is no one best model that will serve all students, and emphasize the importance of designing services for English learners that consider the community context, the needs of students who will be served, and the resources that are available for implementing the program. A report issued by the New York City Board of Education (2000) on the progress of English learners in New York City Schools indicated that consistency of programmatic approach appeared to be a more important determinant of progress than the specific educational philosophy and methods of the bilingual/ESL programs. Nevertheless, Hakuta, Butler, and Witt (2000) and Hakuta (2011) call for research that can more clearly identify salient features of program types that may be more advantageous when resources are present. In this vein, Tong et al. (2008) and colleagues demonstrated that intervention techniques can facilitate language growth independent of program type (English immersion or bilingual). This is an invitation for policy makers and educators to get in touch with and stay abreast of current research.

APPENDIX

Appendix 1

Table 4.1 Regressions on CELDT Overall Means for Analysis 1 Cohort (Extended)

VARIABLES	OLS (no controls) 1	OLS (controls) 2	FE (Fixed Effects) 3	Mixed Effects (no controls) 4	Mixed Effects (controls) 5
All Years in BB					
Difference 2003	-48.30***	-48.07***	-47.62***	-48.04***	-47.92***
	(8.105)	(8.137)	(8.116)	(9.391)	(9.413)
Difference 2004	-40.17***	-40.46***	-40.01***	-40.22***	-40.69***
	(3.921)	(3.989)	(3.944)	(4.391)	(4.488)
Difference 2005	-29.04***	-29.72***	-28.59***	-28.85***	-29.59***
	(3.754)	(3.783)	(3.754)	(5.686)	(5.639)
Difference 2006	-23.03***	-24.05***	-22.83***	-22.84***	-23.99***
	(3.918)	(3.899)	(3.879)	(6.279)	(6.231)
Difference 2007	-17.36***	-18.81***	-17.21***	-17.21**	-18.83***
	(4.763)	(4.849)	(4.782)	(6.847)	(7.058)
4 or 5 Years in BB					
Difference 2003	-53.71***	-53.51***	-52.98***	-53.50***	-53.42***
	(6.179)	(6.195)	(6.180)	(6.976)	(6.954)
Difference 2004	-35.02***	-34.83***	-34.37***	-35.17***	-35.13***
	(2.761)	(2.761)	(2.758)	(5.524)	(5.511)
Difference 2005	-23.88***	-23.87***	-22.86***	-22.98***	-23.03***
	(3.124)	(3.137)	(3.125)	(4.772)	(4.765)
Difference 2006	-19.73***	-20.08***	-18.92***	-18.82***	-19.27***
	(2.782)	(2.814)	(2.800)	(4.232)	(4.285)
Difference 2007	-15.55***	-15.73***	-14.64***	-14.90***	-15.18***
	(3.121)	(3.148)	(3.133)	(4.547)	(4.580)
1,2,or 3 years in BB					
Difference 2003	-36.61***	-36.39***	-36.44***	-36.20***	-36.03***
	(7.062)	(7.061)	(7.048)	(12.67)	(12.70)
Difference 2004	-20.56***	-20.23***	-20.33***	-20.34***	-20.02***
	(3.838)	(3.862)	(3.799)	(6.691)	(6.703)
Difference 2005	-17.46***	-17.44***	-17.26***	-17.39***	-17.19***
	(3.630)	(3.683)	(3.628)	(4.181)	(4.154)
Difference 2006	-6.635**	-6.643**	-6.474**	-6.332	-6.204
	(3.238)	(3.304)	(3.251)	(4.245)	(4.209)
Difference 2007	-0.0568	0.385	0.257	-0.0223	0.422
	(4.431)	(4.527)	(4.429)	(4.962)	(4.949)
All Years in BB Dummy	0	1.796**		-2.602	-0.991
		(0.850)		(3.174)	(3.074)
4 or 5 Years in BB Dummy	0	2.401***		-0.531	0.609
	(3.63e-07)	(0.728)		(2.217)	(2.203)
1,2,or 3 years in BB Dummy	0	-1.624		1.156	0.847
	(5.31e-08)	(1.315)		(3.777)	(3.999)

Year=2003	243.8***	243.6***	243.1***	243.5***	243.4***
	(1.553)	(1.574)	(1.558)	(1.829)	(1.827)
Year=2004	266.2***	266.1***	265.3***	266.0***	266.0***
	(1.031)	(1.047)	(1.072)	(1.312)	(1.309)
Year=2005	267.1***	266.9***	266.0***	266.9***	266.8***
	(1.131)	(1.161)	(1.165)	(1.375)	(1.377)
Year=2006	295.5***	295.2***	294.5***	295.3***	295.3***
	(1.154)	(1.175)	(1.181)	(1.359)	(1.354)
Year=2007	307.5***	307.0***	306.5***	307.4***	307.1***
	(1.333)	(1.366)	(1.387)	(1.679)	(1.692)
Free Lunch		-1.445	-0.668		-1.110
		(1.436)	(1.227)		(1.401)
Ethnicity (Relative to Native America)					
Asian		44.73***			37.39***
		(4.338)			(5.882)
Filipino		-6.076			-4.935
		(3.899)			(5.751)
Hispanic		53.07***			51.21***
		(13.45)			(14.57)
School District (Relative to 1)					
2		4.534	6.679		3.939
		(2.763)	(9.310)		(3.219)
3		2.409	8.608		7.142*
		(3.474)	(9.064)		(4.263)
4		-3.446	15.08**		-1.077
		(2.288)	(7.416)		(2.925)
5		-7.627***	8.319		-8.339***
		(2.205)	(7.938)		(2.543)
6		1.339	19.54**		0.915
		(2.834)	(9.466)		(3.308)
7		-6.689***	10.32		-8.364***
		(2.356)	(7.790)		(2.766)
8		-1.429	17.55*		-0.551
		(3.974)	(10.26)		(4.269)
9		14.65**	8.698		9.130
		(6.661)	(11.76)		(6.268)
11		-12.02*	11.31		-15.23**
		(6.261)	(11.49)		(6.114)
School Type (Relative to Elementary)					
Magnet Center on Regular Campus		19.82***	11.49**		20.73***
		(4.458)	(5.434)		(2.574)
Elementary Magnet (Self Contained)		-18.02***	-28.59***		-21.53***
		(5.800)	(8.644)		(5.728)
Primary Center		0.381	-11.42***		-0.885
		(2.364)	(3.030)		(4.033)
Span Magnet		-13.20***	-27.77***		-18.02***
		(3.389)	(1.460)		(4.225)
Span School (Not Magnet)		5.907	-6.905		1.448
		(5.602)	(4.421)		(4.219)

Constant	220***	159.9***	211.3***	221.5***	160.0***
	(1.42e-07)	(14.11)	(6.589)	(0.679)	(15.31)
Observations	7,176	7,176	7,176	7,176	7,176
R-squared	0.895	0.897	0.938		
Number of groups				321	321
Number of id			1,196		

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Appendix 2

Table 4.2 Regressions on CELDT Listening/Speaking Means for Analysis 1 Cohort (Extended)

VARIABLES	OLS (no controls) 1	OLS (controls) 2	FE (Fixed Effects) 3	Mixed Effects (no controls) 4	Mixed Effects (controls) 5
All Years in BB					
Difference 2003	-48.30***	-47.87***	-47.67***	-48.13***	-47.85***
	(8.105)	(8.137)	(8.117)	(9.393)	(9.449)
Difference 2004	-41.37***	-41.69***	-41.33***	-41.64***	-42.11***
	(5.451)	(5.514)	(5.473)	(5.509)	(5.660)
Difference 2005	-28.30***	-28.83***	-27.80***	-28.26***	-28.89***
	(4.487)	(4.486)	(4.476)	(6.824)	(6.717)
Difference 2006	-24.13***	-25.42***	-24.08***	-24.27***	-25.65***
	(4.831)	(4.836)	(4.811)	(9.079)	(9.198)
4 or 5 Years in BB					
Difference 2003	-53.71***	-53.30***	-53.02***	-53.62***	-53.36***
	(6.179)	(6.193)	(6.180)	(6.967)	(6.937)
Difference 2004	-39.10***	-38.79***	-38.52***	-39.65***	-39.44***
	(3.632)	(3.630)	(3.629)	(6.704)	(6.658)
Difference 2005	-23.08***	-22.81***	-21.97***	-22.56***	-22.39***
	(3.748)	(3.779)	(3.767)	(4.934)	(4.915)
Difference 2006	-21.84***	-22.26***	-20.98***	-21.65***	-22.11***
	(3.728)	(3.778)	(3.752)	(4.888)	(4.969)
1,2,or 3 years in BB					
Difference 2003	-36.61***	-36.42***	-36.23***	-36.38***	-36.23***
	(7.062)	(7.055)	(7.070)	(12.67)	(12.71)
Difference 2004	-22.80***	-22.41***	-22.19***	-22.88***	-22.43***
	(4.494)	(4.537)	(4.503)	(6.698)	(6.685)
Difference 2005	-20.60***	-20.84***	-20.28***	-20.90***	-20.84***
	(4.086)	(4.115)	(4.113)	(4.170)	(4.099)
Difference 2006	-8.864**	-9.136**	-8.397**	-8.931*	-8.931*
	(4.077)	(4.104)	(4.106)	(4.993)	(4.852)
All Years in BB Dummy	-0	2.393***		-2.592	-0.584
		(0.906)		(3.390)	(3.201)
4 or 5 Years in BB Dummy	-0	2.521***		0.838	2.138
	(1.78e-07)	(0.770)		(2.253)	(2.239)
1,2,or 3 years in BB Dummy	-0	-1.453		2.753	2.135
	(1.92e-07)	(1.299)		(3.603)	(3.821)
Year=2003	243.8***	243.5***	243.2***	243.6***	243.4***

	(1.553)	(1.572)	(1.566)	(1.827)	(1.819)
Year=2004	296.3***	296.2***	295.6***	296.3***	296.3***
	(1.386)	(1.395)	(1.417)	(1.629)	(1.633)
Year=2005	275.3***	275.2***	274.5***	275.3***	275.3***
	(1.425)	(1.446)	(1.468)	(1.634)	(1.649)
Year=2006	307.9***	307.4***	306.9***	308.0***	307.8***
	(1.492)	(1.501)	(1.516)	(1.751)	(1.736)
Free Lunch		-4.759**	-2.378		-3.949**
		(1.867)	(1.639)		(1.996)
Ethnicity (Relative to Native America)					
Asian		44.15***			40.21***
		(5.906)			(7.998)
Filipino		-6.600			-5.014
		(5.543)			(7.900)
Hispanic		48.83***			42.83***
		(7.050)			(8.876)
School District (Relative to 1)					
2		4.353	0.224		4.370
		(3.028)	(12.45)		(3.295)
3		4.283	20.30*		8.454*
		(3.829)	(12.01)		(4.667)
4		-3.226	22.49**		-0.117
		(2.520)	(9.975)		(3.238)
5		-7.048***	15.12		-7.821***
		(2.507)	(10.54)		(2.743)
6		2.947	30.46**		2.329
		(3.254)	(12.16)		(3.942)
7		-3.381	16.82		-5.040*
		(2.574)	(10.92)		(2.928)
8		-1.883	33.35*		-0.774
		(4.234)	(17.44)		(4.751)
9		13.11*	35.29**		9.994**
		(7.362)	(17.02)		(4.995)
11		-2.653	28.68**		-5.422
		(6.136)	(11.79)		(5.395)
School Type (Relative to Elementary)					
Elementary Magnet (Self Contained)		23.51***	13.43		24.70***
		(6.977)	(8.390)		(6.250)
Primary Center		0.818	-8.952***		-1.225
		(2.433)	(3.276)		(4.005)
Span Magnet		-8.442**	-28.24***		-12.11**

		(3.898)	(2.153)		(5.004)
Span School (Not Magnet)		2.500	-10.91**		-2.282
		(7.500)	(5.181)		(4.934)
Constant	220	191.3***	206.3***	221.0***	192.9***
		(7.380)	(8.520)	(0.651)	(9.170)
Observations	5,980	5,980	5,980	5,980	5,980
R-squared	0.885	0.887	0.928		
Number of groups				301	301
Number of id			1,196		

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, *
p<0.1

Appendix 3

Table 4.3 Regressions on CELDT Reading Means for Analysis 1 Cohort (Extended)

VARIABLES	OLS (no controls) 1	OLS (controls) 2	FE (Fixed Effects) 3	Mixed Effects (no controls) 4	Mixed Effects (controls) 5
All Years in BB					
Difference 2005	10.02**	9.416*	10.35**	10.40**	9.624**
	(5.043)	(5.107)	(5.039)	(4.185)	(4.204)
Difference 2006	13.02***	12.58***	13.42***	13.67***	12.80***
	(4.618)	(4.677)	(4.649)	(4.490)	(4.595)
Difference 2007	8.979**	7.647*	9.402**	9.455***	7.883**
	(4.362)	(4.398)	(4.393)	(3.597)	(3.728)
4 or 5 Years in BB					
Difference 2005	3.978	3.606	4.314	5.339	4.694
	(3.656)	(3.650)	(3.673)	(3.829)	(3.618)
Difference 2006	6.613**	6.277*	6.957**	8.333**	7.338*
	(3.353)	(3.379)	(3.383)	(3.829)	(3.838)
Difference 2007	6.476*	6.092	6.778*	7.571**	6.656*
	(3.750)	(3.789)	(3.744)	(3.698)	(3.690)
1,2,or 3 years in BB					
Difference 2005	-3.740	-3.579	-3.661	-3.757	-3.490
	(4.829)	(4.894)	(4.847)	(5.293)	(5.276)
Difference 2006	7.455*	7.688*	7.617*	7.651*	7.981*
	(4.161)	(4.206)	(4.187)	(4.306)	(4.446)
Difference 2007	10.36**	11.18**	10.99**	11.06**	10.88**
	(4.561)	(4.646)	(4.584)	(4.572)	(4.952)
All Years in BB Dummy	-30.14***	-29.19***		-30.36***	-29.63***
	(3.674)	(3.798)		(4.822)	(4.238)
4 or 5 Years in BB Dummy	-22.33***	-19.64***		-29.42***	-22.54***
	(3.056)	(3.177)		(3.649)	(4.371)
1,2,or 3 years in BB Dummy	-10.61**	-13.38***		-16.37***	-13.20**
	(4.262)	(4.347)		(4.974)	(5.568)
Year=2005	19.58***	19.45***	19.53***	19.77***	19.68***
	(1.281)	(1.301)	(1.282)	(1.410)	(1.415)
Year=2006	48.04***	47.89***	47.98***	47.96***	48.08***
	(1.266)	(1.280)	(1.289)	(1.425)	(1.435)
Year=2007	76.62***	76.51***	76.63***	76.60***	76.76***
	(1.375)	(1.424)	(1.415)	(1.660)	(1.691)
Free Lunch		2.547	0.724		2.305
		(2.089)	(1.516)		(1.725)
Ethnicity (Relative to Native America)					
Asian		86.78***			81.41***
		(24.29)			(25.90)
Filipino		15.04			17.07
		(23.99)			(25.78)
Hispanic		47.93*			55.63*
		(28.40)			(28.74)
School District (Relative to 1)					
2		7.392*	0.125		5.838

		(4.196)	(7.639)		(4.964)
3		2.925	-7.352		8.290
		(4.913)	(11.37)		(5.552)
4		-4.802	5.356		-1.966
		(3.559)	(7.058)		(4.405)
5		-8.202**	-0.258		-8.814**
		(3.315)	(7.758)		(3.818)
6		-2.204	10.04		-2.086
		(4.127)	(10.30)		(4.389)
7		-12.52***	5.799		-10.80**
		(3.658)	(7.728)		(4.348)
8		4.297	18.72		6.354
		(5.421)	(14.88)		(5.894)
9		22.61***	-4.061		16.52
		(8.056)	(15.26)		(10.54)
11		-20.23***	1.772		-21.72***
		(7.501)	(11.48)		(5.776)
School Type (Relative to Elementary)					
Magnet Center on Regular Campus		22.09***	4.034		21.07***
		(5.825)	(7.854)		(3.146)
Elementary Magnet (Self Contained)		-33.76***	-69.71***		-36.27***
		(6.836)	(14.36)		(6.886)
Primary Center		-6.552	-1.808		-8.562
		(10.52)	(10.22)		(9.529)
Span Magnet		-6.283	-21.32***		-12.18**
		(4.733)	(1.688)		(5.026)
Span School (Not Magnet)		10.42*	-9.152		6.633
		(5.832)	(6.006)		(5.600)
Constant	443.3***	437.1***	433.9***	446.2***	435.6***
	(1.205)	(29.06)	(6.223)	(1.457)	(29.41)
Observations	4,784	4,784	4,784	4,784	4,784
R-squared	0.365	0.394	0.624		
Number of groups				309	309
Number of id			1,196		

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Appendix 4

Table 4.4 Regressions on CELDT Writing Means for Analysis 1 Cohort (Extended)

VARIABLES	OLS (no controls) 1	OLS (controls) 2	FE (Fixed Effects) 3	Mixed Effects (no controls) 4	Mixed Effects (controls) 5
All Years in BB					
Difference 2005	8.536*	7.793	8.151*	8.655*	8.390*
	(4.856)	(4.949)	(4.883)	(4.490)	(4.508)
Difference 2006	21.14***	20.57***	20.81***	21.44***	21.19***
	(5.555)	(5.630)	(5.608)	(4.792)	(4.803)
Difference 2007	23.92***	22.58***	24.24***	24.12***	23.33***
	(5.572)	(5.657)	(5.623)	(5.098)	(5.001)
4 or 5 Years in BB					
Difference 2005	8.580***	8.027**	8.280***	9.430***	9.259***
	(3.120)	(3.150)	(3.131)	(3.031)	(3.014)
Difference 2006	19.97***	19.50***	19.59***	21.03***	20.80***
	(3.399)	(3.449)	(3.413)	(3.468)	(3.489)
Difference 2007	24.11***	23.59***	23.81***	24.63***	24.38***
	(3.854)	(3.894)	(3.880)	(3.531)	(3.562)
1,2,or 3 years in BB					
Difference 2005	11.78**	11.76**	11.57**	11.67*	11.81*
	(4.755)	(4.789)	(4.796)	(6.346)	(6.242)
Difference 2006	20.51***	20.44***	20.10***	20.47***	20.47***
	(5.271)	(5.329)	(5.322)	(7.545)	(7.459)
Difference 2007	26.86***	26.74***	26.95***	27.14***	26.90***
	(5.632)	(5.680)	(5.758)	(6.919)	(6.891)
All Years in BB Dummy	-47.92***	-47.01***		-49.43***	-48.35***
	(4.988)	(5.091)		(5.206)	(5.015)
4 or 5 Years in BB Dummy	-39.58***	-37.01***		-40.96***	-39.89***
	(3.582)	(3.582)		(4.766)	(4.667)
1,2,or 3 years in BB Dummy	-26.05***	-28.62***		-28.45***	-28.53***
	(5.545)	(5.405)		(8.093)	(7.894)
Year=2005	25.81***	25.83***	26.05***	25.89***	25.92***
	(1.183)	(1.202)	(1.190)	(1.320)	(1.326)
Year=2006	46.08***	46.03***	46.54***	45.98***	46.13***
	(1.379)	(1.404)	(1.371)	(1.598)	(1.588)
Year=2007	55.12***	55.09***	55.43***	54.98***	55.16***
	(1.349)	(1.400)	(1.395)	(1.506)	(1.529)
Free Lunch		1.881	1.025		1.628
		(1.937)	(1.682)		(1.765)
Ethnicity (Relative to Native America)					
Asian		46.73***			47.59***
		(3.936)			(4.411)
Filipino		-9.493***			-0.591
		(2.037)			(4.122)
Hispanic		-4.216			4.673
		(21.38)			(23.27)

School District (Relative to 1)					
2		5.459	24.27		4.619
		(3.718)	(23.01)		(4.967)
3		-5.305	-4.068		1.470
		(5.048)	(20.08)		(5.897)
4		-8.096**	21.46		-4.572
		(3.322)	(16.77)		(4.652)
5		-10.53***	-3.013		-9.114**
		(3.122)	(17.17)		(4.397)
6		-0.828	3.822		-0.318
		(4.250)	(18.10)		(5.063)
7		-16.52***	2.713		-15.22***
		(3.444)	(16.92)		(4.758)
8		2.688	-7.881		3.727
		(4.869)	(19.34)		(5.426)
9		22.25***	23.01		19.54***
		(7.636)	(23.19)		(6.976)
11		-35.62***	-1.598		-33.69***
		(7.226)	(18.08)		(6.240)
School Type (Relative to Elementary)					
Magnet Center on Regular Campus		16.04***	-8.765		9.907*
		(5.861)	(6.971)		(5.400)
Elementary Magnet (Self Contained)		14.77**	4.938		13.82**
		(7.237)	(11.22)		(6.791)
Primary Center		5.539	14.07*		4.392
		(8.665)	(7.964)		(7.815)
Span Magnet		1.419	-5.439***		-5.031
		(4.733)	(1.834)		(5.036)
Span School (Not Magnet)		12.77**	6.308		9.663*
		(5.645)	(6.846)		(5.828)
Constant	470.4***	497.5***	453.5***	472.8***	486.2***
	(1.263)	(22.39)	(15.58)	(1.543)	(24.06)
Observations	4,784	4,784	4,784	4,784	4,784
R-squared	0.309	0.348	0.525		
Number of groups				309	309
Number of id			1,196		

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Appendix 5

Table 4.5 Regressions on CELDT Overall Means for Analysis 2 Cohort (Extended)

VARIABLES	OLS (no controls) 1	OLS (controls) 2	FE (Fixed Effects) 3	Mixed Effects (no controls) 4	Mixed Effects (controls) 5
Average Annual Increase Group X Year (EI relative to BB)	**3.174*** **	**3.259*** **	**3.776*** **	**3.466*** **	**3.548*** **
	(1.025)	(0.991)	(0.700)	(1.128)	(1.087)
Group (EI)	-6,331***	-6,502***		-6,917***	-7,082***
	(2,053)	(1,986)		(2,259)	(2,178)
Year=2003	228.4***	228.2***	170.1***	227.7***	227.6***
	(2.727)	(2.764)	(1.148)	(2.776)	(2.791)
Year=2004	257.6***	257.4***	140.1***	257.1***	256.9***
	(2.335)	(2.327)	(0.810)	(2.449)	(2.424)
Year=2005	261.1***	260.8***	84.37***	260.2***	260.0***
	(2.988)	(2.973)	(0.858)	(3.261)	(3.208)
Year=2006	285.2***	284.7***	52.28***	284.2***	283.8***
	(3.867)	(3.852)	(0.931)	(4.264)	(4.194)
Year=2007	287.1***	286.5***		286.3***	285.6***
	(4.861)	(4.807)		(5.434)	(5.345)
Free Lunch		-2.047	-1.201		-1.960
		(1.515)	(1.218)		(1.561)
Ethnicity (Relative to Native America)					
Asian		35.29***			29.74***
		(7.823)			(11.28)
Black		6.521			9.666
		(5.053)			(9.162)
Hispanic		-6.432			-2.199
		(4.793)			(5.899)
White		26.30***			20.20*
		(8.214)			(10.65)
Filipino		-7.890			-9.984
		(6.450)			(6.931)
Pacific Islander		4.981			-0.192
		(6.986)			(9.090)
School District (Relative to 1)					
2		3.482	6.624		2.734
		(2.375)	(6.095)		(2.419)
3		2.353	7.503		6.786**
		(4.009)	(6.603)		(3.437)
4		-2.120	6.567		-0.0529
		(2.392)	(5.578)		(2.526)
5		-8.427***	-1.791		-8.841***
		(2.120)	(6.163)		(2.263)
6		1.849	15.74**		1.639
		(2.871)	(7.512)		(2.786)
7		-5.043*	1.806		-7.533***
		(2.897)	(6.122)		(2.495)
8		-2.557	2.516		-2.633
		(4.171)	(8.964)		(3.273)

9		0.267	14.70		-0.424
		(2.270)	(11.72)		(3.869)
11		-5.691	-0.909		-9.083
		(5.788)	(9.572)		(6.662)
School Type (Relative to Elementary)					
Magnet Center on Regular Campus		24.80***	20.91***		26.64***
		(4.473)	(4.583)		(4.968)
Elementary Magnet (Self Contained)		18.69	29.83***		17.97
		(13.75)	(6.788)		(14.31)
Primary Center		-0.287	-10.35***		-0.249
		(3.253)	(2.517)		(3.189)
Span Magnet		17.05	-4.845		16.05
		(21.03)	(20.58)		(22.97)
Span School (Not Magnet)		-0.767	-3.703		-4.604
		(3.748)	(5.470)		(4.323)
Constant	201.7***	170.0***	-120,401***	204.9***	169.0***
	(1.872)	(9.736)	(518.4)	(2.107)	(11.65)
Observations	11,977	11,977	11,977	11,977	11,977
R-squared	0.888	0.890	0.929		
Number of groups			2,731	381	381

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Appendix 6

Table 4.6 Regressions on CELDT Listening/Speaking Means for Analysis 2 Cohort (Extended)

VARIABLES		OLS (no controls) 1	OLS (controls) 2	FE (Fixed Effects) 3	Mixed Effects (no controls) 4	Mixed Effects (controls) 5
Average Annual Increase Group X Year (EI relative to BB)		**5.655*** **	**5.575*** **	**5.950*** **	**5.903*** **	**5.853*** **
		(1.328)	(1.348)	(0.797)	(1.400)	(1.403)
Group (EI)		-11,300***	-11,142***		-11,798***	-11,699***
		(2,660)	(2,699)		(2,803)	(2,810)
Year (Relative to 2003)						
Year=2003		226.5***	226.4***	154.1***	225.9***	225.8***
		(2.772)	(2.814)	(1.137)	(2.832)	(2.854)
Year=2004		284.2***	284.3***	138.6***	283.7***	283.9***
		(2.941)	(2.986)	(1.046)	(3.028)	(3.053)
Year=2005		263.5***	263.7***	44.91***	262.8***	262.9***
		(3.916)	(3.998)	(1.095)	(4.127)	(4.157)
Year=2006		288.8***	288.7***		288.1***	287.9***
		(5.217)	(5.371)		(5.490)	(5.570)
Free Lunch			-4.210**	-2.623*		-4.040**
			(1.833)	(1.556)		(1.838)
Ethnicity (Relative to Native America)						
Asian			25.60**			23.75
			(12.46)			(16.19)
Black			9.347			10.45
			(6.033)			(9.725)
Hispanic			-8.109			-3.754
			(5.409)			(7.081)
White			7.937			2.003
			(12.72)			(15.33)
Filipino			-8.615			-12.49
			(6.959)			(8.067)
Pacific Islander			20.06*			18.70
			(11.45)			(13.07)
School District (Relative to 1)						
	2		3.235	4.336		2.791
			(2.439)	(8.540)		(2.643)
	3		4.161	15.75*		8.340**
			(4.056)	(8.966)		(3.616)
	4		-2.145	10.99		0.383
			(2.691)	(7.944)		(2.806)
	5		-8.888***	0.0378		-9.945***
			(2.423)	(8.524)		(2.580)
	6		2.908	21.45**		2.444
			(3.117)	(9.836)		(3.123)
	7		-3.240	3.262		-5.960**
			(3.022)	(8.829)		(2.644)
	8		-3.305	7.880		-2.934
			(4.122)	(13.21)		(3.366)
	9		1.057	36.12**		-0.405

		(2.259)	(15.49)		(3.056)
11		2.622	8.903		-0.277
		(4.844)	(11.15)		(5.728)
School Type (Relative to Elementary)					
Magnet Center on Regular Campus		23.76***	19.69***		25.41***
		(6.932)	(7.149)		(7.517)
Elementary Magnet (Self Contained)		30.93***	38.36**		30.20***
		(6.325)	(18.50)		(6.386)
Primary Center		-0.0244	-9.316***		-0.904
		(2.918)	(2.722)		(2.869)
Span Magnet		33.43	11.44		33.29
		(24.36)	(35.33)		(27.03)
Span School (Not Magnet)		-3.071	-3.392		-7.291*
		(3.613)	(6.779)		(4.118)
Constant	203.3***	199.9***	-152,459***	206.6***	202.7***
	(1.885)	(13.80)	(642.7)	(2.164)	(16.86)
Observations	10,922	10,922	10,922	10,922	10,922
R-squared	0.878	0.880	0.920		
Number of groups			2,731	370	370

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Appendix 7

Table 4.7 Regressions on CELDT Reading Means for Analysis 2 Cohort (Extended)

VARIABLES	OLS (no controls) 1	OLS (controls) 2	FE (Fixed Effects) 3	Mixed Effects (no controls) 4	Mixed Effects (controls) 5
Average Annual Increase Group X Year (EI relative to BB)	**-4.195****	**-3.647****	**-3.238*****	**-4.186*****	**-3.852****
	(1.633)	(1.526)	(1.175)	(1.528)	(1.501)
Group (EI)	8,435**	7,334**		8,418***	7,748**
	(3,274)	(3,059)		(3,063)	(3,008)
Year (Relative to 2004)					
Year=2005	27.83***	27.54***	-4.729***	28.12***	27.93***
	(1.679)	(1.619)	(0.846)	(1.678)	(1.650)
Year=2006	58.56***	58.08***	-0.741	58.93***	58.81***
	(3.009)	(2.777)	(0.892)	(2.859)	(2.779)
Year=2007	82.88***	82.03***		84.10***	83.63***
	(4.554)	(4.241)		(4.252)	(4.191)
Free Lunch		3.176	1.468		2.397
		(2.173)	(1.424)		(1.826)
Ethnicity (Relative to Native America)					
Asian		97.41***			92.90***
		(19.01)			(16.28)
Black		15.69			26.72
		(22.41)			(28.56)
Hispanic		16.69			25.48*
		(15.29)			(14.99)
White		87.92***			86.76***
		(21.53)			(17.63)
Filipino		0			0
		(0)			(0)
Pacific Islander					
School District (Relative to 1)					
2		6.189	1.134		5.042
		(4.257)	(5.945)		(4.083)
3		1.159	-1.098		5.125
		(5.373)	(9.098)		(5.013)
4		-3.537	4.809		-1.520
		(4.052)	(5.887)		(4.185)
5		-8.045**	-1.815		-8.121**
		(3.698)	(6.679)		(3.602)
6		-0.945	1.654		0.0688
		(4.396)	(10.23)		(4.224)
7		-10.20***	4.540		-10.19**
		(3.842)	(6.634)		(4.129)
8		3.347	16.56		1.668
		(6.389)	(12.83)		(5.340)
9		-3.857	0.337		-4.908

		(6.798)	(12.99)		(13.72)
11		-18.27***	3.952		-19.68***
		(5.475)	(14.32)		(5.907)
School Type (Relative to Elementary)					
Magnet Center on Regular Campus		25.32***	7.231		22.72***
		(3.811)	(7.631)		(4.464)
Elementary Magnet (Self Contained)		-9.147	-18.29		-9.012
		(23.87)	(19.19)		(24.54)
Primary Center		3.681	-6.996		3.446
		(6.842)	(7.890)		(8.331)
Span Magnet		34.84	7.488		33.66
		(26.58)	(25.10)		(28.16)
Span School (Not Magnet)		5.108	-2.882		2.623
		(3.924)	(5.684)		(5.220)
Constant	419.9***	377.3***	-52,726***	421.9***	373.3***
	(3.236)	(22.36)	(813.0)	(2.828)	(18.06)
Observations	6,515	6,515	6,515	6,515	6,515
R-squared	0.295	0.324	0.586		
Number of groups			2,250	359	359

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Appendix 8

Table 4.8 Regressions on CELDT Writing Means for Analysis 2 Cohort (Extended)

VARIABLES	OLS (no controls) 1	OLS (controls) 2	FE (Fixed Effects) 3	Mixed Effects (no controls) 4	Mixed Effects (controls) 5
Average Annual Increase Group X Year (EI relative to BB)	-8.293***	-7.834***	-8.574***	-8.833***	-8.467***
	(1.839)	(1.672)	(1.314)	(1.565)	(1.537)
Group (EI)	16,664***	15,741***		17,746***	17,010***
	(3,687)	(3,353)		(3,137)	(3,081)
Year (Relative to 2004)					
Year=2005	38.43***	38.37***	8.558***	39.24***	39.11***
	(1.928)	(1.798)	(0.800)	(1.710)	(1.694)
Year=2006	65.33***	64.93***	11.03***	66.72***	66.41***
	(3.429)	(3.126)	(0.899)	(2.895)	(2.859)
Year=2007	75.95***	75.02***		78.52***	77.75***
	(5.255)	(4.774)		(4.393)	(4.324)
Free Lunch		0.417	0.517		0.00193
		(1.747)	(1.519)		(1.614)
Ethnicity (Relative to Native America)					
Asian		67.67***			66.31***
		(16.09)			(14.48)
Black		-11.10			-4.486
		(7.828)			(11.62)
Hispanic		0.322			8.792
		(8.078)			(10.25)
White		37.35*			38.17**
		(20.82)			(17.78)
Filipino		0			0
		(0)			(0)
Pacific Islander					
School District (Relative to 1)					
2		5.155	25.61*		4.638
		(3.408)	(14.19)		(3.974)
3		-7.873	-4.907		-2.668
		(5.370)	(14.96)		(4.883)
4		-7.112**	19.17*		-4.522
		(3.315)	(11.16)		(3.964)
5		-9.676***	3.541		-8.305**
		(2.762)	(11.89)		(3.476)
6		-0.930	16.78		0.208
		(3.908)	(12.94)		(4.221)
7		-13.12***	14.06		-13.41***
		(3.441)	(11.95)		(4.046)
8		2.866	8.334		0.861
		(5.691)	(16.86)		(5.176)
9		-2.250	15.59		-5.516
		(5.370)	(17.04)		(10.14)
11		-24.40***	6.458		-24.72***

		(7.582)	(13.89)		(8.150)
School Type (Relative to Elementary)					
Magnet Center on Regular Campus		20.17***	6.788		16.55***
		(6.154)	(7.081)		(6.284)
Elementary Magnet (Self Contained)		3.612	14.96		4.946
		(40.07)	(32.87)		(41.45)
Primary Center		20.57***	13.35**		20.08**
		(7.206)	(6.612)		(8.667)
Span Magnet		25.07	-8.569		23.22
		(22.89)	(6.090)		(24.62)
Span School (Not Magnet)		7.123	7.349		3.929
		(4.336)	(6.411)		(4.857)
Constant	429.4***	421.4***	-39,686***	431.0***	410.9***
	(3.344)	(21.68)	(840.7)	(3.627)	(18.40)
Observations	6,515	6,515	6,515	6,515	6,515
R-squared	0.276	0.307	0.485		
Number of groups			2,250	359	359

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

References

Alexander, K., & Alexander, M. D. (2004). *American public school law*. Stamford, Connecticut: Cengage Learning.

August, D., & Hakuta, K. (Eds.). (1997). *Improving schooling for language-minority children: A research agenda* (a report for the National Research Council (NRC)). Washington, DC: National Academy Press. Retrieved April 3, 2009, from http://www.nap.edu/catalog/5286.html

August, D. (2003). *Supporting the development of English literacy in English language learners: Key issues and promising practices* (Reports-Descriptive No. 61). Baltimore, MD: Center for Research on the Education of Students Placed at risk.

August, D., & Shanahan, T. (2006). Introduction and Methodology. In August, D. Shanahan. T.(Eds.), *Developing literacy in second language learners. Report of the National Literacy Panel on Minority-Language Children and Youth* (pp. 1-42). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers

Baker, K. (1987). Comment on Willig's "A meta-analysis of selected studies in the effectiveness of bilingual education." *Review of Educational Research, 57,* 351-362

Baker, K. (1998). Structured English Immersion: Breakthrough in Teaching Limited-English-Proficient Students. *Phi Delta Kappan* 80 (3): 1999-204, November.

Baker, K.,& de Kanter, A. A. (1981). Effectiveness of bilingual education: A review of the literature (Final draft report). Washington, DC: Department of Education, Office of Planning, Budget, and Evaluation.

Baker, K., & De Kanter, A. A. (1983). *Bilingual education: A reappraisal of federal policy*. Free Press.

Baker, K. A., & Pelavin, S. (1984). Problems in bilingual education. In *annual meeting of the American Educational Research Association, New Orleans, LA*.

Bialystok E., (2008). Second-language acquisition and bilingualism at an early age and the impact on early cognitive development. Rev ed. In: Tremblay RE, Barr RG, Peters RDeV, eds. *Encyclopedia on Early Childhood Development*. Montreal, Quebec: Centre of Excellence for Early Childhood Development; 2008:1-4. Available at: http://www.childencyclopedia.com/documents/BialystokANGxp_rev.pdf. Accessed July 2013.

Bialystok E. (2010). Bilingualism. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1, 559–572.

Bialystok E., & Craik F. I. M. (2010). Cognitive and linguistic processing in the bilingual
mind. *Current Directions in Psychological Science*, 19, 19–23.

Bleakley, H., & Chin, A. (2004). Language Skills and Earnings: Evidence from Childhood Immigrants*. *Review of Economics and Statistics*, *86*(2), 481-496.

California Department of Education, California English Language Development Test (CELDT). (2013, August). Retrieved from http://www.cde.ca.gov/ta/tg/el/cefceldt.asp

California Department of Education, Standardized Testing and Reporting (STAR) Results. (2013, September). Retrieved September 5, 2013 from http://star.cde.ca.gov/

California Department of Education, California English Language Development Test. (2013, August). Retrieved September 5, 2013from http://www.cde.ca.gov/ta/tg/el/

California Proposition 63, English is the Official Language Amendment (1986). (25 March 2014). In *Ballotpedia*. Retrieved from http://ballotpedia.org/California_Proposition_63,_English_is_the_Official_Language_Amendment_(1986)

California Proposition 187, Illegal Aliens Ineligible for Public Benefits (1994). (18 March 2014). Retrieved from http://ballotpedia.org/California_Proposition_187,_Illegal_Aliens_Ineligible_for_Public_Benefits_(1994)

California Affirmative Action, Proposition 209 (1996). In *Ballotpedia*. (18 March 2014). Retrieved from http://ballotpedia.org/California_Affirmative_Action,_Proposition_209_(1996)

Collier, V. P., & Thomas, W. P. (2004). The astounding effectiveness of dual language education for all. *NABE Journal of Research and practice*, *2*(1), 1-20.

Crawford, J. (1999). *Bilingual education: History, politics, theory, and practice*. Trenton, NJ: Crane Publishing Company.

Cummins, J. (1981). Age on arrival and immigrant second language learning in Canada: A Reassessment1. *Applied linguistics*, *2*(2), 132-149.

Cummins, J. (2000a). Academic language learning, transformative pedagogy, and information technology: Towards a critical balance. *TESOL Quarterly*, *34*(3), 537-548.

Cummins, J. (2000b). *Language, power, and pedagogy: Bilingual children in the crossfire* (Vol. 23). Multilingual Matters.

Dannof, M.N., Coles, G.J., McLaughlin, D. H., & Reynolds, D.J. 1978. *Evaluation of the impact of ESEA Title VII Spanish/English Bilingual Education Programs. Volume III: Year two impact data, educational process, and in-depth analysis*. Palo Alto, CA: American Institutes for Research.

DeAvila,E., Cervantes,R. & Duncan,S. (1978) Bilingual Exit Criteria. *CABE Research Journal*, *1*(2).

De Cos, P. L. (1999). Educating California's Immigrant Children: An Overview of Bilingual Education. Sacramento, CA: California Research Bureau. Retrieved from ERIC database. (ED434511)

Durgunoglu, A. Y., & Oney, B. (2001). Literacy development in two languages: Cognitive and sociocultural dimensions of cross-language transfer. In *Research symposium on high standards in reading for students from diverse language groups: Research, practice, and policy*.

Gandara, P., & Hopkins, M. (2010). *Forbidden Language: English Learners and Restrictive Language Policies.* NY: Teachers College Press

Genesee, F. (1999). Program alternatives for linguistically diverse students. Santa Cruz, CA: University of California. Center for Research on Education, Diversity & Excellence. Retrived March 3, 2009, from http://www.cal.org.crede/PUBS/edprractice/EPRI.pdf

Genesee, F. (Ed.). (2006). *Educating English language learners: A synthesis of research evidence*. Cambridge University Press.

Gersten, R., & Baker, S. (2000). What we know about effective instructional practices for English-language learners. *Exceptional children*. 66: 454-470.

Gersten, R., & Woodward, J. (1995). A longitudinal study of transitional and immersion bilingual education programs in one district. *The Elementary School Journal*, 223-239.

Goldenberg, C. (2008). Teaching English Language Learners. *AMERicAN EdUcATOR*. Retrieved from: http://scholar.google.com/scholar?cluster=16764778777285497858&hl=en&as_sdt=0,5

Gonzalez, A. (2000). The acquisition and labor market value of four English skills: new evidence from NALS. *Contemporary Economic Policy*, *18*(3), 259-269.

Gordon, N., & Hoxby, C. (2002). Achievement effects of bilingual education versus English immersion: Evidence from California's Proposition 227. *Unpublished paper, University of California, San Diego*.

Gottlieb, M. (2006) *Assessing English Language Learners: Bridges From Language Proficiency to Academic Achievement*. Thousand Oaks, CA: Corwin Press Thousand Oaks, CA: Corwin Press.

Graham, C., & Acosta, S. (1979). Comparison of state-adopted oral language tests. In *National Association for Bilingual Education, 8th Annual Conference, Seattle, Washington*.

Greene, J. P. (1998). *A meta-analysis of the effectiveness of bilingual education*. Claremone, CA: Tomás Rivera Policy Institute.

Hakuta, K., Butler, Y., & Witt, D. (2000). How Long Does It Take English Learners to Attain Proficiency?, University of California Linguistic Minority Research Institute: 1-28.

Hakuta, K. (2011). Educating Language Minority Students and Affirming Their Equal Rights Research and Practical Perspectives. *Educational Researcher*, *40*(4), 163-174.

Howard, E. R., Christian, D., & Genesee, F. (2003). *The development of bilingualism and biliteracy from grade 3 to 5: A summary of findings from the Cal/CREDE study of two-way immersion education.* Santa Cruz, CA, and Washington, DC: Center for Research on Education, Diversity, and Excellence.

Katz, A., Low, P., Stack, J., & Tsang, S. L. (2004). A study of content area assessment for English language learners. *Oakland, CA: ARC Associates*.

Krashen, S. (2005). The acquisition of academic English by children in two-way programs: What does the research say. *NABE Review of Research and Practice*, *3*, 3-19.

Kuhlman, N., & Nadeau, A. (1999). English language development standards: The California model. *The CATESOL Journal*, *11*(1), 143-160.

Jepsen, C. (2009, January). Bilingual education and English proficiency. University of Kentucky Center for Poverty Research Discussion Paper Series, DP2009-01. Retrieved July 2013 from http://www.ukcpr.org/Publications/ DP2009-01.pdf.

Lam, T. C. (1992). Review of practices and problems in the evaluation of bilingual education. *Review of educational Research*, *62*(2), 181-203.

Lindholm-Leary, Kathryn and Fred Genesee. 2010. "Alternative Educational Programs for English Learners." In *Improving Education for English Language Learners: Research-Based Approaches*. Sacramento, CA: California Department of Education.

López, M. G., & Tashakkori, A. (2006). Differential outcomes of two bilingual education programs on English language learners. *Bilingual Research Journal*, *30*(1), 123-145.

Los Angeles Unified School District. (2013). "Language Acquisition Branch, Master Plan for the Education of English Language Learners." Los Angeles Unified School District. Retrieved September 5, 2013, from http://notebook.lausd.net/pls/ptl/docs/PAGE/CA_LAUSD/FLDR_ORGANIZATIONS/FLDR_PLCY_RES_DEV/PAR_DIVISION_MAIN/SISB/SEC_SIS/REF-5945%20SECONDARY%20MP%20SURVEY.PDF

Losen, D. (2010). Challenging limitations: The growing potential for overturning restrictive language policies and ensuring equal education opportunity. In Forbidden Language , 86-101. New York: Teachers College Press.

MacSwan, J., & Pray, L. (2005). Learning English bilingually: Age of onset of exposure and rate of acquisition among English language learners in a bilingual education program. *Bilingual Research Journal*, *29*(3), 653-678.

Martiniello, M. (2009). Linguistic complexity, schematic representations, and differential item functioning for English language learners in math tests. *Educational Assessment*, *14*(3-4), 160-179.

Medina, M., & Escamilla, K. (1992). English acquisition by fluent- and limited-Spanish-proficient Mexican-Americans in a 3-year maintenance bilingual program. *Hispanic Journal of behavioral Sciences 14*(2), 252-67

Meyer, M. M., & Fienberg, S. E. (Eds.). (1992). *Assessing evaluation studies: The case of bilingual education strategies*. National Academies Press.

New York City Board of Education (2010) *EPIC Facilitators' Guide A Manual for the ELL Parent Information Case (EPIC)*. Retrieved July 13, 2013, from http://schools.nyc.gov/NR/rdonlyres/08F60DF2-38EC-432C-8B76-FB643B3D960A/0/EPICFacilitatorsGuide_FINAL2010.pdf

Parrish, T. B., Merickel, A., Perez, M., Linquanti, R., Socias, M., Spain, A., Speroni, C. & Delancey, D. (2006). Effects of the Implementation of Proposition 227 on the Education of English Learners, K-12: Findings from a Five-Year Evaluation. Final Report for AB 56 and AB 1116. *American Institutes for Research*.

Peregoy, S. F., & Boyle, O. F. (2001). Reading, writing, and learning in ESL. New York: Addison Wesley Longman.

Purcell-Gates, V. (2001). Emergent literacy is emerging knowledge of written, not oral, language. *New directions for child and adolescent development*, *2001*(92), 7-22.

Ramírez, J. D., Yuen, S. D., & Ramey, D. R. (1991). *Executive summary, final report: Longitudinal study of structured English immersion strategy, early-exit and late-exit transitional bilingual education programs for language-minority children*. Aguirre International.

Roberts, T., & Neal, H. (2004). Relationships among preschool English language learner's oral proficiency in English, instructional experience and literacy development. *Contemporary Educational Psychology*, *29*(3), 283-311.

Rolstad, K., Mahoney, K. S., & Glass, G. V. (2005a). Weighing the evidence: A meta-analysis of bilingual education in Arizona. *Bilingual Research Journal*, *29*(1), 43-67.

Rolstad, K., Mahoney, K., & Glass, G. V. (2005b). The big picture: A meta-analysis of program effectiveness research on English language learners. *Educational Policy*, *19*(4), 572-594.

Rolstad, K., Mahoney, K ., & Glass, G. (2008). The Big Picture in Bilingual Education: A Mata-analysis Corrected for Gersten's Coding Error. *Journal of Education Research & Policy Studies*, *8*(2) 1-15

Rolstad, K., MacSwan, J., & Mahoney, S. (2012). The Ineffectiveness of English Immersion in Arizona. *International Journal of Language Studies*, *6*(2), 139

Rossell, C. H., & Baker, K. (1996). The educational effectiveness of bilingual education. *Research in the Teaching of English*, 7-74.

Rossell, C. (1990). The effectiveness of educational alternatives for limited-English proficient children. In G. Imoff (Ed.), *Learning in two languages* (pp.71-121). New Brunswik, NJ: Transatlantic Publishers. Learning in two languages (pp. 71-121). New Brunswik, NJ: Transatlantic Publishers.

Rossell, C. (2005). Teaching English through English. *Educational Leadership*, *62*(4), 32-36.

Rossell, C. & Kudar, J. (2005). Meta-murky: A rebuttal to recent meta-analyses of bilingual education. In J. Söhn, Ed., *The effectiveness of bilingual school programs for immigrant children.* Berlin, Germany: Wissenschaftszentrum Berlin für Sozialforschung.

Rumberger, R., & L. Tran. (2010). State language policies, school language practices, and the EL achievement gap. In P. Gádara & M. Hopkins (Eds.) *Forbidden Language* , 86-101. New York: Teachers College Press.

Salazar, J. J. (2007). Evaluation of English Learner Programs 2006-07. *Master Plan Evaluation Report for English Learner Programs*. O. O. S. P. A. Accountability. Los Angeles, Los Angeles Unified School District: 1-35.

Saunders, W. M., & O'Brien, G. (2006). Oral Language. In *Educating English language learners: A synthesis of research evidence*. F. Genesee, K. Lindholm-Leary, W. M. Saunders & D. Christian (Eds.). New York: Cambridge University Press.

Saville-Troike, Muriel. (1984). What <u>really</u> matters in second language learning for academic achievement? *TESOL Quarterly, 18,* (2) 199-219.

Shaftel, J., Belton-Kocher, E., Glasnapp, D., & Poggio, J. (2006). The impact of language characteristics in mathematics test items on the performance of English language learners and students with disabilities. *Educational Assessment*, *11*(2), 105-126.

Slavin, R. E., & Cheung, A. (2005). A synthesis of research on language of reading instruction for English language learners. *Review of Educational Research*, *75*(2), 247-284.

Solano-Flores, G. (2008). Who is given tests in what language by whom, when, and where? The need for probabilistic views of language in the testing of English language learners. *Educational Researcher*, *37*(4), 189-199.

Thomas, W., & Collier, V. (2002) *A national study of school effectiveness for language minority students' long-term academic achievement final report: Project 1.1* Santa Cruz, CA, and Washington, DC: Center for Research on Education, Diversity, and Excellence.

Thomas, W.P., & Collier, V.P. (2004). *A national study of school effectiveness for language minority students' long-term academic achievement.* Washington, DC: U.S. Department of Education, Center for Research on Education, Diversity, & Excellence (CREDE) and the Office of Educational Research and Improvement.

Tong, F., Lara-Alecio, R., Irby, B., Mathes, P., & Kwok, O. M. (2008). Accelerating early academic oral English development in transitional bilingual and structured English immersion programs. *American Educational Research Journal*, *45*(4), 1011-1044.

Tregar, B. & Wong, B. F. (1981). The relationship between native and second language reading comprehension and second language oral ability. Paper presented at the Language Proficiency Assessment Symposium, Airlie House, Virginia. [ERIC Document Reproduction Service No. ED 228-857].

Trejo, S. J. (2003). Intergenerational progress of Mexican-origin workers in the US labor market. *Journal of Human Resources*, *38*(3), 467-489.

Ulibarri, D. M., Spencer, M. L., & Rivas, G. A. (1981). Language proficiency and academic achievement: A study of language proficiency tests and their relationship to school ratings as predictors of academic achievement. *NABE journal*, *5*(3), 47-80.

U.S. Department of Education, Class Size Reduction. (2013, January). Information about federal and state class size reduction (CSR) programs established to improve educational achievement for all students by reducing class size ratios per teacher. Retrieved September 5, 2013 from http://www.cde.ca.gov/ls/cs/

U.S. Department of Education, National Center for Education Statistics. (2013). *The Condition of Education 2013* (NCES 2013-037), English Language Learners. Retrieved from http://nces.ed.gov/fastfacts/display.asp?id=96

Wiese, A. M., & Garcia, E. E. (1998). The Bilingual Education Act: Language Minority Students and Equal Educational Opportunity. *Bilingual Research Journal*, *22*(1), 1-18.

Williams T., Kirst M., & Haertel E. (2005). *Similar students, different results: Why do some schools do better? A large-scale survey of California elementary schools serving low-income students*. Mountain View, CA: EdSource.

Williams T., Hakuta K., & Haertel E, (2007). *Similar English learner students, different results: Why do some schools do better? A follow-up analysis, based on a large-scale survey of California elementary schools serving low-income and EL students*. Mountain View, CA: EdSource.

Willig, A. C. (1985). A meta-analysis of selected studies on the effectiveness of bilingual education. Review of Educational Research, 55(3), 269-318.

Wolf, M. K., & Leon, S. (2009). An investigation of the language demands in content assessments for English language learners. *Educational Assessment*, *14*(3-4), 139-159.