IDENTIFYING RECLASSIFICATION BARRIERS AND POST RECLASSIFICATION GROWTH FOR ENGLISH LANGUAGE LEARNERS

A Thesis
Presented
to the Faculty of
California State University, Chico

In Partial Fulfillment
of the Requirements for the Degree
Master of Arts
in
Psychology
Applied Psychology Option

by
© Clifford Hauenstein 2014
Spring 2014
IDENTIFYING RECLASSIFICATION BARRIERS AND POST RECLASSIFICATION GROWTH FOR ENGLISH LANGUAGE LEARNERS

A Thesis

by

Clifford Hauenstein

Spring 2014

APPROVED BY THE DEAN OF GRADUATE STUDIES AND VICE PROVOST FOR RESEARCH:

______________________________
Eun K. Park, Ph.D.

APPROVED BY THE GRADUATE ADVISORY COMMITTEE:

______________________________
Leesa V. Huang, Ph.D., Chair

______________________________
James N. Wolfe, Ph.D.
PUBLICATION RIGHTS

No portion of this thesis may be reprinted or reproduced in any manner unacceptable to the usual copyright restrictions without the written permission of the author.
DEDICATION

This thesis is dedicated to my parents, who could not have done more to support me through my master's program. I would not be in the position I am without all their help and encouragement.
ACKNOWLEDGEMENTS

Firstly, I would like express much gratitude to Dr. Huang and Dr. Wolfe for offering their guidance, and also being so patient with me throughout this process. Additionally, I cannot overestimate the assistance and contributions of David Philhour, who took a great deal of time to meet with me and fine tune the statistical end of the analysis.

A great deal of thanks is also warranted for the faculty and staff at Orland Unified. This project would not have been completed without all their input, recommendations, and assistance: Luanne Stewart, for taking the time to inform me of all ELL assessment procedures; Tracy Sailsbury, for consulting with me, listening to my ideas, and providing valuable feedback; Joey Brett for supervising me as an intern and being so generous as to allow me to conduct my study in the district; Jennifer Bevers for being so responsive in helping me obtain additional data for the project; and of course, Paula Ikerd and Tammi Kibby for helping me organize my data.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Publication Rights</th>
<th>iii</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dedication</td>
<td>iv</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>v</td>
</tr>
<tr>
<td>List of Tables</td>
<td>viii</td>
</tr>
<tr>
<td>List of Figures</td>
<td>ix</td>
</tr>
<tr>
<td>Table of Contents</td>
<td>vi</td>
</tr>
<tr>
<td>Abstract</td>
<td>x</td>
</tr>
</tbody>
</table>

## CHAPTER

I. Introduction ........................................................................................................ 1

   Background and Theoretical Basis and Orientation ........................................ 1
   Statement of the Problem ............................................................................. 13
   Purpose of the Study .................................................................................. 15
   Limitations of the Study .......................................................................... 16
   Terms and Definitions ................................................................................ 19

II. Literature Review ......................................................................................... 20

   Defining Academic English Proficiency ..................................................... 20
   The Interaction of ELD and Academic Content Standards/Assessments ........ 23
   Post-Reclassification Growth .................................................................... 31

III. Methodology ................................................................................................. 37

   Design of the Investigation ....................................................................... 37
   Population ................................................................................................... 38
   Treatment .................................................................................................... 39
   Data Analysis Procedures .......................................................................... 41
# LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Percent of Sampled Students Reclassified in Each Grade According to District</td>
<td>52</td>
</tr>
<tr>
<td>2. Student CELDT Performance by District of Enrollment</td>
<td>56</td>
</tr>
<tr>
<td>3. Mean Scores for CST ELA Clusters (Aggregate Student Performance)</td>
<td>59</td>
</tr>
<tr>
<td>4. Section Performance Levels According to Reclassification Year</td>
<td>61</td>
</tr>
<tr>
<td>5. A Comparison of the Linear and Logarithmic Models for 2nd and 3rd Grades</td>
<td>67</td>
</tr>
<tr>
<td>6. Model vs Expected CST Scores for Reclassification Threshold</td>
<td>68</td>
</tr>
<tr>
<td>7. Model vs Expected CST Scores for Proficient Threshold</td>
<td>69</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. CELDT section performance, aggregated across two years prior to reclassification</td>
<td>54</td>
</tr>
<tr>
<td>2. CST cluster performance, aggregated across two years prior to reclassification</td>
<td>59</td>
</tr>
<tr>
<td>3. CST cluster performance, two years prior to reclassification</td>
<td>61</td>
</tr>
<tr>
<td>4. CST cluster performance, one year prior to reclassification</td>
<td>62</td>
</tr>
<tr>
<td>5. CST ELA by CELDT performance, second grade</td>
<td>63</td>
</tr>
<tr>
<td>6. CST ELA by CELDT performance, third grade</td>
<td>64</td>
</tr>
<tr>
<td>7. Linear and logarithmic models of CST ELA by CELDT performance, second grade</td>
<td>66</td>
</tr>
<tr>
<td>8. Linear model of CST ELA by CELDT performance, third grade</td>
<td>67</td>
</tr>
<tr>
<td>9. Time series chart for RFEP period: August through November</td>
<td>69</td>
</tr>
<tr>
<td>10. Time series chart for RFEP period: December through March</td>
<td>70</td>
</tr>
<tr>
<td>11. Time series chart for RFEP period: April through June</td>
<td>70</td>
</tr>
<tr>
<td>12. ELL – RFEP performance trend for CST ELA</td>
<td>72</td>
</tr>
<tr>
<td>13. ELL-RFEP performance trends for CST ELA, by grade of RFEP</td>
<td>73</td>
</tr>
<tr>
<td>14. ELL-RFEP performance trends for CST ELA by district</td>
<td>74</td>
</tr>
</tbody>
</table>
ABSTRACT

IDENTIFYING RECLASSIFICATION BARRIERS AND POST RECLASSIFICATION GROWTH FOR ENGLISH LANGUAGE LEARNERS

by

© Clifford Hauenstein 2014

Master of Arts in Psychology

Applied Psychology Option

California State University, Chico

Spring 2014

There is growing recognition that protracted enrollment in English language development curriculums leads to undesirable student outcomes, including drop-out (ostensibly due to lower expectations, diminished access to a challenging curriculum, etc.). There is a significant research base supporting the notion that English Language Development (ELD) curriculums may interfere with the academic growth trajectories of English Language Learner (ELL) students. At the same time, insufficient access to a sheltered, English support curriculum may leave students unprepared for English only instruction. Ostensibly, such a limited and sheltered academic environment is only beneficial if students graduate to a mainstream classroom as soon as is warranted; when English fluency is advanced enough to allow appropriate navigation within an English only academic environment. However, determining when students are ready for such promotion is obfuscated by many factors. Critically, there is no universally accepted theory or definition of academic English proficiency. Thus, there is not a well-established theoretical framework with which to align English proficiency assessments.
Correspondingly, there is growing concern in the literature that the exit criteria for ELL students is not validly distinguishing those capable from those incapable of independently navigating an English only academic environment. Given these concerns, this research project aims to provide an updated review of the validity of current reclassification criteria, and how the different criteria components interact to affect student reclassification rates. Specifically, growth rate changes pre/post reclassification across districts and reclassification grade, the relationship between English proficiency and language achievement measures, and particular areas of difficulty along these assessments for ELL students were examined.

*Keywords:* ELL, reclassification, CELDT, academic achievement
CHAPTER I

INTRODUCTION

Background and Theoretical Basis and Organization

In the former half of the 20\textsuperscript{th} century, a philosophy of segregation and stringent stratification for marginalized groups with intense needs dominated the educational sphere. In fact, this philosophy was pervasive in society, and affected opportunities, for employment and health services. Galvanized by the efforts of advocacy groups and the civil rights movements, federal legislation began to address issues of discrimination in academic settings. Following the landmark decision in Brown vs. The Topeka Kansas Board of Education (1956), educational institutions have progressively adopted a more inclusive philosophy of service delivery, making active efforts to appropriately serve children with a broad array of needs (Hardman, Drew, & Egan, 2010). However, over the next twenty years, it was realized that educational institutions had no formal structures in place to address the particular needs of language minority students. These students faced a unique challenge within the academic setting, as they were charged with the task of acquiring English, while at the same time progressing in academic content areas.

In 1974, the next milestone was achieved in terms of non-discriminatory educational practice, as a Supreme Court ruling mandated that schools take active efforts to address the needs of second language learners (Malakof & Hakuta, 1990). Fueled by the experiences of nearly 1800 native Chinese speaking students who did not receive targeted English development instruction in San Francisco, the Supreme Court cited
Section 601 of the Civil Rights Act of 1964 in their decision (Sugarman & Widess, 1974). Essentially, inaccessibility of targeted, English language development supports represented a discriminatory practice, as these students were not being afforded the same educational opportunities available to their native-English speaking peers. Although this was an important step in ensuring positive outcomes for English language learners, many issues in serving these students remain unresolved after 40 years of practice and research. Additionally, the recent focus on accountability reform, and the major tenets of No Child Left Behind (which mandate high standards for all students, regardless of linguistic origins) have only instigated more debate and contention. With population growth rates for language minority students increasing every year, the magnitude of these issues has been amplified.

Current literature regarding instruction of second language learners reveals several salient points of contention; these include: 1) the most efficient model of English language development instruction, 2) the validity of classification instruments and criteria, 3) access to an appropriately challenging curriculum and necessary school resources, and 4) the operationalization of academic English proficiency. A cursory review of each of these matters will be discussed in order to provide context for the current research project.

**The Most Efficient Model of English Language Development Instruction**

Two models of English language development instruction have dominated the field of education over the past 30 years. Bilingual models advocate for continued development (or maintenance) of the native language while also offering targeted instruction in English. A proportion of academic instruction is made available in the
native language, thus providing access to challenging coursework (Krashen, 2001). Use of the native language for much or part of academic instruction is presumed to reduce the potential for a “watered down” curriculum, which might be necessary if English is overwhelmingly preferred in the academic setting (Wong-Fillmore & Valdez, 1986). Although bilingual education was a preeminent force in the formative years of English language development programs, momentum began to swing towards an English Immersion model in the early 1980’s.

Concerns regarding students’ inability to access employment opportunities were a prevailing concern, providing the impetus for many to reconsider the bilingual models (Lyons, 1990). Specifically, it was argued that students would never be sufficiently challenged to acquire the requisite English skills, and therefore struggle mightily in post-secondary endeavors. Additionally, a number of studies found that, when students were completely immersed in a culture with a novel language, oral proficiency was attained with relative expediency (Rossell & Baker, 1996). Consequently, advocacy for a model wherein students were instructed overwhelmingly in English gained momentum; this model eventually took on the moniker of “structured/sheltered English immersion.”

In 1998, California legislature passed Proposition 227, which formally established structured/sheltered English Immersion as the state’s primary means of serving limited English proficient students. Consistent with the original tenets of the model, students were expected to enter mainstream English classes within one year of being enrolled in the English Language Development program (Jepsen & De Alth, 2005). Generally, students identified as limited English proficient are provided two types of services: 1) English Language Development (ELD) which is devoted to facilitating academic English
proficiency and 2) Specifically Designed Academic Instruction in English (SDAIE), which offers academic instruction overwhelmingly in English, albeit with modifications and supports appropriate to emerging bilinguals (CA Department of Education, 2013). It should be noted that ELL learners are generally separated from their English only peers for these instructional placements.

Unfortunately, since the passage of Proposition 227, the objective of students acquiring basic English skills within one year has proven elusive to most districts. Tracking three cohorts of second grade ELL students, Grissom (2004) found that after four to five years of schooling, only 30% of the students had achieved a relative level of English proficiency to enter mainstream classes. Other studies have also reported a general inability for districts to adhere by the one year objective specified in the law. Salazar (2005) found that in Los Angeles Unified School District, only 9% of ELL students were able to enter mainstream English classes after one year, and only 61% of the sample entered mainstream English classes after six years. Flores (2009) also found similar numbers in a secondary analysis of Los Angeles Unified School District. The overarching goal of proposition 227 was to expedite the process of English language development for the ELL population, yet there has actually been little improvement in the rate of exit to mainstream classes, or improvements on standardized assessments of English language proficiency (Grissom, 2004; Hill, 2006; Jepsen & De Alth, 2005).

These discouraging numbers have led many to argue for a return to bilingual models of English language development; a growing body of research indicates that the bilingual models may lead to more positive outcomes over the sheltered/structured English Immersion models (see Cummins, 2000; Green, 1998; Mortensen, 1984; as a few
Furthermore, because the sheltered/structured English Immersion models require instruction overwhelmingly in the secondary language, there is concern that the breadth and depth of academic content coverage will be compromised. With few opportunities to utilize the native language to access the curriculum, the cognitive resources of ELL individuals may be substantially exhausted just in their attempts to navigate a predominantly English-only environment (Wong-Fillmore & Valdez, 1986).

The Validity of Classification Instruments and Criteria

Federal legislation mandates that educational institutions actively identify students with limited English proficiency, provide appropriate instruction to develop academic English skills, and monitor the progress of English language learners towards academic and linguistic objectives (Jespent & de Alth, 2005) Consistent with the prevailing philosophies of inclusion and education in the “least restrictive environment”, states have established procedures for exiting students from English language Development programs when they demonstrate a basic command of academic English. When an ELL is exited from such programs, supplementary language supports are withdrawn, placement in mainstream English classes is made, and the title of the student is formally changed from ELL to “Reclassified Fluent English Proficient” or “R-FEP” (this process is commonly referred to as “reclassification” in the educational lexicon) (Linquanti, 2001).

As can be anticipated, both the initial and reclassification decisions have monumental implications for students’ academic futures. Placement in a language support program when not actually warranted may stifle a student’s academic growth, ostensibly by shifting instructional focus to unnecessary linguistic development and
compromising the breadth and depth of content instruction. Conversely, placement in an English-only academic setting for a student with substantial language support needs (either at the initial or reclassification checkpoint) may overwhelm the student and effect negative academic outcomes. As such, the predictive validity of the criteria chosen as a basis for these decisions is a critical factor. Generally, a standardized English language proficiency assessment or “ELPA” forms the cornerstone of language proficiency criteria, though other assessments and factors are often considered for these decisions.

Title III of the No Child Left Behind Act (2002) specifically directs educational agencies to develop a set of English proficiency standards, and adopt a standardized assessment of English proficiency aligned to the standards. However, beyond that, states and local educational agencies are given much latitude in specifying exact guidelines and thresholds as the basis of initial identification or reclassification decisions (Jepsen & De Alth, 2005). For initial decisions, California uses a “Home Language Survey” to compliment standardized assessment with an ELPA. Specifically, guardians of all newly enrolled students are asked to complete a survey that solicits information regarding linguistic exposure (Jepsen & De Alth, 2005). As currently implemented, one can obtain data related to the language in which the child first learned to speak, the language the child and the parents most frequently use at home, and which language is used when speaking to the child in the home environment (CA Department of Education, 2005).

Generally, any answer in the affirmative results in designation of language minority status, and the student is required to take the ELPA to determine need for an English language development curriculum. However, concern has been raised that the predictive validity of the Home Language Survey in identifying potential ELL students is
grossly deficient (García Bedolla, & Rodriguez, 2011). Indeed, there is a dearth of research examining the predictive quality of such surveys. One of the items on the California’s Home Language Survey has represented a particular point of contention, as many fear it has the potential to recruit many English proficient students for testing (Littlejohn, 1998). It can be speculated that many students could be, for all intents and purposes, native English speakers with only one relative in the home (aunt, uncle, grandparent) who primarily speaks another language. These students would be identified as linguistic minority students, and subsequently administered the ELPA, even though they may speak proficient English.

This in and of itself would not be problematic, since the Home Language Survey acts only as a screener to trigger further assessment with an ELPA. However, the validity of ELPA’s for making educational placement decisions remains questionable. For example, Pray (2000) found that among 40 native English speakers, no student was able to score in the proficient range on the Woodcock Munoz Language Survey. The IDEA Proficiency Test of English skills proved more valid, though 13% of native English speakers were still identified with limited fluent proficiency. Additionally, studies of the Language Assessment Scales have revealed meaningful variability within one proficiency level, indicating that a more discrete classification system may be warranted (Butler & Castellon-Wellington, 2000).

The validity of the California English Language Development Test or CELDT (the particular ELPA used in California) for educational placement and progress monitoring has been similarly questioned. Consider a study performed by the CELDT publishers in 2005, in which English language development experts rated the English
proficiency levels of 1,384 ELL students. The experts were blind to the students’ previous CELDT performances, and the resultant ratings were compared against pre-existing proficiency levels, determined by CELDT performance. Results revealed only 40% agreement between ratings and CELDT scores.

Additionally, another study commissioned by the state of California evaluated the performance of native English speakers against the pre-existing cut scores for English proficiency on the CELDT. At the kindergarten level, 67% of the English only students scored below Early Advanced for the Listening domain. For the Speaking domain, 42% of the English only students scored below Early Advanced. These results are significant, given that Early Advanced performance is generally established as the criterion for English language proficiency.

Thus, the CELDT’s potential to over-identify ELL students remains a valid concern. When considered alongside the potential for the Home Language Survey to over-identify language minority students, it seems possible that a significant number of students, who could otherwise navigate a mainstream English class without additional supports, are in fact offered an English Language Development curriculum. As will be discussed shortly, issues of educational tracking, inadequate resources, and low expectations for ELL students may create an academic trajectory with unfavorable outcomes; this only amplifies the concerns of over-identification of ELL students.

The same concerns regarding initial ELL identification are also relevant for reclassification decisions. However, California state law requires three additional criteria to be considered during the reclassification process; these include: at least one measure of academic achievement (usually the California Standards Test or CST, a state-mandated,
standardized test administered each Spring to evaluate academic performance), teacher evaluation of curricular progress, and parent consultation (Jepsen & De Alth, 2005). Reclassification to fluent English proficient indicates that the former ELL learners “have exhibited the ability to participate successfully with English-speaking peers in mainstream classes” (p. 1, CA Department of Education, 2007).

Thus, the multiple criteria guard against premature withdrawal of English language supports, and inclusion of an academic criterion ensures that students have developed basic linguistic skills and higher order academic skills before English-only mainstreaming. State law makes the latter criterion explicit in its statement that [Specialized language services be provided to ELLs until they] have recouped any academic deficits which may have been incurred in other areas of the core curriculum as a result of language barriers” (pg.2, CA Department of Education, 2013) However, some argue that these multiple criteria place an undue burden on English language learners, as they impede access to a more challenging, stimulating curriculum in a general education setting (Xiong & Zhou, 2006). For example, in the spring of 2012, 15% of 2nd grade and 17% of 3rd grade English only students scored below proficient on the CST Language Arts section (CA Department of Education, 2012). ELL students scoring at the same level would generally be retained in an English language development curriculum.

Furthermore, in an undisclosed state, Albedi (2008) found that achievement tests accounted for very little of the variance in ELL classification status; an average of only 3.5 % for grades 3, 6, and 8. Analyses in other sites revealed variances of 7.6% and 12%. Contributing to this dilemma are notions of inadequate resources, low expectations, and an overly simplified curriculum for English learners; this ostensibly creates a perennial
cycle of low achievement scores and continued placement in a low performing educational track for these students. The following section will elaborate upon the concerns of curricular tracking for ELL learners.

**Access to an Appropriately Challenging Curriculum and Necessary School Resources**

One of the prevailing concerns of structured/sheltered English immersion models is limited academic content coverage. Previous studies have noted that administrators may interpret ELL status as an indicator necessitating a more basic curriculum. Unfortunately, the curriculum offered may be oversimplified and rudimentary, affecting access to post-secondary academic institutions. In 1994, Harklau completed an ethnographic study of high school ELL learners in northern California; she concluded that the academic tasks within ELL courses predominantly involved decoding and repetition of information. However, access to challenging academic coursework may augment, rather than mitigate, second language development.

In a study of achievement gap differences between ELL students and native English speakers, Wang and Goldschmidt (1999) found that placement in a higher track had a positive, moderating effect on academic outcomes for ELL students. In fact, Callahan (2005) studied the effect of several variables on GPA, credits earned, standardized reading and math scores, and high school exit exams for ELL students. Predictor variables included the number of college preparatory courses enrolled in and the level of English language proficiency. Consistent with the study by Wang and Goldshmidt (1999), educational track served as a better overall predictor for the criterion variables than did English proficiency.
Concerns have also been raised regarding teachers serving ELL students. Aside from lowered teacher expectations potentially interfering with academic progress (Callahan & Gandara 2004; Katz 1999; Schiller & Muller, 2000), reports have been made of less qualified teachers serving ELL classrooms. Albedi (2006) surveyed opportunity to learn variables in 21 school districts in Southern California, and found that teachers with less content knowledge were more frequently assigned to ELD classes. Also, Gandara, Rumberger, Maxwell-Jolly, and Callahan (2003) found that ELL students in California were four times more likely than their general education peers to be taught by an uncredentialled teacher.

Also affecting high school ELL learners, mandatory enrollment in SDAIE and ELD classes may preclude opportunity to complete the courses requisite for college application. Many SDAIE classes may not meet college preparatory requirements; additionally, enrollment in ELD classes may result in scheduling constraints that limit college preparatory or advanced placement registration (Gandara et al., 2003). For example, in Callahan’s (2005) study sample, taken from a rural district in northern California, 98% of ELL students had not taken the requisite college preparatory coursework to apply to either the California State University or University of California systems. In a study of college preparatory enrollment for ELL students in a Central California school district, it was found that only 10-14% of ELL students met University of California graduation requirements, compared to 23-32% of English only students (Xiong & Zhou, 2006).

Together, these factors may create a situation where ELL individuals find it very difficult to be reclassified to fluent English proficient. As previously mentioned, low
academic tracks for ELL students with lowered expectations may instill a perennial cycle of low achievement and continued placement in an ELL curriculum. If ELL learners are not being adequately challenged, they may never make sufficient academic progress to score proficient on standardized tests of achievement. This, in turn, blockades access to a more challenging, general education curriculum. As one administrator noted, “The [ELD] track is a kind of purgatory, a holding place to nowhere” (pg. 12; Gutierrez, 2005).

The Operationalization of Academic English Proficiency

As of now, there is no universally accepted theory or definition of academic English proficiency. Cummins (1980) offered a useful and widely received framework for defining language development and levels of language demand. Briefly, he distinguished between a more informal, context laden discourse (Basic Interpersonal Communication Skills or BICS) and more formal, context independent discourse (Cognitive Academic Language Proficiency Skills or CALP). While his terms have become fixtures in ELD research and practice, other frameworks exist and are differentially adopted in the development of English proficiency assessments.

Many traditional assessments have used different theories and definitions of academic English proficiency to guide item construction, obscuring a universal profile of fluent English proficient students. Variability in the perception of English language proficiency is also reflected in the English Proficiency standards that states have established. As of February 2012, 18-20 different sets of English proficiency standards existed across the 50 states and Washington D.C. Much ambiguity appears present in the determination of language abilities required for success in English only classrooms. For example, in Albedi’s (2008) review of current ELL assessment procedures, it was
questioned whether, “ELP assessment should be focused on the language of the content areas (such as mathematics and science) or the language that facilitates content learning” (p. 20).

Further obfuscating the question of what language skills are required for access to an English only curriculum, many states now mandate that ELL learners demonstrate adequate performance on proficiency assessments and academic achievement measures of language skills before reclassification. These mandates lead to a series of additional questions, including: How do the constructs measured between the assessments differ? What language skills are common to both measures? Can ELPA performance levels serve as indicators of readiness for standardized achievement assessment in English? How do these measures differentially serve as gatekeepers to mainstream English-only instruction? A small body of research has revealed that ELPA’s tend to focus more on basic linguistic skills (grammar, phonology), stress comprehension at the literal level, and contain less academic vocabulary. Language achievement assessments, conversely, focus on higher order, more complex linguistic functions and comprehension at the inferential level.

**Statement of the Problem**

Taken together, this host of issues points to the fact that current English proficiency assessment systems may over-identify ELL students; that once in an ELD curriculum, there is potential for lowered expectations and a “watered down” curriculum; and that this may encumber academic progress and result in low achievement scores, thus blockading access to the general education curriculum. This would create a perennial cycle of low achievement and continued placement in an ELD curriculum. Limited
understanding of the nature of language proficiency required for successful learning in an English only classroom imparts more controversy.

In light of these issues, it appears that two fundamental conclusions can be drawn. First, the earlier a student is able to become English proficient and given access to a challenging curriculum, the more positive the outcome. ELL outcome data generally support this claim. For example, ELL students at the elementary level are much more likely to be retained than their English-only counterparts, and retention has been shown to correlate with drop-out. Long term ELL learners (generally those who retain ELL status for six years or more) generally exhibit anemic academic performance (GPA < 2.0) (Olsen, 2010). Salazar (2005) also found that high school ELL learners are 25% more likely to repeat ninth grade, and only 3% of ELL students who repeated ninth grade were able to graduate. Students with ELL backgrounds have also shown declining performance on the California High School Exit Exam from 2004 to 2007, with only 34% and 26% passing the math and language portions, respectively.

The second conclusion that can be drawn from the background research is that the predictive validity of reclassification criteria remains equivocal. Different perspectives of academic English proficiency and questions of appropriate cut scores for standardized measures create a reclassification process that is mired in ambiguity. Although it has been established that earlier reclassification is desired, it should be noted that the argument is not made that ELL status, in and of itself, is debilitating. ELD supports may be vital during a student’s initial efforts to acquire and utilize novel language structures. However, it appears that protracted ELL status, or retention in an ELD curriculum beyond what is necessary to access English only instruction, may affect
negative outcomes. Thus, developing a reclassification assessment protocol that exits students from ELD services not too early, and not too late, is imperative. Given that there are still many questions as to how to define and operationalize academic English proficiency, it is critical that educators continually review how language minority students perform post-reclassification, and that administrators reconsider reclassification policy accordingly.

**Purpose of the Study**

In addressing the first conclusion cited above (that earlier attainment of English proficiency and prompt access to the English only curriculum is desired), it behooves administrators and educators to understand the particular roadblocks to reclassification. Every year, ELL students in California are administered tests of English proficiency (CELDT) and English academic skills (CST) as part of the reclassification process. Review and inspection of student performance on these assessments can provide a wealth of information regarding areas of relative struggle. Because each of these assessments organizes performance across several language subdomains (Listening, Speaking, Reading, and Writing for the CELDT; and Word Analysis, Reading Comprehension, Literary Analysis, Written Conventions and Writing Strategies for the CST), the opportunity is available to obtain a more nuanced understanding of the barriers to reclassification. Knowledge of how these two assessments interact to limit the probability of reclassification, and which particular subdomains ELL students struggle with most, can inform future intervention and instructional efforts. Thus, the current research project evaluates student ELL performance data across six grade spans to determine which assessments (CELDT or CST) and which particular subdomains emerge as glaring
obstacles to reclassification. Such information may assist in future intervention instructional efforts for ELL students.

In addressing the second conclusion (the lack of established predictive validity for reclassification criteria), Robinson (2011) presented an analytic model for determining the appropriateness of reclassification cut scores, which we adopt in the current study. Based on the logic of the model, the ideal scenario for reclassification should be a smooth trend line of academic growth through the reclassification period. If there is a significant negative change in academic growth after reclassification, it is likely that current criteria for reclassification are too liberal, and that language supports are being prematurely withdrawn. Conversely, a significant, positive change in academic growth post-reclassification may suggest that cut-scores are too stringent; students could have been given earlier access to the English only curriculum and still realized positive academic outcomes. Reclassification represents a significant event for ELL learners and influences their subsequent academic trajectory. As such, the current research project aims to evaluate the extent to which current reclassification cut scores are appropriately established, and the extent to which they indicate readiness to access an English only curriculum. Results from this analysis will assist educators and administrators in modifying reclassification criteria in order to best serve the ELL population.

**Limitations to the Study**

Before proceeding, it is important to note a few limitations that constrain the application of findings. Firstly, the ELL population has been widely recognized as a diverse and heterogeneous population. Pooling students by language proficiency status ignores critical elements of variability that effect each student’s academic growth in...
idiosyncratic ways. Unfortunately, the sample size available was not sufficiently large to allow regression along such important variables as SES, parent education, and differential language exposure in the home. Furthermore, our sample was overwhelmingly composed of students whose native language was Spanish. The findings emerging from this study may have limited application to the reclassification process for students of other native languages. Another concern with respect to sample composition is our exclusive study of ELL students from grades 2-7. The literature is clear regarding negative outcomes for long-term ELL students (Olsen, 2010), so the objective of this research project was to assist educational agencies in appropriately serving and reclassifying ELL students before they enter high school. However, this comes at the expense of understanding how long-term ELL students respond to English proficiency assessment and reclassification at the secondary level.

There are a few methodological issues that warrant attention as well. For the analysis of changes in academic growth post-reclassification, only one measure was used as a proxy for achievement level (CST ELA scores). Aside from reliability concerns with only one outcome measure for language arts performance, math performance pre/post reclassification was not evaluated in the current study. This was primarily due to the fact that by 7th grade, there is stratification of math placement. Thus, math performance for seventh grade and beyond would have to be disaggregated by course enrollment, and separate analyses performed. The sample available was not of sufficient size to support such disaggregated analysis. Also, because the CST scales are not vertically aligned, growth could not be evaluated simply with the obtained scaled scores. Rather, ELL performance had to be defined as the scaled score discrepancy in standard deviation units
from native English speakers. This is a normative evaluation, and thus year to year changes in the performance of native English speakers will affect analyses for the ELL sample. For example, a rapid, disproportionate increase in the scores of native English speakers would result in a negative observed growth trend for ELL students.
Definitions of Terms

1. English Language Learners (ELL) are students who often come from primarily non-English-speaking homes and background. These individuals cannot communicate fluently in English which in turn creates difficulties learning effectively in an English only academic setting. ELL students typically need modified or specialized teaching in both the English language and in their academic content areas.

2. English Language Proficiency Assessment (ELPA) is a category of instruments utilized to identify, document progress in English language acquisition, and determine when proficiency in English has been established.

3. California English Language Development Test (CELDT) is the ELPA specific to the state of California that is used for educational placement and progress monitoring. There are four sections: Listening, Speaking, Reading, and Writing for the CELDT.

4. California Standards Test (CST) is a state-mandated, standardized test administered each Spring to evaluate academic performance. Though there are multiple sections of the CST, only the English Language Arts portion (Word Analysis, Reading Comprehension, Literary Analysis, Written Conventions and Writing Strategies) was utilized for this research study.

5. Reclassified Fluent English Proficient (R-FEP) is a process commonly referred to as “reclassification” in the educational lexicon.
CHAPTER II

REVIEW OF THE LITERATURE

As previously established, the current research is focused on how language proficiency and standardized achievement assessments intersect and impact access to the English only curriculum for ELL students in California. Thus, to begin this section, a review of the definitions of academic English proficiency is provided. Afterwards, an analysis of the standards and assessments ELD and academic content offered which includes a discussion of the impacts of both topics. Finally, the literature on post-reclassification growth is explored.

Defining Academic English Proficiency

Central to reclassification assessment (and therefore, the specific research objectives) is the question of what particular English language skills are necessary to access the English only curriculum. In a review of the second language acquisition process and the development of an ELL instructional framework, Chamot and O’Malley (1994) offered the following definition of academic language proficiency: “the language that is used by teachers and students for the purpose of acquiring new knowledge and skills . . . imparting new information, describing abstract ideas, and developing students conceptual understanding” (p. 40). Clearly, any assessment efforts for reclassification purposes need to consider the minimal language demands necessary for success in an English only environment. Many efforts have been made to develop more nuanced and
discrete conceptualizations of English language proficiency, leading to much variation and inconsistency in ELL assessment. For example, some language proficiency frameworks have specified up to 64 linguistic domains; others have argued for a single-order proficiency factor to explain the majority of performance variance (Cummins, 1980). The lack of a universally accepted construct of academic English proficiency has obfuscated much of the reclassification process.

Nonetheless, most attempts at defining academic English proficiency can be classified according to two fundamental approaches. The first approach can be traced to the theoretical framework promoted by Cummins (1980), who distinguished “Basic Interpersonal Communication Skills” from “Cognitive Academic Language Proficiency”. Cummins noted that previous research failed to account for critical English acquisition processes for second language learners, namely the varied linguistic demands of different settings. It was Cummins’ objective to develop a framework which could take into account how different settings, with disparate linguistic climates, differentially affected language development. To this end, Cummins’ model was predicated upon two orthogonal continua. The first defines language demands in terms of cognitive complexity; or the volume of information that must be simultaneously or sequentially processed in a given task. The second continuum defines language in terms of contextual dependence. Context dependent language shows greater reliance for paralinguistic elements and situational cues in communication. Conversely, context independent language is more verbose, and requires greater use of linguistic devices to accurately delineate a concept or idea.
Basic Interpersonal Communicative Skills or BICS refers to performance for context dependence, cognitively simple language tasks. Cognitive Academic Language Proficiency, or CALP, on the other hand, refers to competence in more context independent, cognitively complex language functions. Generally, formal academic discourse falls within the CALP domain, while BICS is largely represented in informal, social communicative exchanges. This distinction was important, as it underscored the need for explicit instruction in context-independent and cognitively demanding discourse. Reflecting Cummins’ model, a series of early case studies revealed the efficiency of second language acquisition for young children immersed within a novel linguistic climate. Conversant proficiency was often obtained within a few months to a year. However, the language was invariably noted to be rudimentary, highly dependent on social context, and lacking advanced linguistic structures. From Cummins’ model, BICS were observed to develop relatively quickly and naturally for language minority students, though CALP took a greater amount of time and explicit instruction to be realized.

The second approach to defining academic language has generally evaluated academic language in terms of discrete language functions, or how linguistic elements are organized to accomplish a given communicative task at the discourse level. Such language functions include analyzing, comparing, predicting, persuading, solving problems, and evaluating. In the same vein, Halliday (1978) described academic language as being organized among different “registers”; each register being idiosyncratic to a particular academic discipline. Register can be thought of as a linguistic style or language aesthetic that has been socially constructed for specific contexts and tasks; registers can be differentiated from one another at the lexical and syntactical levels. At the syntactical
level, there are highly stylized elements of English grammar that vary across different domains and tasks. For example, mathematical reasoning problems often contain a higher frequency of comparatives (greater/less than), logical connectors, and passive voice (Spanos, Rhodes, & Dale, 1988). Conversely, analysis of classroom history instruction reveals a dominance of active voice, simple past, historical present, and temporal/causative signal syntax types (Short, 1994). At the lexical level, Stevens, Butler, and Castellon-Wellington (2001) elucidated three principle word categories: high-frequency general words (words that are commonplace among informal speech patterns); nonspecialized academic words (words that are exclusive to academic domains, yet cut across various disciplines); and specialized content-words (words that are exclusive to particular academic disciplines).

**The Interaction of ELD and Academic Content Standards/Assessments**

These models and evaluative frameworks have shaped the development of English proficiency standards for schools; this has been in response to federal legislation requiring educational agencies to delineate a set of language criterion objectives for ELLs. Subsequently, English language proficiency assessments are to be aligned with these standards. California developed its first set of ELD standards in 1999, based on input from 15 language experts and comprehensive review of extant research. As currently envisioned, the ELD standards in California are to represent, “pathways” to proficiency in the general education language arts standards (CA Department of Education, 1999). The intention is to circumvent a sequential model of ELL instruction, where grade-level curriculum standards become the focus only after basic academic English proficiency has been attained. Proficiency levels along both sets of standards are
coordinated such that advanced performance along ELD standards is expected to correlate with proficient performance along ELA standards. Elucidation of how ELD standards and ELA standards interact to impact ELL assessment has been attempted in previous research, and these efforts provide an important empirical basis for the current project. It should be noted that a portion of the following studies evaluated relationships between ELD standards/assessments and academic achievement standards/assessments in several domains (Reading, Math, Science, Social Studies). Others focused solely upon the relationship between ELD standards/assessments and language arts standards/assessments. The latter set of studies is more relevant to reclassification processes and the present investigation, since achievement assessments in non-language domains (i.e., math, science, social studies) do not directly inform reclassification decisions. However, both types of studies will be reviewed, as both provide critical information regarding how the language from ELPAs correlates with the academic language demands in the general education setting.

In 2006, the State of California commissioned CTB/McGraw Hill to evaluate the extent to which ELD standards were linked to content standards (including ELA, math, and science) and how well the CELDT was aligned to the ELD standards. Specifically, commonalities between the ELD and CST standards, and alignment between the CELDT and ELD standards, were evaluated along three dimensions: modality (i.e., Listening, Speaking, Reading, and Writing); language demand; and complexity. Language demands at both the linguistic level (i.e., phonology, morphemic analysis, vocabulary words) and functional level (i.e., definition, explanation, inquiry) were included in the analysis. Each language demand was subsequently given a complexity rating of either low, medium, or
high. These same dimensions were also applied to the analysis of alignment between the ELD standards and CELDT.

Findings indicated that the ELD standards placed a premium on linguistic skills versus higher order academic language functions. In turn, the content standards were dominated by a greater proportion of broad language tasks, including labeling/enumeration, explanation, and analysis/synthesis. Additionally, the standards were found to be weakly linked for complexity across all grades. At grade two, for example, half of the ELD standards were at the low complexity level, while 33% of the content standards were at the low complexity level. A greater discrepancy emerged for grade five, where 53% of the ELD standards were of low complexity, while only 5.2% of the content standards were at low complexity. Analysis of the degree of alignment between the CELDT items and ELD standards revealed that the Listening and Writing sections showed the strongest mismatch, both being poorly aligned for complexity and modality. Additionally, a disproportionate number of CELDT items were written at the beginning proficiency level, indicating that the instrument may not discriminate well at the higher ELD proficiency levels. For example, in the 6-8 and 9-12 grade spans, only one item per test represented the advanced proficiency level. Overall, then, there appears a trend for the ELD standards and assessments to address lower complexity and more basic linguistic functions in comparison to the academic content standards. On top of this, CELDT items often did not represent the full range of proficiency levels as reflected in the ELD standards. Thus, students scoring at higher levels on the CELDT may not actually reach proficiency on content assessments. The predictive validity of the CELDT may be stronger at low proficiency levels, where more items are represented.
Studies of other language proficiency and content assessments have generally established similar relationships, while also revealing a few novel patterns. For example, in 2000, Stevens et al. analyzed the Language Assessment Scales-Reading (LAS-R) and the Iowa Tests of Basic Skills-Social Studies (ITBS) for commonalities among distinct language features and student performance levels. As with other studies, language comparisons were made at the lexical (general vocabulary, academic-general vocabulary, and discipline-specific vocabulary); grammatical, and discourse levels. Performance ranges along the two assessments were also investigated, and quantitative performance correlations were calculated.

Generally, the ITBS contained more specialized academic vocabulary, more complex grammatical structures, more demanding language functions, and more varied discourse. The Pearson Product-Moment Correlation for the total LAS-R and ITBS scores was .4482 (p<.001), thus accounting for 20% of the variance in performance levels. This would indicate that as students’ language proficiency increases, they are in a better position to access the language of the content assessments. Splitting ELL learners into a high, middle, and low group based upon LAS-R scores and comparing performance levels on the ITBS reinforced this claim.

The highest performing ELL students on the LAS-R exhibited a response pattern similar to English only students on the ITBS. Even when items were marked incorrectly, the distracters chosen usually reflected an underlying logic in the approach to the task. Thus, it was evident that these ELL students were adequately deciphering the language of the test and had a clear understanding of the nature of each item. Conversely, the lowest performing ELL students exhibited response patterns that had no clear underlying logic
or rationale. If anything, some students adopted a strategy of avoiding distracters with unfamiliar vocabulary; others preferentially selected or avoided the longest distracters. In this manner, tests of language proficiency may have the potential to serve as “readiness” assessments for academic content assessment. In other words, student performance on an ELPA may help educators determine the extent of language confounds on standardized academic assessments. In fact, Kim and Herman (2010) found a significant, positive quadratic correlation between ELPAs and academic reading assessment across three states. The quadratic relationship may indicate a particular, baseline level of language proficiency, below which improvements in English proficiency have little effect on academic reading performance.

Also in 2005, Butler and Castellon-Wellington evaluated the performance level relationships between the LAS Reading/Writing assessments and the Stanford 9 Reading and Language assessments. Unique in this study was availability of LAS Reading and Writing scores for English only students (who are not normally administered tests of English proficiency). The findings were consistent with previous indications that ELPA are more sensitive at lower performance levels. For example, there was much variability in Stanford 9 Reading performance for students scoring in the competent reader range for the LAS-Reading. The average Stanford 9 Normal Curve Equivalent (NCE) score for English only students at the competent level was 58.5 (just above the national average), while the average NCE for ELLs at the competent level was 35.0 (below the national average). However, the LAS Reading performances appeared to correlate more strongly with Stanford 9 Reading performances at lower proficiency levels. For English only and ELL students scoring in the Limited Reader range on the LAS-Reading, there was only a
five point average NCE difference between the two groups on the Stanford 9 Reading assessment. Similar results were found for the relationship between LAS-Reading and Stanford 9 Language performance levels. It should be noted that the LAS Writing subtest was more difficult for students of all language proficiencies, and appeared to correlate with a more limited performance range on the Stanford 9 assessments.

Overall, however, quantitative performance correlations between ELPAs and language arts achievement assessments have remained moderate to moderately high. For example, in Butler and Castellon-Wellington study (2005), Pearson Product Moment Correlations for the LAS and Stanford 9 assessments were significant and positive; LAS reading accounted for 31% and 51% of the variance in Stanford 9 Language and Reading scores, respectively. The LAS Writing accounted for 14% to 25% of the variance in Stanford 9 Reading and Language scores, respectively. In 2009, Parker, Louie, and O’Dwyer examined overall performance correlations between Utah’s language proficiency and achievement assessments. Again, significant, positive correlations were found. The state ELPA was able to account for 23-30% of the variance in Reading and Writing achievement assessments, after controlling for relevant student and school characteristics. Similar relationships were identified for ELPAs and academic content assessments in Minnesota (Kato, Albus, Liu, Guven, & Thurlow, 2004) and Colorado (Mahon, 2006).

Taken together, these studies reveal consistent trends in how language proficiency assessments and standardized achievement assessments might interact to affect the reclassification process. Principally, although most states attempt to coordinate ELD and academic content standards such that performance levels are consistent across
assessments, this may not always be the case. ELD standards may not be adequately linked to content standards; this would create situations wherein English proficient students (as determined by ELPAs) may not always reach expected benchmark levels on Reading and language content assessments. Indeed, language demands for other ELPAs show much less complexity and elaboration than content assessments. Additionally, there is evidence that the CELDT has lacked alignment with ELD standards, predominately by under-representing higher proficiency level items. Thus, there may be a wide range of performance levels on language content assessments for students at advanced levels on the CELDT; some of these students may score well above state benchmark levels, while others may score well below. In sum, although significant, positive correlations have been established between ELPAs and content assessments; it is likely that content assessments (rather than English proficiency assessments) represent the major obstacles to reclassification.

Within California, this generally seems to be the case. Salazar (2005) found that in the 2003-04 school year, 52% of ELLs in Los Angeles Unified School District scored proficient on the CELDT (early advanced or advanced), while only 10% scored proficient on the CST ELA assessment. Looking at statewide data, from 2001 to 2005, the percentage of students scoring at or above the CELDT proficiency threshold has increased. In the 2001-02 school year, 25% of ELLs scored at early advanced or advanced; over the next three academic school years, percentages progressively increased to 34%, 43% and 48%. However, the number of ELLs reclassified each year remained fairly stable, fluctuating between 7.7 and 9.0%. A much larger percentage of ELLs are scoring proficient on the CELDT than are being reclassified. While other factors
influence reclassification (teacher input and parent consultation), it is likely that a substantial number of these students are failing to meet the reclassification threshold on the CST ELA assessment.

While it is seems well established that academic content assessments represent a greater encumbrance to reclassification than ELPAs, there is a dearth of research evaluating how individual clusters and subdomains affect reclassification. However, a study by Robinson (2011) provides an initial look at how individual CELDT sections impact reclassification for ELL learners. As part of a study evaluating the validity of reclassification cut scores, Robinson analyzed the manner in which the CELDT and CST differentially serve as “gatekeepers” to reclassification. Specifically, Robinson used a binding score technique, wherein a student’s minimum earned score among six reclassification test variables (CELDT overall, CELDT Listening, CELDT Speaking, CELDT Reading, CELDT Writing, and CST ELA) was standardized and re-centered around the respective cut score. This allowed for determination of how far above or below a reclassification cut score a student performed, in standard deviation units.

Results for a large, urban California school district revealed that between 6th and 7th grade, the proportion of ELLs failing to meet the CST ELA cut score increased substantially. For grades 4 through 10, 17-25% of ELLs failed to meet the CST ELA reclassification cut score, while 39-46% failed to meet the CST cut score in grades 7 through 10. Conversely, the CELDT Reading section represented the “gatekeeper test” at the younger grades, and progressively had less and less influence as grade level increased. The CELDT Listening, Speaking, and Writing tests maintained a stable influence on reclassification; between 16 and 26% of students failed to meet
reclassification thresholds for these sections across all years. However, while Robinson was able to provide initial insight into how the different CELDT sections uniquely impact reclassification efforts, individual CST ELA clusters were overlooked. Elucidating the relative difficulty of these clusters, and how cluster performance interacts with the individual CELDT sections, can provide a more discrete understanding of ELL language deficits. Thus, the hope then is for this present study to expand upon Robinson’s initial findings by incorporating analysis of both the individual CST ELA clusters and the individual CELDT sections.

**Post-Reclassification Growth**

After elucidating the distinct impact of ELPA and content test subdomains on reclassification assessment, a critical follow up question pertains to the validity of pre-existing reclassification cut scores. As described in the previous sections, presently there is no universally accepted definition of academic English proficiency. Dozens of different ELD standards exist across the country, and different frameworks of English proficiency have informed development of the multitude of available ELPAs. Additionally, concerns have been raised regarding the validity of current CELDT cut scores, as well as the requirement of proficient performance on an academic content assessment for reclassification. Given the equivocal nature of reclassification assessment, it is imperative that educational agencies continually validate the reclassification criteria they have employed. If criteria are too liberal, language supports may be prematurely withdrawn, subjecting students to an unnecessarily challenging academic environment. Conversely, reclassification criteria that are too stringent may retain students in an ELD curriculum.
longer than needed, thereby prolonging access to more stimulating and complex subject matter.

Although there is a significant body of research examining the performance levels of reclassified students in relation to ELLs and English only students, much variability exists in the findings. Ostensibly, this is primarily due to differing reclassification criteria across and within states. For example, Kim and Herman (2010) analyzed achievement gaps for students of differing language proficiency status across three states. Two of the states differed in the ELPA cut scores required for reclassification. Students reclassified in the state with more lenient criteria performed substantially below their English only counterparts upon exiting the ELD program. Conversely, students exiting the ELD curriculum in the state with more stringent criteria performed above their English Only counterparts. However, the authors observed that two years after being reclassified, most students tended to catch up to their English only peers, regardless of the state they were reclassified in.

In California, data have indicated declining performance for reclassified student groups as a function of grade. Linquanti (2001) reported SAT-9 reading data across the state of California in 1999, stratified by language proficiency status. While a higher percentage of reclassified students scored above the 50th national percentile than English only students in grades 2 through 4, a rapid performance decline was observed between grades 4 and 7 for that year. In grades 6 through 11, a substantially smaller percentage of reclassified students met the 50th national percentile, compared to their English only counterparts. In tracking a single California cohort of students beginning second grade in

Declining performance trends for reclassified students have occurred in more recent years as well. In 2006, Xiong and Zhou discovered 69% of reclassified students scored proficient or above on the CST ELA assessment in second grade. This percentage progressively dropped across grades for that year, with only 36% of eleventh grade reclassified students scoring proficient or above. In 2012, 88% of 2nd grade reclassified students scored at proficient or above, while 49% of 11th grade students were able to meet the proficiency cut score (CA Dept of Education, 2012). It should be noted, however, that the percentage of reclassified students scoring proficient or above in second grade was substantially higher than the percentage of English only students scoring proficient or above in second grade. By eleventh grade, however, the percentage of reclassified students scoring proficient or above had dropped below that of the English only group.

Although this data points to the consistent declining performance of reclassified students as a function of grade, it is important to note that the composition of the reclassified student pool is continually changing. Every year, former ELL students who meet the reclassification thresholds are added to the reclassified student group, thus introducing another variable to consider when analyzing performance trends. It is not unrealistic to assume that this pool is initially dominated by the most academically precocious language minority students. Each year, a new sample of ELL students is reclassified and joins this group, progressively defining the reclassified pool to be more representative of the greater language minority community. Thus, decreasing scores for reclassified student groups may primarily stem from a continually changing subject pool,
rather than individual student performance declines. Additionally, such broad
descriptions of reclassified student performance fail to consider the heterogeneity of the
ELL group, as well as individual student growth patterns.

Several studies have applied more refined analyses in order to control for these
interfering variables. Flores (2009) examined six academic outcomes (8th grade SAT9
Reading and Math scores, 9th grade retention, AP enrollment, CAHSEE performance, and
high school dropout) for students of different language proficiency status in Los Angeles
Unified School District. Reclassified students were stratified by grade of reclassification,
and sixth grade achievement scores were controlled for. Controlling for individual
achievement levels at sixth grade provided a purer measure of reclassification effects.
Overall, reclassified students performed similarly to English only students across all
outcome measures. Reclassified students scored just below English only students on the
SAT9 measures in eighth grade, but were also less likely to drop out, more likely to pass
the CAHSEE, and more likely to take an AP course than English only students. Notably,
those students who were not reclassified by eighth grade performed substantially worse
across these outcome measures. When sixth grade achievement was accounted for, the
positive effect of reclassification was expectedly diminished, though it was still
significant. Reclassification before 4th grade predicted greater achievement levels in 8th
grade, as compared to those students reclassified in 5th grade. Overall, the authors were
able to conclude that reclassification, with the pre-existing criteria, generally led to
favorable outcomes. However, for those students reclassified late in school (sometime
after eighth grade) more academic struggles and challenges were observed.
Robinson (2011) introduced a novel, binding-score regression discontinuity design to evaluate the validity of established cut scores in a large, urban California school district. Specifically, Robinson focused on the probability of reclassification, given both assessment and student variables (i.e., gender, race, special education status, free and reduced lunch status). By controlling for the effect of student level variables on reclassification probability, Robinson was able to obtain a less confounding estimate of the validity of reclassification cut scores. Additionally, only students who performed just above or below the reclassification criteria were pooled for the study. This also helped mitigate any individual variables that might confound the analysis of reclassification impact. For example, the assumption was made that the profile of students scoring just below the reclassification threshold would not differ appreciably from the profile of students scoring just above the reclassification threshold. In the initial analysis, Robinson conducted an ordinary least squares regression for CST ELA performance following reclassification. A comprehensive list of covariates was included in the analysis, including gender, special education status, and free and reduced lunch status. Robinson also controlled for prior achievement levels, similar to Flores (2009). Results were similar to those of Flores, revealing that reclassified students scored significantly better than ELL students; this suggested a positive impact of reclassification.

However, when Robinson (2011) utilized his unique analysis as previously described, a smooth transition was observed across the reclassification period. On the CST ELA assessment, those reclassified did not perform substantially better or worse than their peers who just failed to meet reclassification criteria. According to Robinson’s logic, the lack of appreciable gains or losses post-reclassification would indicate
appropriately established reclassification thresholds. Matsudaira (2005) employed a similar strategy in a large, urban school district and obtained comparable results.

Overall, there appears to be much variability across and within states in terms of the impact of reclassification. However, within California, a few consistent trends have emerged. The reclassified student pool seems to exhibit declining academic performance levels with grade; this may partially be due to group composition changes. Each year, an additional set of language minority students are deemed fluent English proficient, and are subsequently added to the reclassified student pool. It is also worth noting that language minority students represent a heterogeneous group, with a wide range of social support, language exposure, and baseline achievement levels. Thus, more fine-tuned analyses are required to better understand the unique impact of reclassification for ELL students. Matsudaira (2005), Flores (2009), and Robinson (2011) offered more refined investigations, which generally found positive to null effects of reclassification. The objective is to compliment this research base by offering a similarly discrete level of analysis of reclassification impact.
CHAPTER III

METHODOLOGY

Design of the Investigation

The current research project primarily concerns how current language assessments impact the reclassification process. Specifically, this study examined how English Language Performance Assessments (ELPAs) and standardized achievement assessments differentially blockade access to the English only curriculum for ELL students. Additionally, how students recently reclassified to fluent English proficient respond to mainstream instruction was also analyzed. Implicit in this second question is an evaluation of the validity of reclassification criteria. These objectives can be summarized in four explicit research questions:

1) Which test domain of the California English Language Development Test (CELDT) do ELL students experience the most difficulty with: Listening, Speaking, Reading, or Writing?

2) Which cluster of the standardized English Language Arts (ELA) assessment do ELL students experience the most difficulty with: Word Analysis and Vocabulary, Reading Comprehension, Literary Response and Analysis, Written Conventions, or Writing Strategies/Applications?

3) What is the relationship between the English Language Development (ELD) standards and ELA standards, as reflected in CELDT and California Standards
Tests (CST) scores? How does proficiency on the CELDT relate to performance on the ELA?

4) Is there a significant difference in ELA growth rates before and after reclassification?

Multiple statistical techniques were performed to examine and analyze the research questions including: mixed design analysis of variance, scatterplots and Loess fit lines, and exploratory regression.

**Population**

In the current effort to answer each of these questions, CELDT and CST student performance data was obtained from two rural school districts in Northern California. The initial data included a total of 190 subjects which were comprised of 131 subjects collected from District 1 and for 59 subjects from District 2. However, due to missing data for CELDT administrations, our sample size was reduced to a total of 154 subjects (97 from District 1, and 57 from District 2). Specifically, the subject pool included current 10th, 11th, and 12th grade language minority students who were reclassified by either 4th, 5th, 6th, or 7th grade. The focus of the study remained on students reclassified by junior high, as the literature is clear regarding negative outcomes for long-term ELL students. The primary objective is to illuminate reclassification phenomena during the time students are still in a position to realize substantial benefit from this promotion. Three cohorts of students were utilized in order to improve the sample size.
Treatment

Procedure

Prior to the study, approvals were obtained from two rural school districts in California and the CSU, Chico Human Subject Review Board. Data was extracted from the available district electronic records, which included a student’s gender, current grade, reclassification grade, CELDT and CST ELA full scaled scores, CELDT individual section scaled scores, and CST ELA individual cluster raw scores. All subjects were assigned a random code, and data was exported and reorganized into a series of electronic templates appropriate to the nature of this study.

Instruments

The study analyzed data from two instruments: 1) California English Language Development Test (CELDT) and 2) California Standards Tests (CST).

California English Language Development Test (CELDT). The CELDT is a measure determines how well a student can listen, speak, read, and write in English. California state law requires that the CELDT be annually to those individuals whose primary language is other than English or ELLs. Its purpose is threefold: identification of those needing ELD services, to monitor progress in acquiring English, and to aid in the decision making process of deciding when a student is fully proficient in academic English (CA Dept. of Education, 2012). The CELDT includes four individual sections (Listening, Speaking, Reading, and Writing) all of which are weighted equally towards the overall score for grades 2 through 12. Raw scores are converted to scaled scores for each section, using a calibration procedure based on item response models. Scaled scores are then organized into five performance ranges; Beginning, Early Intermediate,
Intermediate, Early Advanced, and Advanced. Although an Intermediate score on any individual section does not immediately disqualify a student from attaining overall English proficiency status, scores in the Early Advanced range are generally considered the primary objective for ELL’s being reclassified. To determine a student’s overall CELDT score, a simple average is taken of all individual section scaled scores.

**California Standards Tests (CST).** All California students enrolled in grades 2 through 11 are tested annually in various academic content areas. Grades 2 through 8 tests cover mathematics and English/language arts (CA Department of Education, 2002ab). Only the ELA section was utilized for the purposes of this research study. For the CST ELA, five to six individual clusters comprise the overall score, depending on grade of administration. These clusters include Word Analysis and Vocabulary Development, Reading Comprehension, Literary Response and Analysis, Written Conventions, Writing Strategies, and Writing Applications (administered in grades four and seven, only). Using item response models, performance levels are equated to a reference form based on 2002 data, and raw scores are transformed to a pre-existing scale ranging from 150-600 for an overall score. Similar to the CELDT, scaled scores are demarcated by five descriptive performance levels, including Far Below Basic, Below Basic, Basic, Proficient, and Advanced (Educational Testing Service, 2004).

In contrast to the CELDT, though, individual cluster performance for the CST is only reported at the raw score level, and less guidance is offered in establishing criterion levels of performance. Technical reports do offer statewide cluster averages for students scoring just at the minimally Proficient or minimally Advanced levels; this provides some indication of expected cluster performances for students scoring within the upper
proficiency ranges. However, most districts establish reclassification cut scores around the Basic performance level, rendering the reported expected cluster performance levels less relevant for the purposes of the current research. Thus, expected cluster performance levels for students just meeting the minimal overall ELA score for reclassification were identified through evaluation of the test structure itself (as opposed to a normative basis). Each cluster is given differing emphasis in terms of its contribution to the overall scaled score. In the second grade ELA test form, for example, Word Analysis and Vocabulary contributed 22 raw score points, roughly 33% of the overall scaled score. In contrast, Literary Response and Analysis contributed only six points, representing only about 9% of the overall score. These relative cluster weightings were re-applied to the specific, minimum overall raw score required for reclassification. This resulted in a new set of expected raw scores for each cluster, which then formed the basis for the present evaluation of ELL performance strengths and weaknesses.

Data Analysis Procedures

Research Questions 1 and 2

In addressing the first two research questions, the primary interest was in how performance distances from benchmark levels differed among the individual CELDT sections and CST clusters. Unfortunately, specific benchmark levels for individual section and cluster performances were not established by the test-makers. In other words, while students must earn a particular overall score on the CELDT and CST ELA for reclassification purposes, the range of necessary total raw score points is free to vary
among the individual sections\(^1\). Without established benchmark scores for the individual sections and clusters, there was no available value to judge student performance levels against. Thus, a set of benchmark performance levels were developed for each CELDT section and CST ELA cluster, based on the unique relationship of each section and cluster to the overall score. The steps taken to establish such benchmark levels will be delineated for each assessment (CELDT and CST), as a clear understanding of how individual scores profiles were evaluated is critical for appropriate interpretation of the study results.

To determine the individual section benchmark scores, the scaled score ranges just surrounding the Early Advanced proficiency level were evaluated for each section within each grade. The combination of scores within this range that averaged to the minimal overall CELDT reclassification criterion were subsequently identified. These identified individual section scores represented the benchmark levels against which subject performances were evaluated.

Having established benchmark reclassification scores for the CELDT and CST subdomains, it was then possible to examine for relative performance differences among these subdomains. Additionally, there was an opportunity to investigate whether these differences remained stable across districts and years of administration.

\(^1\) *Note, however, that one stipulation is made for CELDT proficiency; students cannot earn a score lower than Intermediate on any individual section, even if the total score is above the overall cut score.*
A mixed design analysis of variance was utilized and applied via the IBM software program, Statistical Package for the Social Sciences (SPSS version 20). This allowed for ascertainment of performance differences across the various subdomains, and whether these differences were stable across district of enrollment and year of test administration. With this design, the individual clusters and year of test administration represented the within subjects factors, while district of enrollment represented the between subject factor. Thus, variance in subdomain performance within each assessment (CELDT and CST) could be partitioned by the main effects of: 1) the differing nature of language tasks for each subdomain, 2) individual maturation factors and alternate test forms across subsequent years, and 3) instruction in different academic environments. However, the primary concern was with the main effect of the particular subdomain in affecting performance levels, and how differing instructional environments and assessment across subsequent years interacted with this variable.

While reported scaled score differences represented the metric of analysis for individual CELDT sections, cluster performances for the CST ELA assessment were converted to percentages since score reports only include raw scores for the individual clusters. Specifically, the ratio of earned raw scores for each cluster to expected cluster performance for reclassification was computed and entered into the mixed design analysis of variance. Thus, for the CST ELA assessment, student cluster performance was analyzed as percentage earned of the criterion raw score for reclassification.

Contrasts among the individual levels of the within subjects variables were carried out via paired samples t-tests, with Bonferroni adjustments. Use of paired-samples t-tests circumvented the complications sphericity might impose for the individual
contrast tests. For the omnibus F tests, violations of sphericity were addressed with the Huynh-Feldt adjustment.

**Research Question 3**

The main area of interest was with the particular relationship between ELD and CST ELA standards, within the context of reclassification. The effort has been made to link ELD standards to ELA standards such that the former represent, “pathways” to the latter. The objective is avoid a sequential model of ELD/ELA instruction, whereby students begin to master mainstream ELA standards only after attaining full proficiency in the ELL standards. If standards have been linked appropriately, gains along CELDT standards should correspond to simultaneous and proportional gains along ELA standards. If, however, ELD standards are linked to mainstream ELA standards in a “sequential” manner, (thus generally representing prerequisites to accessing mainstream English language arts instruction), then growth along ELD standards would not lead to consistent gains along ELA standards. In this case, a clear, simultaneous growth pattern along both sets of standards might only be observed at lower language levels. As students progress towards higher mastery of ELD standards, progress toward ELA standards may not follow in a concordant manner. Essentially, a non-linear function may dominate this latter case, whereas a linear function may better represent the former case. As for the form of non-linear relationship for the “sequential” case, a logarithmic function may well characterize an initial, steep growth relationship between the two standards. Additionally, a logarithmic function may capture the subsequent flattening of the relationship, as higher levels of ELD mastery offer less predictive value for ELA mastery.
The CELDT and CST ELA assessments have been specifically aligned to the ELD and ELA standards, respectively. Thus, the relationship between scores on these assessments is a direct reflection of the overall relationship between the two sets of standards, and whether the ELD standards do indeed represent a “pathway” to proficiency along ELA standards. As such, both qualitative (visual inspection of scatterplots and Loess lines) and quantitative (linear and non-linear regression) methods were used to determine the function that best captures the nature of the relationship between CELDT and CST ELA scores. It should be noted that, because the CST assessments are not vertically equated, separate analyses were carried out for individual grade levels. Specifically, CELDT-CST score relationships were analyzed for the sample of 2nd and 3rd grade students.

For research question 3, which specifically addressed the quantitative relationship between CELDT and CST ELA performance, scatterplots and Loess fit lines were initially analyzed for deviations for linearity. For significant deviations from linearity, follow-up, exploratory regression procedures were performed. Specifically, hierarchical regression analysis was performed, with the inclusion of the linear term in the first block and the logarithmic term and polynomial terms in subsequent blocks.

**Research Question 4**

The changes in academic growth post-reclassification were examined and used as an indicator of the validity of reclassification criteria. Adopting Robinson’s (2011) logic, a smooth transition through the reclassification period would suggest appropriately established reclassification cut scores. If cut scores are overly stringent, then a rapid increase in academic growth would be observed post-reclassification, as students were
being retained in a curriculum not commensurate with their academic potential. In this case, academic development may be encumbered through an inappropriately simplified curriculum. Conversely, premature withdrawal of language supports (via overly lenient reclassification criteria) may result in a drop in academic growth. Students in this latter case would be better served with continued placement in an English language development curriculum.

To quantify academic growth rates, a student’s overall CST ELA score was used as an indicator of pre and post reclassification academic language skills. Because the CST assessments are not vertically equated, scaled scores for subsequent grades cannot be validly compared. Thus, analysis of performance changes over subsequent years was on a normative basis; ELL performance levels were defined in relation to statewide English only performance levels. Specifically, for each grade level and year, ELL scaled scores were translated to standard deviation units from the English only mean. Average changes in standard deviation units from the English only mean were then calculated over the three year period before reclassification, as well as the three year period after reclassification. This resulted in two distinct growth rates for each ELL student; pre reclassification growth and post reclassification growth.

Because the student data entered included year of reclassification and district of enrollment, an opportunity emerged to evaluate how changes in academic growth rates differed along these variables. Evaluating growth rate changes across years of reclassification is important, as a different profile of student remains an ELL for seven years, compared to those reclassified at the earliest opportunity. For this analysis, the differences in growth rate changes were evaluated across students reclassified in 5th, 6th,
and 7th grade. Also notable to these additional analyses is the different reclassification criteria employed by the two sampled districts. As mentioned, District 1 utilized slightly more rigorous performance thresholds (CST ELA scaled score of 330, between Basic and Proficient), while District 2 employed a more lenient cut score (CST ELA scaled score of 300, minimally Basic). Comparing differences in growth rate changes across the two districts allows for a possible determination of whether more or less rigorous cut scores represent more valid indicators of readiness for reclassification.

At this point, it is important to note a caveat in the design to address this fourth question. Students were not reclassified at the same point in the school year. In fact, there was great variability in terms of when reclassification decisions are made, and when former ELL students are promoted to the mainstream curriculum. Data obtained from the two districts for this study indicate that ELL students are reclassified as early as August and as late as June; this constitutes a range of nearly 10 months. This means that the end of the year CST assessments were administered to students with vastly different experiences in the mainstream curriculum. Some will have been enrolled in the mainstream curriculum for nearly eight months, whereas others might have only been enrolled in the mainstream curriculum one week prior to taking the CST ELA. Additionally, those students reclassified in late April, May, or June will not have had any experience in the mainstream curriculum before taking the CST ELA assessment (the CST is administered early to mid-April every school year). Obviously, it would be inappropriate to disregard month of reclassification when assessing changes in CST ELA performance.
To address this issue, CST ELA performance trends (across several academic years) were analyzed for students reclassified at various points within their year of reclassification. Essentially, the goal was to determine which CST change interval served as the most appropriate proxy for reclassification effects. That is, how many months does it take before the effects of reclassification manifest in CST ELA performance? For example, it is obvious that the effects of reclassification for students entering mainstream English classes in April, May, or June will only be detected in the following year’s CST ELA scores. The CST ELA is administered early to mid-April, so any changes in academic programming for these students will not be observed until the following administration of the CST (roughly one year later). However, one might argue that the effect of academic promotion for students reclassified in March would also not be detected until the following year’s CST ELA. That is, perhaps one month of mainstream academic programming would have only negligible effects on CST ELA scores for the current academic year. Similarly, perhaps two months of mainstream English instruction also offers little change in scores for that year’s CST ELA. Thus, to determine which administration of the CST ELA offers the most valid indication of reclassification effects, qualitative analyses of CST ELA growth trends were carried out, with students disaggregated by month of reclassification.

A mixed design analysis of variance was again applied to the data set via SPSS 20 software. The dependent variable in this analysis was change in CST ELA performance, represented in standard deviations from the English Only mean. The within subjects factor included program status for an individual student (ELD or mainstream curriculum enrollment) while the between subjects factors included district of enrollment and grade
of reclassification (5th, 6th, or 7th). Subsequently, variance in CST ELA growth rates could be partitioned out by three main effects: 1) differing coursework and levels of language support between ELD and mainstream curriculums, 2) instruction in differing academic environments, and 3) length of time as an ELL student. However, of particular interest was the manner in which the between subjects variables interacted with the within subjects variable. Because the primary concern was the change in achievement growth rates pre and post reclassification, analyses which collapsed growth rates along language proficiency status were less relevant to the current study. Thus, the focus was on the overall main effect of language proficiency status on CST ELA growth rates, and how this effect interacts with district of enrollment and reclassification grade.

Additionally, it was of critical importance to examine the effect of month of reclassification on CST ELA performance. It would be inappropriate to evaluate CST ELA change scores across a single time interval for students who varied greatly in their date of reclassification. To address this issue, time series graphs were constructed to reveal trends in CST ELA performance across five administrations (from two years prior to reclassification to two years post reclassification). Importantly, separate time series graphs were plotted for groups of students reclassified at different points within a single grade level. For example, for students reclassified in 5th grade, time series charts included mean CST ELA scores for efforts in 3rd, 4th, 5th, 6th, and 7th grade. However, separate time series charts were constructed for three groups of students: 1) those reclassified in August through November, 2) those reclassified in December through March, and 3) those reclassified in April through June. These visuals provided insight into the relationship between time of reclassification within a given academic year and CST ELA
performance that same year. Furthermore, these charts provided some indication as to whether it was necessary to use different CST ELA change intervals in the analysis, dependent upon the month of student reclassification.
CHAPTER IV

RESULTS AND DISCUSSION

Presentation of Findings

This research study is attempting to investigate how language proficiency assessments and standard achievement impact the reclassification process. Furthermore, the validity of reclassification criteria was evaluated through the progress of students who were recently reclassified to English only curriculum. Four research questions were generated:

1) Which test domain of the California English Language Development Test (CELDT) do ELL students experience the most difficulty with: Listening, Speaking, Reading, or Writing?

2) Which cluster of the standardized English Language Arts (ELA) assessment do ELL students experience the most difficulty with: Word Analysis and Vocabulary, Reading Comprehension, Literary Response and Analysis, Written Conventions, or Writing Strategies/Applications?

3) What is the relationship between the English Language Development (ELD) standards and ELA standards, as reflected in CELDT and California Standards Tests (CST) scores? How does proficiency on the CELDT relate to performance on the ELA?

4) Is there a significant difference in ELA growth rates before and after reclassification?
As previously mentioned, the two districts sampled differed in the reclassification criteria implemented. Specifically, District 1 required a CST ELA score of 330 for reclassification, while District 2 required a CST ELA score of 300. Both districts, however, established the same CELDT cut scores for reclassification (an overall score of Early Advanced or Advanced, with no individual section score lower than Intermediate). Given that the discrepancy in reclassification criteria may impart significant differences in ELL performance profiles and post-reclassification growth, a more discrete understanding of the reclassification process is obtained by considering district of enrollment in the analyses. In fact, the effect of differing cut scores and academic environments is immediately recognized in the reclassification rate discrepancy between the two districts. The general trend is that students in District 2 (with the more lenient reclassification criteria) are reclassified earlier than students in District 1. Table 1 reveals the percent of sampled students reclassified as fluent English proficient (RFEP) in each grade between the two districts.

Table 1

Percent Of Sampled Students Reclassified In Each Grade According To District

<table>
<thead>
<tr>
<th>District</th>
<th>%RFEP in 4&lt;sup&gt;th&lt;/sup&gt;</th>
<th>%RFEP in 5&lt;sup&gt;th&lt;/sup&gt;</th>
<th>%RFEP in 6&lt;sup&gt;th&lt;/sup&gt;</th>
<th>%RFEP in 7&lt;sup&gt;th&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>31</td>
<td>17</td>
<td>32</td>
<td>20</td>
</tr>
<tr>
<td>2</td>
<td>13</td>
<td>25</td>
<td>19</td>
<td>43</td>
</tr>
</tbody>
</table>

As can be seen, roughly 50% of the students pooled from District 2 were reclassified by sixth grade. Conversely, 80% of the sample pooled from District 1 was reclassified by sixth grade. With these critical differences in mind, the results section will be organized according to each of the specific research questions posed; for each
question, the effect of the main variable of interest will be reported, with subsequent
disclosure of the manner in which it interacts with district of enrollment.

**Research Question 1**

This research question sought to determine which test domain of the CELDT do
ELL students experience the most difficulty with: Listening, Speaking, Reading, or
Writing. This question was addressed through a mixed design analysis of variance, with
CELDT section performance and year of administration as the within subject variables,
and district of enrollment as the between subject variable.

Results from the analysis revealed a significant differentiation of performance
across CELDT sections, when data was collapsed across year of administration and
district of enrollment, $F(2.3,370.66) = 29.85, p < .001, n^2 = .164$. Specifically, Reading
tasks emerged as the most difficult for ELL students, while Speaking tasks were
accomplished with the most relative ease. For section scores collapsed across the two
years prior to reclassification, ELLs on average scored above the reclassification
threshold for Listening ($M = 9.55, SE = 3.68$) and Speaking ($M = 17.19, SE = 4.22$).
Conversely, students tended to score below the reclassification threshold for the Reading
($M = -20.08, SE = 2.64$) and Writing ($M = -5.90, SE = 2.81$) sections. These performance
trends are depicted below in Figure 1.
Figure 1. CELDT section performance, aggregated across two years prior to reclassification.

As can be seen, there is a clear performance discrepancy between the oral language domains (Listening and Speaking) and the text-based language domains (Reading and Writing). A value of zero indicates performance right at the reclassification threshold for any given section. Mean Listening and Speaking scores fall above the reclassification threshold, while mean Reading and Writing scores fall below the reclassification threshold. Furthermore, the standard error bars indicate that the observed performance discrepancies between the two language domains (oral versus text-based) lie outside the margin of error. These initial findings were confirmed with paired t-tests (with Bonferroni corrections applied). Averaged across both years prior to reclassification, performance levels were significantly higher for the Listening and Speaking sections than either the Reading or Writing sections (all p’s <.001). Of
additional note, students scored significantly higher on the Reading section than the Writing section.

Examining the main effect of a one year CELDT re-administration on performance levels, there was a significant increase in scores after a year of additional academic preparation, $F (1, 152) = 27.71, p < .001, \eta^2 = .154$. This would be expected, as the purpose of ELD instruction is to facilitate progression towards mastery of ELD standards, which are embedded within the CELDT language tasks. However, of interest to this study is the observed non-significant interaction between CELDT re-administration and differentiation of individual section performance levels, $F (2.50, 379.62) = .715, p = .519$. Thus, while scores increase significantly across subsequent years, the relative profile of scores itself remains stable. In other words, there is proportional increase across all four sections after an additional year of English language development. Also, it should be noted here that Mauchly’s test indicated a violation of the assumption of sphericity for all within-subject tests and interactions. As such, the degrees of freedom for the analyses were corrected using the Huynh-Feldt estimate.

District of enrollment had a significant effect on CELDT performance levels, $F (1, 152) = 14.18, p < .001, \eta^2 = .085$. Across all CELDT sections, students from District 1 (which implemented the more stringent ELA CST reclassification criteria) outscored students from District 2. This trend is depicted in the following table, which disaggregates student CELDT performance by district of enrollment. The values displayed indicate the average difference from a student’s obtained scaled score to the reclassification threshold score. Negative values reflect underperformance, while positive values indicate performance above and beyond the threshold.
Table 2

Student CELDT Performance by District of Enrollment

<table>
<thead>
<tr>
<th>CELDT Section</th>
<th>District 1</th>
<th>District 2</th>
<th>Mean Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SE</td>
<td>Mean</td>
</tr>
<tr>
<td>Listening</td>
<td>17.31</td>
<td>4.51</td>
<td>-5.71</td>
</tr>
<tr>
<td>Speaking</td>
<td>27.53</td>
<td>4.95</td>
<td>-3.15</td>
</tr>
<tr>
<td>Reading</td>
<td>-14.00</td>
<td>3.32</td>
<td>-32.01</td>
</tr>
<tr>
<td>Writing</td>
<td>-3.58</td>
<td>3.42</td>
<td>-10.45</td>
</tr>
</tbody>
</table>

As can be seen, students in District 1 were already meeting reclassification criteria for the Listening and Speaking sections across the two years prior to reclassification. In contrast, students from District 2 generally failed to meet any of the reclassification thresholds during the two years prior to reclassification. For the Reading and Writing sections, students from District 1 performed under the reclassification thresholds, but were closer to attaining proficiency than students from District 2.

In addition, the interaction term for the district of enrollment and individual CELDT section was verging on significance. With the Huynh-Feldt correction, the F statistic was just shy of significance at the .05 level, $F(2.53, 357.66) = 2.72, p < .058, n^2 = .018$. With the standard analysis (no correction for violations of sphericity) the F statistic was significant. Thus, there is possibility of a true shift in the CELDT score profile between the two districts. In other words, although District 1 students performed significantly better on the CELDT sections than District 2 students, the increase may not be equally proportional across all sections. Because of the approaching significance of the term, contrasts were performed to compare each of the CELDT section performance levels within the two districts to identify pattern differences. Again, paired t-tests with Bonferonni corrections were used for the contrasts. For District 1, CELDT performance
patterns perfectly matched the trends for the main analysis; that is, Listening and Speaking scores were significantly greater than either Reading or Writing scores (all p’s < .001). Additionally, students scored significantly higher on the Writing section than the Reading section.

However, for District 2, discrepancies in performance between the Listening/Speaking sections and the Writing section were no longer significant. Ostensibly, students in District 2 (with the more lenient reclassification criteria) exhibited substantially lower performance levels for the Listening and Speaking sections, relative to District 1 (differences of 22.71 and 30.68, respectively). However, the discrepancy in writing scores was not as marked between the two districts (only 6.87 scaled score units lower for District 2). Thus, the loss of significance for the Listening - Writing and Speaking - Writing contrasts in District 2 seemed to be primarily a function of changes in the Listening/Speaking section performance levels. The table above depicts the varying degrees of section performance differences between the two districts.

**Research Question 2**

This question attempted to analyze which cluster of the standardized English language arts (ELA) assessment do ELL students experience the most difficulty with: Word Analysis and Vocabulary, Reading Comprehension, Literary Response and Analysis, Written Conventions, or Writing Strategies/Applications. To address this specific research question, a repeated measures analysis of variance was employed. An analysis of district level effects was excluded, as different CST ELA cut scores were implemented by the two districts. Because the criteria used to directly calculate the dependent variable differed between the two districts, any significant terms for district
level effects would have primarily resulted from the divergent dependent variable specifications.

As with the significant differentiation of CELDT section performance levels, the overall analysis of CST ELA clusters also revealed a significant differentiation of performance, $F(4,620) = 25.42, \ p < .001, \eta^2 = .141$. Pairwise contrasts among the individual CST clusters (via paired samples t-tests with Bonferroni corrections) revealed that students experienced the most difficulty meeting the reclassification benchmark for Reading Comprehension and Writing Strategies. The data was collapsed across both districts as well as the two years prior to reclassification. Results indicated that students scored significantly lower on Reading Comprehension and Writing Strategies than Written and Oral Conventions, Literacy Response and Analysis, and Word Analysis and Vocabulary (all $p$’s < .001). The performance difference between the Reading Comprehension and Writing Strategies clusters was not significant. Figure 2 and Table 3 summarize the cluster performance patterns for ELL students, collapsed across district of enrollment and years of administration. The particular values presented represent the percent of raw score points earned for each cluster, relative to the expected benchmark level.
Figure 2. CST cluster performance, aggregated across two years prior to reclassification.

Table 3

Mean Scores for CST ELA Clusters (Aggregate Student Performance)

<table>
<thead>
<tr>
<th>ELA Cluster</th>
<th>Mean</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word Analysis and Vocabulary</td>
<td>1.12</td>
<td>.017</td>
</tr>
<tr>
<td>Reading Comprehension</td>
<td>1.04</td>
<td>.018</td>
</tr>
<tr>
<td>Literacy Response and Analysis</td>
<td>1.15</td>
<td>.020</td>
</tr>
<tr>
<td>Written and Oral Conventions</td>
<td>1.17</td>
<td>.017</td>
</tr>
<tr>
<td>Writing Strategies</td>
<td>.996</td>
<td>.019</td>
</tr>
</tbody>
</table>
As can be seen, although students generally exhibited significantly lower performance on the Reading Comprehension and Writing Strategies clusters, the observed scores were still near or above the reclassification benchmark (104% and 99.6% respectively). Thus, at least when collapsing data across district and year of administration, the CST seems to present less of a reclassification hurdle than the CELDT.

There was also a significant improvement in scores after repeated administration of the CST ELA one year later, F (1.155) = 42.74, p < .001, n² = .216. Across all clusters, students generally increased their performance by 12.7 percentage points in subsequent years. Similar to the analysis with CELDT scores, there was also a significant interaction between subsequent ELA administrations and cluster differentiation, F(3.87, 599.5) = 6.92, p < .001, n² = .043. Among the individual clusters, there was markedly less growth observed between years for the Literary Response/Analysis and Writing Strategies clusters. Significantly less growth was observed for these clusters relative to the Word Analysis/Vocabulary and Written/Oral Convention clusters. Table 4 presents cluster performance levels (again as a percentage of items correct relative to the reclassification benchmark). Difference values marked with a double asterisk indicate significant improvement at p<.005. Figures 3 and 4 summarizes cluster performance patterns for a given test administration year.
Table 4

Section Performance Levels According To Reclassification Year

<table>
<thead>
<tr>
<th>CST Section</th>
<th>2 Yrs Prior to Reclassification</th>
<th>1 Yr Prior to Reclassification</th>
<th>Mean Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SE</td>
<td>Mean</td>
</tr>
<tr>
<td>Word Analysis/Vocab</td>
<td>1.02</td>
<td>.026</td>
<td>1.20</td>
</tr>
<tr>
<td>Reading Comprehension</td>
<td>.986</td>
<td>.029</td>
<td>1.10</td>
</tr>
<tr>
<td>Literary Response/Analysis</td>
<td>1.13</td>
<td>.030</td>
<td>1.16</td>
</tr>
<tr>
<td>Written Conventions</td>
<td>1.06</td>
<td>.024</td>
<td>1.29</td>
</tr>
<tr>
<td>Writing Strategies</td>
<td>.950</td>
<td>.027</td>
<td>1.02</td>
</tr>
</tbody>
</table>

Figure 3. CST cluster performance, two years prior to reclassification.
Figure 4. CST cluster performance, one year prior to reclassification.

Research Question 3

This question sought to clarify the relationship between the ELD standards and ELA standards, as reflected in CELDT and CST scores. Additionally, the research question was designed to determine whether proficiency on the CELDT is related to performance on the ELA Scatterplots for overall CELDT and CST ELA scores were initially analyzed to determine the degree of departure from linearity. Scatterplots were augmented with Loess fit lines to facilitate interpretation. Provided below are the plots, separated by grade of administration (Loess lines utilize a smoothing parameter of .60).
Figure 5. CST ELA by CELDT performance, second grade.
Both scatterplots generally show positive relationships between CELDT and CST ELA scores. However, a linear component seems more pronounced for the third grade administration than the second grade administration. Furthermore, the Loess line for the second grade administration indicates the possible presence of a non-linear relationship between the two assessments. As can be seen, improvement on the CELDT from 400 to 500 scaled score units generally corresponds to proportional improvement in CST scores. However, the relationship flattens when scores exceed 500 on the CELDT. Conversely, for the 3rd grade administration, the linear trend seems to be preserved throughout the range of data.

Subsequently, regression analyses were performed with each of the two data sets to evaluate the strength and form of the relationships. For the second grade data set,
polynomial and logarithmic regression analyses were performed in addition to linear regression to capture the curve identified in the scatterplot.

The linear model for the second grade data set was significant, $F(1,144) = 16.758, p<.001$, accounting for 10.4% of the variance in CST ELA scores. The logarithmic model offered significant improvement over the linear model, $F(1,143) = 5.715, p = .018$, and accounted for an additional 3.4% of the variance in CST ELA scores. Thus, the logarithmic model offered utility by capturing the attenuated relationship for CELDT scores beyond 500. However, the incremental utility beyond the simple linear model was modest. The quadratic term was unable to offer significant improvement in prediction beyond the linear and logarithmic models, indicating relatively little to no reversal in the relationship at the highest CELDT scores.

As might be expected, the linear model offered stronger prediction for the third grade data set. The model was significant, $F(1,147) = 38.024, p<.001$, and accounted for 20.7% of the variance in CST ELA scores; double that observed for the second grade data set. The scatterplot and Loess line failed to indicate any substantial non-linear component to the relationship; thus, additional regression analyses were not carried out. Scatterplots of the two data sets are provided below, this time with the appropriate regression functions overlaid (second grade includes both linear and logarithmic, while third grade includes only the linear).

As can be seen in Figures 7 and 8, the difference in the fitted linear versus logarithmic function is almost negligible within the context of the full data set. That is, the bend in the fitted logarithmic function is so slight that it seems to approximate the linear model of best fit.
Figure 7. Linear and logarithmic models of CST ELA by CELDT performance, second grade.
Figure 8. Linear model of CST ELA by CELDT performance, third grade.

Following, the linear and logarithmic regression models are provided. These models offer some initial insight with regards to how reclassification thresholds on the CELDT correspond to reclassification thresholds on the CST ELA assessment.

Table 5

A Comparison Of The Linear And Logarithmic Models For 2\textsuperscript{nd} And 3\textsuperscript{rd} Grades

<table>
<thead>
<tr>
<th>Grade</th>
<th>Linear Model</th>
<th>Logarithmic Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>2\textsuperscript{nd} Grade</td>
<td>CST = 148.772 + .311 (CELDT)</td>
<td>CST = -717.729 + 164.471*Ln(CELDT)</td>
</tr>
<tr>
<td>3\textsuperscript{rd} Grade</td>
<td>CST = 10.65 + .584 (CELDT)</td>
<td></td>
</tr>
</tbody>
</table>

The following tables compare the CST scores predicted from regression models with those that might expected given the specified relationship between CELDT and CST. Table 6 uses the regression models to predict CST scores at the reclassification
CELDT score. The information contained within the first table essentially addresses the question, “do students scoring at the reclassification threshold on the CELDT perform above or below the reclassification threshold on the CST?” Although districts vary in the CST ELA cut scores used, many utilize the basic category (cut score of 300) as a minimum for reclassification; thus, a CST score of 300 serves as the value for comparison. Results show that students scoring at the reclassification threshold for the CELDT score 12 to 20 points above 300 on the CST ELA. The difference is significant relative to the standard error of the predicted value.

Table 6

*Model vs Expected CST Scores for Reclassification Threshold*

<table>
<thead>
<tr>
<th>Grade</th>
<th>Predicted Value from Model</th>
<th>Expected Value</th>
<th>Difference</th>
<th>Standard Error of Predicted Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>2nd Grade</td>
<td>312</td>
<td>300</td>
<td>12</td>
<td>2.80</td>
</tr>
<tr>
<td>3rd Grade</td>
<td>320</td>
<td>300</td>
<td>20</td>
<td>3.08</td>
</tr>
</tbody>
</table>

Table 7 uses the regression models to predict CST scores at the minimally Advanced level on the CELDT. The ELD and ELA standards are linked such that Advanced performance on the CELDT should correspond to minimally proficient (cut score of 350) performance on the CST ELA. Results indicate substantial departure from this expectation at the second grade level; however, the relationship seems to hold at the third grade level.
Table 7

*Model vs Expected CST Scores for Proficient Threshold*

<table>
<thead>
<tr>
<th>Grade</th>
<th>Predicted Value from Model</th>
<th>Expected Value</th>
<th>Difference</th>
<th>Standard Error of Predicted Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>2\textsuperscript{nd} Grade</td>
<td>324</td>
<td>350</td>
<td>26</td>
<td>4.79</td>
</tr>
<tr>
<td>3\textsuperscript{rd} Grade</td>
<td>343</td>
<td>350</td>
<td>7</td>
<td>5.64</td>
</tr>
</tbody>
</table>

**Research Question 4**

The final research question attempted to determine if there was a significant difference in ELA growth rates before and after reclassification. The first step in addressing this question was determining the effect month of reclassification had on subsequent CST ELA scores. For an initial exploration of this question, time series charts of CST ELA performance were constructed, disaggregated by interval of reclassification. Figures 9 through 11 are the time series charts for the three RFEP periods.

*Figure 9. Time series chart for RFEP period: August through November.*
Figure 10. Time series chart for RFEP period: December through March.

Figure 11. Time series chart for RFEP period: April through June.
These charts indicate that students exhibit a similar adjustment pattern to reclassification, regardless of the month they were enrolled in an English only curriculum. In other words, students reclassified as early August and as late as March all show decrements in performance for the CST ELA administered that same academic year. As would be expected, those reclassified in April or later in a given academic year (after the CST ELA has been administered) do not show performance decrements until the following year. As such, the examination of changes in growth rates pre and post reclassification moves forward with students reclassified from August through March combined in a single analysis. The interval of interest for CST change scores lies between one year pre-reclassification and the year of reclassification for this group of students (in other words, the interval directly preceding reclassification).

Students reclassified in April, May, or June are also included within the same analysis. However, change scores will be examined for the interval directly following reclassification. As mentioned, effects of reclassification do not manifest in the CST scores for that academic year, since these students entered mainstream English classes after taking the assessment.

The repeated measures analysis of variance revealed that ELA performance growth, collapsed across both districts, was significantly different across the reclassification period, \( F(1,141) = 9.230, p = .003, \eta^2 = .061 \). While enrolled in an ELD curriculum, students, on average, increased CST ELA scores by .17 standard deviation units relative to the English Only mean. However, subsequent to being enrolled in the mainstream English language arts curriculum, average scores decreased by .06 standard deviation units. Growth, therefore, changed by .23 standard deviation units in the
negative direction. Figure 12 is a time series chart which showcases this trend for the entire sample of valid cases (n = 149).

![Time Series Chart](Image)

**Figure 12. ELL – RFEP performance trend for CST ELA.**

The plot clearly shows the negative change in CST ELA performance after at least one month of English only instruction. Furthermore, the standard error bars show little overlap between the first two time points, indicating that performance increases while enrolled in the ELD curriculum are substantial. The standard error bars for the second two time points, however, overlap greatly. Thus, there is no significant decrease in CST ELA performance after being promoted to an English only curriculum; progress essentially stalls.

There were no significant differences in overall ELA growth patterns by grade of reclassification, F (3,141) = .739, p = .530. Although students who were reclassified in 4th grade had significantly higher scores than those reclassified in 6th and 7th grade, the pattern of growth remained statistically indistinct between groups. Below is another time
series chart showing trends in CST performance around the reclassification interval, but this time with data disaggregated by grade of reclassification. As can be expected, students reclassified in 4th grade are the most academically precocious of the sample, and exhibit higher initial CST ELA scores. Students reclassified in 6th and 7th grades show the lowest initial performance levels. Of additional note, all 4 groups of students show a similar pattern of performance, whereby scores initially rise, then decrease post-reclassification, consistent with the non-significant interaction of CST ELA change scores and grade of reclassification. However, it is apparent that students who are reclassified in later grades (6th and 7th) show a more pronounced rise and fall in performance. Those reclassified earlier (4th and 5th grades) make fewer gains prior to reclassification, but also exhibit minimal (or even positive) changes in performance post-reclassification.

Figure 13. ELL-RFEP performance trends for CST ELA, by grade of RFEP.
Growth patterns also did not change significantly by district, F(1,141) = 2.508, p = .116. Below is a time series chart for CST ELA performance, disaggregated by district of enrollment. District 1 (with the more stringent reclassification criteria) exhibited higher initial CST ELA scores throughout all testing intervals (the difference was significant, p<.001). However, students in both districts showed similar gains while enrolled in the ELD curriculum. Additionally, students across both districts showed a similar change in performance after being reclassified. While students in District 2 showed a slightly more positive adjustment to reclassification, the difference was small and non-significant.

![Figure 14. ELL-RFEP performance trends for CST ELA by district.](image-url)


Discussion of Findings

Thus far, several of the more divisive and contentious issues surrounding ELL assessment and intervention have been identified and reviewed. Subsequently, four specific research questions were posed in an effort to provide some initial resolution, and guide future practice. The first part of this chapter has generally been organized according to these four fundamental research questions, and this arrangement will continue in the interpretation of findings.

Research Question 1

The first question aimed to determine which test domain of the CELDT did ELL students experience the most difficulty? Significant differences among CELDT section performance levels were found. Generally, the Speaking and Listening sections proved more difficult than the Reading and Writing sections. Furthermore, students consistently scored below the reclassification threshold for both of these sections (by an average of 20 and 5 scaled score units, respectively). The distance from the reclassification cut score for Reading was more profound for District 2 with students scoring under the threshold by an average of 32 scaled score units. Thus, the tasks inherent in the Reading section seem to pose the most significant hurdle towards reclassification; thereby keeping students from accessing a more challenging and cultivating curriculum. As discussed earlier, protracted retention in an ELD curriculum may potentially incur poor academic outcomes. ELD curriculums have been observed to constitute de facto tracking systems, and it is important to understand and address barriers to a timely exit from ELL status.

Few studies have addressed specific testing barriers to reclassification, though the findings generally confirmed those of Robinson (2011), who also found Reading to be the
primary gatekeeper to mainstream language arts instruction. However, attention has been
given to the relationship between oral language skills and literacy skills at a more general
level. It is helpful to place the converging findings within the context of these more
general paradigms, in order to better elucidate the academic and linguistic trajectory
towards reclassification for ELL students.

In the simple view of the reading model (Hoover & Gough, 1990), literacy
development involves the use of two fundamental language domains: grapheme decoding
and general language comprehension. Literacy can, in its most distilled form, be
conceptualized as a combination of decoding and semantic understanding.
Correspondingly, the Speaking and Listening sections primarily correspond to tasks of
semantic understanding and comprehension, while the Reading tasks involve a
combination of decoding and comprehension. The question remains, then, in what
fashion do students acquire these skills? Is it constricted a sequential process?

For native English speakers, oral fluency has been identified as moderate
(Cunningham & Stanovich, 1997; National Early Literacy Panel, 2008) to strong (Snow,
Porche, Tabors, & Harris, 2007) predictor of later decoding and reading comprehension
ability. These studies point to a sequential process of skill development, whereby oral
language abilities develop prior to literacy skills. Although the English language learners
are uniquely distinct from native English speaker along many facets of language
development, it seems reasonable to assume the same general trajectory applies to both
groups, especially given that lack of oral vocabulary is commonly cited as a source of
limitation for reading comprehension among ELL students (see Lsaux & Kieffer, 2010
and Whitley, Smith & Conners, 2003). In fact, while native English speakers generally
begin kindergarten with a relatively large vocabulary as a basis for reading instruction, ELL students may begin with a much more limited vocabulary (Ortiz, 2014). Thus, ELL students are faced with the task of developing English oral language skills while simultaneously developing decoding skills. Given this state of affairs, it is reasonable to assume that one would observe oral language skills develop before literacy skills, as one is a prerequisite for the other. And, this is what was generally observed in the current analysis of CELDT performance trends.

Additionally, while ELL students have many opportunities to develop oral language skills through socialization within a generally monolingual environment, communication is heavily contextualized. Reading tasks, by in large, represents a communication medium devoid of context. Students do not have the opportunity to assess gestures, voice inflection, and the immediate surroundings in deciphering language.

It seems reasonable to conclude, then, that a series of linguistic tasks stands between reclassification for English language learners. While the immediate results indicate that Speaking and Listening tasks are developed relatively early and sufficiently to earn promotion to the mainstream curriculum, reading comprehension skills lag behind. An interesting question at this point becomes, to what extent does the Reading cut score play in ensuring ELL students are sufficiently prepared to navigate an English only academic environment? In other words, with basic Listening and Speaking skills mastered, would ELL students be able to apply these oral language abilities towards decoding within a more rigorous, mainstream language arts curriculum? If so, then the Reading tasks of the CELDT may be unnecessarily holding ELL students back. This question is partially addressed in this study (see section 4.4 and the discussion to follow).
Research Question 2

As mentioned in the previous section, there remains little to no discussion of specific cluster performance profiles for ELL learners on language assessments, or how these patterns impact reclassification. Thus, the second question this study sought to answer was which cluster of the standardized English language arts (ELA) assessment do ELL students experience the most difficulty? Studies have considered the general question of which assessment overall (CELDT or CST ELA) presents the most substantial impediment to reclassification; and the research has been mixed. Robinson (2011) found that the CST offered less of a reclassification barrier relative to the CELDT at younger grades. Conversely, others have presented evidence that the CST generally serves as the primary gatekeeper to reclassification (Salazar, 2005; Xiong & Zhou, 2006). Additionally, McGraw Hill (2006) and Stevens, Butler, and Castellon-Wellington (2001) showed that language proficiency assessments include items of lower complexity and abstraction, relative to academic content assessments (such as the CST ELA). However, in any of these cases, specific patterns of CST cluster performance were not examined.

At the level of global performance, the present study’s results are in agreement with those of Robinson. By two grade levels prior to reclassification, students were generally scoring at or above the reclassification threshold for all CST ELA clusters. This is contrasted with the CELDT, in which District 2 students scored nearly one full performance category below the threshold for Reading.

The fact that students were generally scoring at or above reclassification thresholds for all CST clusters renders the exact scoring profiles less relevant for this particular sample. The Reading Comprehension and Writing Strategies clusters contained
the lowest scores relative to the reclassification threshold. However, students, on average, were only scoring between 2.5% and 5% below the cut score at 2 grade levels prior to reclassification. It is interesting that students scored significantly lower on the Literary Response/Analysis cluster than the Reading Comprehension cluster, when both clusters require students to draw inferences from text. However, the Literary Response/Analysis requires students to identify general characteristics of a text and organizational patterns (e.g., determining form: poetry, non-fiction, drama; identifying the speaker; determining how a character is generally portrayed). In contrast, the Reading Comprehension cluster seems to require a more discrete understanding of the text (e.g., answering questions regarding specific events, following multiple step instructions, supporting answers by connecting prior knowledge with literal information from the text). It may be that students have multiple cues at their disposal to complete the more general tasks of the Literary Response/Analysis cluster. If students are unable to decipher one aspect of the text, other components may be utilized to draw a general conclusion. In contrast, the Reading Comprehension cluster may require accurate decoding of particular details to correctly answer items.

The low relative score for Writing Strategies might be expected, given the generally low Writing scores for the CELDT, as well as the notion that literacy skills develop after oral language skills. It may initially seem counterintuitive that students perform significantly lower on the Writing Strategies cluster than the Written Conventions cluster. However, inspection of the respective tasks for these clusters reveals that Writing Strategies involves a more complex and abstract understanding of written prose. While the Written Conventions cluster generally assesses knowledge of specific
text-based customs (e.g., identifying subjects, and verbs, identifying correct tenses, basic punctuation and grammar), the Writing Strategies cluster requires students to develop original sentences and paragraphs that have appropriate organization and focus. This is likely a skill that develops only after mastering basic writing customs.

It was also interesting to note that there was no significant growth across years for the Writing Strategies cluster. Although students make significant gains for Reading Comprehension, Word Analysis/Vocabulary, and Written Conventions; progress along the Writing Strategies stagnated. Again, given the relatively complex nature of the tasks inherent in this cluster, it is reasonable to assume that students take a number of years to develop full competency. It is important to note that the Word Analysis/Vocabulary cluster represented one of the stronger domains; consistent with results from the analysis of CELDT scores, it seems students develop oral language and vocabulary skills early, but experience difficulty applying these skills to decoding and comprehending text.

Although the observed scoring profile for the CST ELA clusters offers a fruitful source for making inferences of ELL linguistic trajectories, it is important to remind the reader that, within the context of the immediate findings, such analysis provides less utility. For the younger grades sampled in this study, the CST ELA represented much less of a hurdle relative to the CELDT. Thus, discrete analysis and conversation of CELDT performance trends are more appropriate with the immediate findings. However for that group students who do fall below CST reclassification thresholds, these scoring profiles may offer some initial guidance into early measures that can be taken for a timelier reclassification.

Research Question 3
The third research question of the present study was twofold: What is the relationship between the ELD standards and ELA standards, as reflected in CELDT and CST scores? How does proficiency on the CELDT relate to performance on the ELA? Results from the regression analyses revealed a generally weak to moderate relationship between CELDT scores and CST ELA scores. Between 10.4% and 20.7% of the variance in CST ELA scores could be accounted for by CELDT scores, depending on grade level. This relationship is slightly weaker than what has emerged in similar studies, which revealed $r^2$ values between .20 to .30 (Butler & Castellon-Wellington, 2005; Parker et al., 2009; Stevens et al., 2000).

Given that specific efforts have been made to link ELA and ELD standards, one might expect a stronger relationship between the CELDT and CST assessments. As mentioned, standards have been linked such that development along ELD skills should represent a “pathway” to ELA proficiency. In fact, the California Department of Education specifically notes that Advanced performance along ELD standards should correspond to emerging proficiency along ELA standards (CA Department of Education, 1999). This claim also implicitly denotes the form of the relationship; simultaneous and proportional gains should be observed along the CELDT and CST ELA. If the CELDT should be equally sensitive at all ability levels in predicting CST ELA scores.

In general, results support the implicit claim of a proportional relationship. Linear trends tended to provide the strongest fit to CELDT/CST ELA data for both 2nd and 3rd grades. While the relationship was clearly linear for the 3rd grade data set, the 2nd grade data set did show a slight bend, as indicated by the curved Loess line. This bend would indicate that the CELDT is more sensitive at lower ability levels. Indeed, both
logarithmic and quadratic functions provided significant fit above and beyond the linear model. The fact that the bend did not lead to a reversal in trend rendered the logarithmic function the most appropriate of the two (the quadratic function did not provide significant incremental benefit over the logarithmic function in terms of predictive power). However, abstracting the finding of the significant logarithmic fit to more substantive conclusions seems unwarranted. The increment in the coefficient of determination was relatively small ($r^2$ change = .034), and the difference in predicted scores between the linear and logarithmic models was almost negligible, especially when considering the general range of scores. The bend in the fitted logarithmic function is slight, and a general linear trend is preserved.

However, although the form of the relationship coincides with expectations, the strength of the relationship does not necessarily follow. At the second grade level, the linear model was only able to account for 10.4% of the variance in CST ELA scores, using CELDT scores a predictor (with an improvement of only 3.4% using the logarithmic model). This seems to be an uncharacteristically small relationship between the assessments, especially given the intended function of ELD standards in preparing students for ELA proficiency. What is even more interesting is the marked improvement in the relationship for third grade students. It may be that correlations between the assessments improve as more skills become available to test. The range of skills that can be assessed obviously increases with grade level, and it may be that a more elaborate sequential ordering of skills along the assessments improves performance correlations. In any case, this establishes the need for more research in this area; clarifying the extent to
which the relationship between ELD and ELA assessments improves with grade level has implications for improving the linkage between ELD and ELA standards.

Additionally, results from the regression analyses indicate some deviance between predicted and expected CST ELA performance levels from CELDT scores. While a linear trend does seem to define the performance relationship between the assessments as expected, the slope of the trend itself may not coincide with expectations. As mentioned, the claim is made by California Department of Education that Advanced performance along ELD standards should correspond to Proficient performance along ELA standards. The regression models derived from this data indicate that 2nd grade students performing at the Advanced level on the CELDT are still scoring within the middle of the Basic category of the CST ELA (26 points lower than expected). This may indicate that, for higher level items on the CELDT, item complexity is lower than intended. Conversely, for third grade students, the two performance categories show better alignment.

Applying the regression models to the CELDT reclassification cut scores, the predicted CST scores fall beyond the minimum Basic threshold score (300). Most districts use Basic performance on the CST ELA as a minimum criterion for reclassification; thus, these results indicate that students in both 2nd and 3rd grades are achieving minimum reclassification CST scores before minimum CELDT reclassification cut scores. As discussed previously, these findings contribute to a research base that is mixed with regards to the relative difficulty of the CELDT and CST for reclassification. It is interesting that both this study and Robinson (2010) obtained converging results with a sample of younger grades. The apparent contradiction in the literature may stem from
the fact that the relative difficulty of the two assessments for reclassification may switch at a certain developmental level.

**Research Question 4**

The final question examined whether there was a significant difference in ELA growth rates before and after reclassification? As a prerequisite analysis in addressing this research question, CST ELA performance trends over 5 years were examined for 3 groups of students; those reclassified between August and November, those reclassified between December and March, and those reclassified between April and June. This allowed a determination as to whether month of reclassification had any differential effects on CST ELA performance that year. This preliminary exploration informed the selection of intervals for the analysis, and whether subjects should be separated by reclassification month in the final analysis.

For those students reclassified between August and November, there is a clear stall in CST ELA growth beginning at one year prior to reclassification. Students exhibit substantial growth from two years prior to reclassification to one year prior to reclassification; however, performance stagnates from one year prior to reclassification through the following interval. Given that this group of students was reclassified early in the school year, the stall in growth beginning at one year prior to reclassification may be attributed to the change in the academic program. In other words, by the time these students were administered the CST ELA in their year of reclassification, they had already been enrolled in a mainstream language arts curriculum for up to eight months. This interval of time in a novel academic program may have been sufficient to affect test scores later that spring. It should be noted that an upward trend is re-established in the
following interval (though not near as pronounced as the initial interval recorded) and may indicate the effect of an “adjustment period.” That is, CST ELA scores may initially stall within 5 to 8 months post-reclassification, as students continue to learn to navigate the mainstream curriculum with limited language support. However, with time, students adapt to the new academic environment and re-establish a growth pattern.

For students reclassified between December and March, several growth patterns are consistent with those observed in the previous plot. First, growth is observed from 2 years prior to reclassification to 1 year prior to reclassification. Second, growth ceases between one year prior to reclassification and the year of reclassification; in this case, performance actually drops for this interval. Lastly, positive growth is re-established by the following interval; again, possibly pointing to the presence of an adjustment period.

The third plot includes students who were reclassified after taking the CST ELA assessment that particular school year. If the assumption is correct that reclassification affects changes in academic performance, then one would not expect any substantial decrease in academic performance for the CST ELA assessments that year. Indeed, the graph shows that students reclassified between April and June continue to improve their CST ELA scores until the following academic year. At that point (after being enrolled in a mainstream language arts program for a maximum of nine academic months), a drop in performance becomes apparent. Distinct from the other student groups, this drop continues throughout the following interval.

Overall, these plots indicate that stalls or declines in CST ELA performance are observed for students reclassified from August through March in that academic year. However, there is a slight difference between the two trends that may have interpretive
value. Students reclassified earlier in the year (August – November) exhibit CST ELA performance at a level almost identical to that of the previous year. As such, although growth stalled, absolute performance levels remained constant. Conversely, those students reclassified in intermediate months (December through March) exhibited an observed drop in performance in the year of reclassification. That is, growth did not just stall; it reversed direction. Assimilating these two findings is relatively straightforward if one considers adjustment periods. It may be that students reclassified earlier in the year have more time to adjust to the mainstream curriculum, and show less pronounced performance changes on the CST ELA that year. However, those students reclassified later are not offered the same adjustment period, and CST ELA scores suffer more that year. By the year following reclassification, both groups of students have had sufficient time to adjust and re-establish positive growth trends.

The implications of these findings for the immediate research design were delineated in presentation of findings and the following discussion is concerned with integrating these initial results with outcomes from the ANOVA tests to draw more general conclusions regarding ELL assessment. Thus, a continued review and interpretation of the ANOVA results, with the previous findings is integrated throughout the discussion.

There was a significant drop in CST ELA growth post-reclassification, collapsing across district of enrollment and reclassification grade. The significant decrement in CST ELA growth post-reclassification would initially indicate premature promotion to an English only academic environment. That is, performance is not preserved post-reclassification because students are not yet linguistically prepared to navigate the
mainstream language arts curriculum. This finding would seem to contradict previous outcomes from similar studies (Flores, 2009; Matsudaira, 2005; Robinson, 2011), which showed null to positive effects of reclassification.

However, further inspection of the immediate results actually reveals more consistency than discrepancy with the prior studies. Firstly, the effect size of the change in growth rate was modest at best, with an estimated $\eta^2 = .061$. Thus, the change in growth rate, by all intents and purposes, does not translate to drastically lower performance levels. In fact, the observed drop in score is only by a margin of $1/17$ standard deviation units, and growth itself only declines by only roughly $1/5$ of a standard deviation unit. From a general perspective, the substantive negative effects of reclassification seem minimal. This is especially true considering the plethora of benefits accessible from a more challenging academic environment, with potentially greater resource availability.

Secondly, for the majority of students, the initial drop in CST ELA performance is immediately followed by re-establishing a positive growth trend. While slight, these positive trends may become more pronounced as students continually adjust to the English only academic environment. Overall, then, it would appear that the reclassification criteria are not exiting students prematurely, at least by analysis of CST ELA score changes. The results from the immediate analysis do not indicate a total mismatch between instructional level and student ability level once reclassified.

To close this discussion, two additional findings are highlighted with respect to the effects of reclassification on academic performance. The nature of the reclassification effect did not change across districts or grade of promotion. Although differences in
overall CST ELA performance were significantly higher for those reclassified in fourth grade, and those enrolled in District 2, growth trends themselves remained relatively indistinct across these factors. The implications of these findings are two-fold. Firstly, ELL students who enter mainstream language arts classes in later grades (6th and 7th) are affected by reclassification in a similar manner as those promoted in earlier grades (4th and 5th). Thus, while retaining students in an ELD curriculum for a several additional years may extend provision of a remedial academic curriculum (and more opportunity to fall behind) the adjustment pattern is similar to those who are promoted in a more timely fashion. Thus, students with differing experiences in ELD curriculums all seem to adjust similarly and relatively well to academic promotion. It should be noted that those reclassified in later grades seem to show slightly more pronounced academic growth prior to reclassification, relative to those reclassified in earlier grades. Subsequent to reclassification, those reclassified in later grades also show steeper drops in performance. While this might initially point to the increasing negative effects of reclassification with increasing grade, the actual observed decrements are relatively small. For example, comparing the most extreme groups (those reclassified in 7th grade to those reclassified in 4th grade) the difference in post-reclassification growth differs by roughly 1/10 of a standard deviation unit.

Secondly, applying more stringent reclassification criteria did not alter reclassification growth trends by any substantive amount. District 1, which imposed slightly higher cut scores for reclassification, clearly promoted students who were functioning at a higher academic achievement level. However, this higher level of achievement had little bearing on growth patterns post-reclassification. This indicated an
important consideration when establishing cut scores for reclassification: shifting the CST ELA cut score for reclassification along the Basic and Proficient ranges does not seem to augment adjustment to the mainstream curriculum. Based on the immediate findings, increasing or decreasing reclassification thresholds for the CST ELA will only ensure higher achievement levels for newly designated RFEP students. It will not necessarily ensure a smoother transition to the English only academic environment.
CHAPTER V

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

Since the decision in Topeka vs Board of Education (1954), educational agencies have progressively adopted a more inclusive attitude toward serving students with diverse needs. That is, the mainstream curriculum should be accessible to all students to a maximal extent that still provides benefit. This is, undoubtedly, a noble goal which has improved life outcomes for many individuals with unique needs. However, sustaining this pillar of educational service delivery has proved increasingly difficult over the last 30 years. In particular, the population of language minority students has increased at a rate nearly seven times that of the general population (Flynn & Hill, 2005). Obviously, this has placed increased demand on schools to secure bilingual personnel, employ teachers who are competent addressing diverse language needs, and adopt appropriate language curriculums. Additionally, English language learners are a heterogeneous group, and generally follow an erratic and discontinuous trend of language growth (Kohnert, 2010). Thus, determining the manner and extent to which these students should be educated with the mainstream curriculum is a formative challenge for educators; and as might be expected, fraught with much debate.

Thorough review of the literature addressing ELL education reveals four issues that define the fundamental complexities in serving these students. First, there is substantial controversy and debate regarding the most efficient form of instruction for second language learners. While the political landscape seems to favor an English
immersion model, there is emerging research pointing to the effectiveness of bilingual models. Second, the operationalization of “academic English” remains unresolved; that is, no single framework exists that universally defines classroom level English. The relative complexity is due in part to discourse stylings changing by academic area. In any event, the lack of a universally accepted definition of academic English obfuscates valid assessment of academic English proficiency. It is difficult to determine if ELL students are ready for the mainstream instruction when the definition of academic English proficiency itself remains nebulous. Extending from these states of affairs, the validity of reclassification criteria remains equivocal. Concerns have been voiced that the criteria used to determine ELL status over-identify this population. By returning an inflated rate of false positives, it may be that some students are unnecessarily subjected to a remedial, ELD curriculum. Lastly, there have been numerous claims and observations made that ELD problems themselves are oversimplified, harbor fewer resources, fewer qualified teachers, inappropriately low expectations, and may ultimately hinder the progress of ELL students.

The point in reviewing these seemingly bleak qualities is not to accuse ELL assessment and instruction as fundamentally flawed or argue that it has no redeeming qualities as currently implemented. Rather, these potential issues are presented in order to define a context for relevant and future research directions. We derive two essential truths from this research base to more explicitly define this context: 1) it is imperative to reclassify students as soon as appropriate since protracted ELL instruction beyond what is absolutely necessary is undesirable, and 2) we do not yet have a clear and unequivocal definition of what “appropriate” is; that is, determining if students are ready for
reclassification remains a muddled affair. The current research endeavor seeks to provide some initial insight into addressing the situation by asking four basic research questions. First, what are the most significant obstacles to reclassification in terms of assessment? In other words, which of the assessment criteria seems to be the primary “gatekeeper” to English only promotion? Second, how do the assessments interact to affect the reclassification process? Lastly, how does academic growth change after reclassification? It is taken for granted that marked changes in academic growth post-reclassification can serve as a proxy for the validity of the reclassification criteria; if it is too liberal, students may show a marked decrement in growth. Conversely, if reclassification criteria are overly stringent, students may show a spike in performance post reclassification, indicative of an ELD curriculum that has been holding them back.

In addressing these questions, archival assessment data from 190 former ELL students was obtained from educational records. Specifically, language proficiency scores (from the CELDT) and language arts achievement performance (from the CST ELA) were recorded across a span of five years for all students. Specifically, scores were recorded for the two years prior to being reclassified through two years post-reclassification. The students differed in the grade of reclassification; though all students were reclassified between 4th and 7th grade. Additionally, the students were sampled across two districts. Of primary interest to this study was the fact that these districts differed in terms of the reclassification criteria utilized. While the same exact assessments were utilized for reclassification decisions, the two districts employed different cut scores. One district required a score of 330 on the CST ELA, while the other
the district required a score of 300 on the CST ELA. This allowed comparison of how different cut scores differentially affected academic performance post-reclassification.

The first set of questions (regarding relative “gatekeepers” to reclassification) was analyzed at both the overall score level, as well as subdomain level for both the CELDT and CST ELA. In other words, we asked not only which of these assessments proved more difficult for reclassification overall, but also which particular cluster/section seemed to be source of most difficulty. This was done by identifying an expected “benchmark” performance level for reclassification, for each individual cluster or section across both assessments.

Results showed that averaged across grades 2 through 4 (2 years and 1 year prior to reclassification), ELL students were scoring at or above the reclassification thresholds for all CST ELA clusters. Although students scored significantly lower on the Writing Strategies and Reading Comprehension clusters, they were still performing at the expected level for reclassification.

Conversely, overall scores on the CELDT fell below the reclassification threshold across two years prior to reclassification. Thus, our data indicated that the CELDT, rather than the CST, proved to be the gatekeeper assessment for most students. This was also partially confirmed by regressing CST ELA scores on CELDT scores. Examination of the individual CELDT sections revealed performance patterns that were somewhat consistent with those of the CST ELA. Both the Reading and Writing sections were the only two areas where students consistently scored below the reclassification threshold. Of the two, the Reading section returned the lowest scores. Thus, while ELL students were generally
exhibiting Listening and Speaking skills at a level necessary for reclassification, Reading performance was primarily impeding promotion to the mainstream curriculum.

Reading performance was the primary obstacle to reclassification was expected, given general models of literacy acquisition. Reading development can be understood as a sequential model, whereby individuals first master oral comprehension. Subsequently, these general comprehension skills become applied to text once decoding skills are acquired. The findings from the current study seem to be consistent with such a model; students seem to show proficient skills in oral comprehension before developing more advanced reading skills.

This observation leads to another question of particular importance in the reclassification process. Once ELL students have developed a sufficient set of oral language skills, would they be able to make a smooth transition to the mainstream curriculum while still developing grade level literacy skills? In other words, does a basic set of oral language skills represent the minimum competence level to benefit from mainstream instruction? If this is the case, then the relative difficulty of the Reading section may be unnecessarily retaining students in a less challenging curriculum. Decoding and reading comprehension skills seem to be the next steps after oral language skills. The essential question is, can students develop these latter abilities within the mainstream curriculum after developing the former in a more restricted educational setting?

Although the validity of reclassification criteria is partially addressed within this research project, a more in-depth analysis addressing this question alone is immediately pertinent. While the results of the current study provided some initial insight into the
validity of reclassification criteria, a more discrete analysis is warranted. Only one measure (CST ELA) was examined as a proxy for overall academic adjustment; additional studies may improve the reliability of findings by including several outcome measures. Future studies examining the relative importance of individual sections may be difficult, given that district and state level mandates would disallow reclassification of students at different CELDT levels. However, the question of differential validity for sections may be partially addressed by administering CELDT assessments to English only students. If all students score high on the Listening and Speaking sections, but show a range of performance levels on Reading and Writing sections, this may provide preliminary evidence that Listening and Speaking should be weighted much heavier in reclassification decisions. From another point of view, if there are English only students who are progressing and benefiting from the mainstream curriculum, but who score below the reclassification threshold for the Reading section, then it would be inappropriate to hold ELL students to a higher standard for access to the mainstream curriculum.

For the second analysis in this project, the main focus was on the relationship between the two primary assessments used in reclassification decisions (the CELDT and CST ELA). Several components of this relationship were of primary interest; including the form of the relationship (linear, quadratic, logarithmic, etc.), the strength of the relationship, and the correspondence of proficiency levels across both assessments. The fact that the California State Department of Education has made an effort to link ELD to ELA standards in a particular manner provided additional context and relevancy for these questions. Specifically, the ELD standards are to represent pathways to ELA proficiency,
such that Advanced performance on the CELDT should roughly correspond to Proficient performance on the CST ELA.

At both the 2nd and 3rd grade levels, the relationship between the assessments was generally dominated by a linear function. This would confirm the notion that one set of standards leads to simultaneous and proportional gains along the other set, as implicitly intended by the state of department of education. However, the strength of the relationship for second grade students was generally less pronounced than might be expected. Roughly 10-13% of the variance in CST ELA scores could be accounted for by CELDT. Given that the ELD and ELA standards (and therefore the CELDT and CST ELA) have been explicitly linked, one might expect the CELDT to have more power in predicting CST ELA scores. Additionally, the regression model showed that second grade students scored half a performance category below what was expected on the CST ELA, given an Advanced score on the CELDT. This further supports a relatively weak CELDT/CST ELA relationship for second grade students.

However, the strength of the linear relationship at the third grade level nearly doubled in size, accounting for over 20% of the variance in CST ELA scores. This is more consistent with previous studies, which have shown language proficiency assessments to account for 20-30% of variance in language arts achievement assessments. Of particular note, these other studies included a range of higher grade levels in the analysis. Taken together, this may indicate stronger language proficiency/language achievement relationships with increasing grade level. Future studies examining this particular notion in greater detail would offer an important and practical contribution to the literature base. The structure and linkage of ELD and ELA standards are of critical
importance, as they interact to define the linguistic developmental trajectory of ELL students. The standards should be carefully constructed and linked to offer the most expedient pathway to academic English proficiency. Smaller than expected relationships between the CELDT and CST ELA at particular grade levels may indicate misalignment between the respective standards.

For the last analysis, the validity of current reclassification cut scores was examined by evaluating changes in CST ELA performance patterns across the reclassification period. The assumption was made that a smooth transition from ELD classes to the mainstream curriculum offered the strongest evidence for appropriately set cut scores. If reclassification cut scores were set too low, one might expect to observe negative changes in CST ELA growth. In this situation, students would be reclassified before academic English skills could be developed sufficiently. Conversely, if reclassification criteria were too stringent, we might expect students to show substantial increases in academic growth once reclassified. The logic here is that these students would have been unnecessarily held back in a curriculum that was overly rudimentary, given their immediate level of ability. As such, once given access to an appropriately challenging curriculum, academic growth would be augmented.

The findings revealed significant, though small changes in growth patterns following reclassification. Generally, decrements in academic growth were observed post-reclassification. However, these decrements were so small that we consider them negligible. Furthermore, whatever minor decrements were incurred on growth seemed to have reversed by the following academic year. Thus, as currently implemented, reclassification criteria do not seem to be overly liberal. Although the relatively smooth
transition also indicates that criteria are not overly stringent, more refined analyses are required to arrive at a more tenable conclusion.

It is important to note that negligible changes were also observed for students reclassified across all grades. This finding further confirms the fact that reclassification criteria seem to be generally appropriate. For example, suppose that those reclassified in fourth grade experienced substantially negative changes in growth rates, but those reclassified in seventh grade did not. This might indicate that reclassification criteria are promoting a sizeable number of students too early in their linguistic development. The fact that this is not the case would imply that those students reclassified in fourth grade are indeed well prepared to access the English only curriculum. Furthermore, this would indicate that promotion has not been inappropriately delayed for those reclassified in seventh grade. Again, more sophisticated and advanced analyses are necessary to draw a stronger conclusion in this second case.

Growth patterns also remained similar across districts. This implies that lower reclassification cut scores on the CST ELA may still affect similar outcomes. Obviously (and as observed in the current analysis), a higher CST ELA cut score will reclassify students at a higher overall level of academic performance. However, and important to the particular research question, absolute growth remains relatively unchanged. Decisions regarding specific cut scores, should take these findings into account. At least within the range of Basic to Proficient scores on the CST ELA, reclassification thresholds can be shifted up and down without necessarily hindering or augmenting the adjustment process. Within the context of these immediate findings, and the apparent benefits of mainstream instruction, utilizing a cut score of 300 on the CST ELA seems to be optimal. This allows
for earlier reclassification, while still preserving a smooth transition process that is not significantly different from those reclassified with higher cut scores.

There are a few noteworthy points regarding this last analysis. First and foremost, is the recognition that seemingly sweeping conclusions are made based on a single outcome measure. This claim is indeed justified, though conclusions are meant to be initial and tentative. In other words, it would be desirable that these findings provide inspiration for subsequent research endeavors to provide additional confirmation with advanced or alternative statistical analyses. This study, in some form or another, can be viewed as a pilot study, providing a foundation for the development of future research efforts.

Second, the ELL community is diverse, and these students follow a course of linguistic development that is difficult to predict. Thus, any study which aggregates these students into a single group immediately loses a wealth of information. These students differ on a range of individual, family, community, and cultural differences which substantially impact their response to assessment and intervention. These variables were uncontrolled in the current study, which does obscure results to a certain extent.

Similarly, we did not factor district level differences into the analyses. It was assumed that the non-significant interaction between post-reclassification academic growth and district of enrollment was indicative of the negligible effect of increasing the reclassification threshold by half a performance level.

One final highlight of the last analysis determined through a more qualitative analysis of data. As mentioned, changes in growth rates across the reclassification period were significant, though the effect size was minor. However, when trends were
disaggregated by months of reclassification, a clear pattern was observed that may have implications for ELL programming and decision making. Students reclassified in the earlier months (August – November) showed a pattern of stalled CST ELA growth for the year of reclassification. That is, scores remained relatively unchanged from the previous year. Conversely, those students reclassified later (December – March) demonstrated a clear drop in performance. These results imply a slight negative effect of mid-year reclassification.

The notion of an “adjustment” period may offer some explanation for these findings. Those reclassified in earlier months may have more opportunity to adapt to the new educational surroundings and workload, and may show less pronounced decrements in CST ELA performance by April. However, those students reclassified middle to late in the year may have less opportunity to adapt to the new instruction and curriculum. These students would exhibit a pronounced drop in CST ELA performance that April.

Although the size of these effects was relatively small in the current study, this phenomenon should be followed up with future studies. The effect of the “adjustment” period may very well be more pronounced in other districts with different demographics and curricula. If this finding can be replicated with a larger effect size, the implications seem particularly noteworthy. First off, it is reasonable to assume that performance on summative achievement measures (such as the CST ELA) has the potential to influence educational placement decisions. If reclassifying students mid-year consistently attenuates performance relative to those reclassified early in the year, administrators would be wise to reconsider reclassification intervals. The implications also extend to issues of reporting adequate yearly progress and maintaining compliance with the tenets
of the Elementary and Secondary Education Act. Reclassification in earlier academic months may affect more positive accountability reports to state and federal level educational agencies.
REFERENCES


http://www.edcoe.org/departments/curriculum_instruction/documents/082207CILC_SectionV.pdf

California State Department of Education (2012). *2012 Star Test Results*. Retrieved from:

http://dq.cde.ca.gov/dataquest/


http://www.cde.ca.gov/ds/sd/cb/cefelfacts.asp


