UCLA
UCLA Electronic Theses and Dissertations

Title

English Language Proficiency Classification, Reclassification, and Educational Programming Decisions for Language Minority Students: A Mixed Methods Study

Permalink

https://escholarship.org/uc/item/5cz284p6

Author

Carroll, Patricia Elaine

Publication Date

2015

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA

Los Angeles

English Language Proficiency Classification, Reclassification, and Educational Programming Decisions for Language Minority Students: A Mixed Methods Study

A dissertation submitted in partial satisfaction

of the requirements for the degree Doctor of Philosophy

in Special Education

by

Patricia Elaine Carroll

2015

© Copyright by

Patricia Elaine Carroll

2015

ABSTRACT OF THE DISSERTATION

English Language Proficiency Classification, Reclassification, and Educational Programming Decisions for Language Minority Students: A Mixed Methods Study

by

Patricia Elaine Carroll
Doctor of Philosophy in Special Education
University of California, Los Angeles, 2015
Professor Alison L. Bailey, Chair

This study investigates the influences and constraints on data use pertaining to initial classification, reclassification, and educational programming for English Learners (ELs). A growing number of language minority students face extended time in EL programs past the typical 4-7 years of language acquisition; however, multivariate data use in reclassification decisions has rarely been investigated as a factor. Using mixed methods, this study describes the variation of reclassification outcomes by different combinations of data (composite and aggregate decision rules) and describes the use of data by decision makers in EL programs. Descriptive analyses were conducted to examine variation in reclassification eligibility according to six composite and two aggregate decision rules applied to performance data of EL students in grades 2-11 (n=667), including high-performing EL students (n=107), EL students with disabilities (n=78), and high-performing EL students with disabilities (n=8). Findings illustrate

the wide variation in reclassification eligibility for all groups, and the broad range of abilities in students not reclassified. Interviews with district decision makers (*n*=16) and analyses of district documents illuminated variety in data use. Interviews also illustrated influences and constraints on noticing, interpretation, and constructing implications of data, especially for EL students at the intermediate level. Taken together, findings suggest that certain district-chosen rules can dramatically reduce reclassification eligibility, and yet the long-term EL phenomenon is more often considered to be an outcome of student characteristics (e.g., motivation, special needs) rather than a result of stringent decision rules. Findings stand to inform assessment validation, policy, and multivariate data use practices.

The dissertation of Patricia Elaine Carroll is approved.

Christina A. Christie

Todd M. Franke

Diane Haager

Lois A. Weinberg

Alison L. Bailey, Committee Chair

University of California, Los Angeles

2015

To my parents, with gratitude.

Table of Contents

CHAPTER ONE: Introduction 1

CHAPTER TWO: Review of Relevant Literature.....9

CHAPTER THREE: Study Methodology 31

CHAPTER FOUR: Quantitative Data Presentation and Analysis.....51

CHAPTER FIVE: Qualitative Data Presentation and Analysis 62

CHAPTER SIX: Discussion 94

Appendices.....105

References.....119

List of Tables

Table 1: State Composite Rule: California English Language Development Test.....5

Table 2: Accountability and Improvement Uses of ELP Levels and Classifications.....14

Table 3: Schedule of Research Activities.....34

Table 4: Inclusion Criteria for Groupings.....41

Table 5: Composite Rules.....42

Table 6: Aggregate Rules.....42

Table 7: Participant Roles in the District.....45

Table 8: Descriptive Data: Students with Disabilities.....51

Table 9: Descriptive Data by Group: Mean Levels and Scale Scores, CELDT and CST .52

Table 10: Percentage Proficient by Composite Rules: EL.....52

Table 11: Percentage Proficient by Composite Rules: EL/SWD.....53

Table 12: Percentage Proficient by Composite Rules: EL/HP 53

Table 13: Number Proficient by Composite Rules: EL/SWD/HP.....54

Table 14: Percentage of Students Meeting CST-ELA Criteria.....54

Table 15: Percentage Eligible for Reclassification by Aggregate Rules: EL 54

Table 16: Percentage Eligible for Reclassification by Aggregate Rules: EL/SWD 56

Table 17: Percentage Eligible for Reclassification by Aggregate Rules: EL/HP.....57

Table 18: Number Eligible for Reclassification by Aggregate Rules: EL/SWD/HP 58

Table 19: The Seasons of Annual Testing and Data Use for EL Status Decisions 63

Table 20: Reclassification Criteria: State and District.....68

List of Figures

Figure 1: Aggregate Decision Rule, EL Reclassification in California Grades 2-12 9

Figure 2: Procedural Diagram: Mixed Methods Concurrent Design.....27

Figure 3: Framework for Data Use (Coburn & Turner, 2011) 28

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Alison Bailey, and my committee members, Tina Christie, Todd Franke, Diane Haager, and Lois Weinberg, for their contributions to my growth as a researcher, scholar, and writer. I would also like to thank the "District" for allowing me to conduct this research in their midst. In addition, I would like to thank the countless faculty, colleagues, friends, and family members who have stood beside me on this journey.

VITA

Patricia Elaine Carroll
patriciacarroll@ucla.edu

EDUCATION

Ph.D. Special Education (Expected, September 2015)
University of California, Los Angeles & California State University, Los Angeles (Joint Doctorate)
Specialization in Language and Literacy Assessment
Concentration in Human Development and Psychology
Dissertation: *English Language Proficiency Classification, Reclassification, and Educational Programming Decisions for Language Minority Students: A Mixed Methods Study*
Committee: Alison Bailey (chair), Tina Christie, Todd Franke, Diane Haager, Lois Weinberg

M.A., Education (2012)
University of California, Los Angeles
Thesis: *Examining the Validity of Classifications from an English Language Proficiency Assessment for English Language Learners and Native English Speakers in Fifth Grade*

M.A., Teaching English to Speakers of Other Languages (2000)
California State University, Los Angeles

California Single Subject Teaching Credential: English with CLAD (1998, current)
California State University, Los Angeles

B.A., Business Administration & Physical Education (Individual Goal Oriented) (1991)
Taylor University, Upland, Indiana

HONORS AND AWARDS

UCLA Research & Inquiry Conference Paper Award, Second Place, 2014
UCLA Graduate Summer Research Mentorship Award, 2012
UCLA Graduate Summer Research Mentorship Award, 2011

RESEARCH EXPERIENCE

Graduate Student Researcher

Community College Pathways Evaluation Project, 2013-2015
Social Research Methodology, University of California, Los Angeles

Evaluating the Validity of English-Language Assessments (EVEA) Project, 2010-15
Human Development and Psychology, University of California, Los Angeles

Learning in Two Languages (LTL) Research and Development Project, 2010
Human Development and Psychology, University of California, Los Angeles

Literacy Learning Cohort (LLC) Project, 2009-2010
Special Education, California State University, Los Angeles

TESOL Project, Graduate Excel Program (2000)
TESOL, California State University, Los Angeles

TEACHING EXPERIENCE

Teaching Fellow, Writing Programs, University of California, Los Angeles, 2013 – 2015
Graduate Teaching Assistant, Human Development & Psychology, UCLA, 2013
Graduate Teaching Assistant, Special Education, CSULA, 2009 – 2010
Graduate Writing Consultant, Graduate Writing Center, UCLA, 2012 – 2013
ESL Instructor, Glendale Community College, CA, 1999 – 2003; 2013 – 2014
Guest Lecturer, Teacher Training Institute, USC, January – February 2009
ESL Lecturer and Marketing Specialist, University of Southern California, 2001 – 2003
ELD/English Language Arts/History Teacher, Pasadena High School, CA, 1997 – 1999
EFL Guest Professor and Central Administrator, Pushkin, **Russia,** 1997
EFL Teacher and Central Administrator, Dunaújváros, **Hungary,** 1992 – 1994

PROJECT LEADERSHIP EXPERIENCE

Recruitment and Placement Director, Los Angeles Urban Teacher Residency at CSULA
Center for Collaborative Education, Boston, MA, 2010
Senior Educational Consultant,
Consortium on Reading Excellence, Inc., Berkeley, CA, 2003 – 2009

PUBLICATIONS

Carroll, P.E., & Bailey, A.L. (2015). Do decision rules matter? A descriptive study of English language proficiency assessment classifications for English-language learners and native English speakers in fifth grade. *Language Testing*.

Bailey, A.L., & **Carroll, P.E.** (2015). The assessment of English language learners in the era of new academic content standards. *Review of Research in Education, 39, 253-297.*

Osipova, A., Prichard, B., Bourdman, A.G., Kieley, M.T., & **Carroll, P.E.** (2011). Refocusing the lens: Enhancing elementary special education reading instruction through video self-reflection. *Learning Disabilities Research & Practice,26*(3),158-171.

PRESENTATIONS (SELECTED)

Carroll, P.E. (2014, October). *Where policy meets practice: Data use in a Title III program*. Paper presented at the Northeastern Educational Research Association Annual Meeting, Trumbull, CT.

Carroll, P.E., & Bailey, A.L. (2014, April). *Classification models and English learner redesignation: High-performing students left behind?* Paper presented at the National Council on Measurement in Education Annual Meeting, Philadelphia, PA.

Carroll, P.E. (2013, April). *Classification model validity for English language proficiency assessments: Insights from native English speaker performances.* Poster presented at the American Educational Research Association Annual Meeting, San Francisco, CA.

Carroll, P.E., & Bailey, A.L. (2013, April). *Language-as-resource in English learner assessment systems: Evaluating the fit of classification models.* Paper presented at the American Educational Research Association Annual Meeting, San Francisco, CA.

Carroll, P.E., & Bailey, A.L. (2013, April). *Combining multiple indicators in classifications of English language proficiency: A descriptive study.* Paper presented at the National Council on Measurement in Education Annual Meeting, San Francisco, CA.

CHAPTER ONE

INTRODUCTION

Decisions are made every day about the education of school-aged children in the United States. At the local level, teachers assess students' responses to instruction, mastery of basic skills, engagement in standards-based content, and growth trajectories towards educational goals. Decisions are made about educational settings and individualized educational plans for students needing specialized instruction, including students who are still acquiring English, students with specialized learning needs and cognitive disabilities, and students with exceptional giftedness. School districts must make decisions about who is eligible for specialized programs, if students' progress within such programs is sufficient, and when students are eligible to exit programs that are meant to be temporary. The standards which govern these decisions are guided by educational accountability policies at federal, state, and local levels. At the heart of today's accountability policies are assessment tools and their infrastructure systems. Underlying the use of assessments are the commonly agreed-upon standards for test development, standards for evaluating test score reliability and validity, and guidelines for the combination and use of multiple measures for high stakes decisions. Multiple measures are used to overcome the known limitations of basing high-stakes decisions on one errorful test score. Combining measures is not without its own limitations, however, and care must be taken to ensure that the combination produces reliable, valid, and fair inferences for each use. Surrounding these assessment systems are improvement efforts in classrooms, schools, districts, and states.

As part of this nationwide reform effort, there is one group of students who are subject to more testing than the general population. English-language Learners (ELs) are a subgroup of students from language minority homes who, through initial testing at school enrollment, have

been found to need additional language acquisition and development support. This support is meant to be temporary. Once ELs reach a language level where they can meaningfully participate in or benefit fully from instruction delivered in English, the English Language Development (ELD) services end and students spend all their schooling hours in mainstream classes. At that point, EL students are reclassified as "Fully English Proficient" (FEP or RFEP), a status that becomes permanent. Thus, language minority students are either permanently classified as Initially Fluent English Proficient (IFEP) and not in need of ELD, temporarily classified as Limited English Proficient (LEP) and eligible for ELD as needed, or reclassified as RFEP, a permanent classification indicating that ELD has been received, but is no longer needed. English Language Proficiency Assessments (ELPAs) are given at initial enrollment, and again annually, by each state to determine status and services. The EL designation, designed to be temporary, activates funding for states to provide an appropriate combination of language development and content-based instruction with the goal of raising student English language proficiency (ELP) to a point where it does not hinder success in English-medium instructional environments.

According to the US Department of Education, an estimated 4.4 million students, or 9.1 percent of the school-aged population, were designated as EL in 2011-12. The proportion of EL students in each state varies from 0.7 percent in West Virginia to 23.2 percent in California (United States Department of Education, 2012). Although more students than ever before are entering EL programs, the number of students able to successfully exit those programs has not kept pace as seen in the number of students participating in initial English language proficiency (ELP) testing versus annual testing (e.g., California, see Appendix E). The so called Long Term

EL[1] population is a growing percentage of language minority students nationwide. This has led to concern about potential errors in the classifications (Ragan & Lesaux, 2006; Bailey & Carroll, 2015), and the reclassification process (Grissom, 2004) leading some to call for greater scrutiny of decision makers at the district level (e.g., California's Senate Bill 1108, see Linquanti & Hakuta, 2012).

To date efforts have been made to ensure the psychometric quality of ELPA and move towards a common or comparable definition of "English Learner" as determined by ELPA performance standards which make state-by-state comparisons less ambiguous. These efforts have recognized concerns about lack of remedies for EL misclassification arising from assessment imprecision for which "current federal law provides no guidance" (p. 11, Linquanti & Cook, 2013).

Multiple sources of evidence are a necessity for high-stakes decisions. Best practices of using multiple measures are crucial for decisions about EL students, for whom misclassification can have long-term consequences (Wolf, Herman, & Dietel, 2010; Wolf, Kao, Herman, et al. 2008). According to the *Standards for Educational and Psychological Testing* (American Educational Research Association [AERA], American Psychological Association [APA], and National Council on Measurement in Education [NCME], 2014, p. 198):

> Standard 12.10: In educational settings, a decision or characterization that will have a major impact on a student should take into consideration not just scores from a single test but other relevant information.

[1] "Long-term English learner" is a term adopted by the California State Board of Education to mean an English learner who is enrolled in any of grades 6 to 12, inclusive, has been enrolled in schools in the United States for more than six years, has remained at the same English language proficiency level for two or more consecutive years as determined by the English language development test identified or developed pursuant to Section 60810, or any successor test, and scores far below basic or below basic on the English language arts standards-based achievement test administered pursuant to Section 60640, or any successor test." (California Educational Code § 313.1, 2012)

Adding multiple sources of evidence does not guarantee that overall validity will be improved. When multiple scores or tests are used to make a decision, the combination rules matter. There are two ways to combine multiple measures: composite and aggregate. A composite is when two or more scores are weighted and summed together into one score, e.g., subtests into an overall score, so that a performance standard can be applied for use in decision making. An aggregate is when two or more measures are evaluated in a concurrent or stepwise fashion towards one decision, e.g., criteria for entering a gifted program. Composites and aggregates follow one of four combination models: conjunctive, compensatory, complementary, or mixed. These models adhere to theories of how the measures relate to one another and often seek to minimize one type of error over another, e.g., minimizing false positives while tolerating false negatives. Within the chosen model, a decision rule is applied to create the combination to which the performance standard is applied. For composites, a decision rule establishes the weighting and summing procedure. For aggregates, a decision rule establishes whether the combination will be concurrent or stepwise, and if stepwise what the ordering of the data will be, how banked scores will be considered, and other decision-specific procedures.

Context

For all language minority students in California with home language influences other than English, the California English Language Development Test (CELDT) is given as an initial screener and, for English Learners, as an annual test to determine readiness for program exit. The composite decision rule is the same for both purposes (see Table 1). This is set to change in 2017-18 when two separate assessments, a new screener and a new annual test, will be introduced.

Table 1

State Composite Rule (Proficient): California English Language Development Test

Grade	Model	Overall Performance Level	Weights for Overall Performance Level				Subtest Performance Levels
			Listening	Speaking	Reading	Writing	
K & 1	Conjunctive	4 or 5	0.45	0.45	0.05	0.05	Listening: 3-5 Speaking: 3-5
2-12	Conjunctive	4 or 5	0.25	0.25	0.25	0.25	Listening: 3-5 Speaking: 3-5 Reading: 3-5 Writing: 3-5

In annual classification decisions to determine readiness of ELs for program exit, California follows an augmented classification model (Abedi, 2008a) where classifications from the CELDT are considered first and followed by other sources of data (see Figure 1). Since California is a state which allows Local Educational Agencies (LEAs) to set the criteria performance levels on the CELDT and other criteria, these state decision rules are to be interpreted as suggestions, not as mandates.

According to the California Department of Education, the CELDT is not meant to stand alone for student-level decisions. As stated in the 2011-12 CELDT training presentation:

> CELDT results can provide information at varying levels of analysis: district, school, grade, and individual student. At the student level, CELDT results must be interpreted with great caution and in conjunction with other information, such as data from teacher observations and checklists, classroom assessments, and other tests of English language proficiency.[2]

[2] From *CELDT 201A: Reports and Data Management 2011-12 Presentation Transcript*, page 4, slide 18.

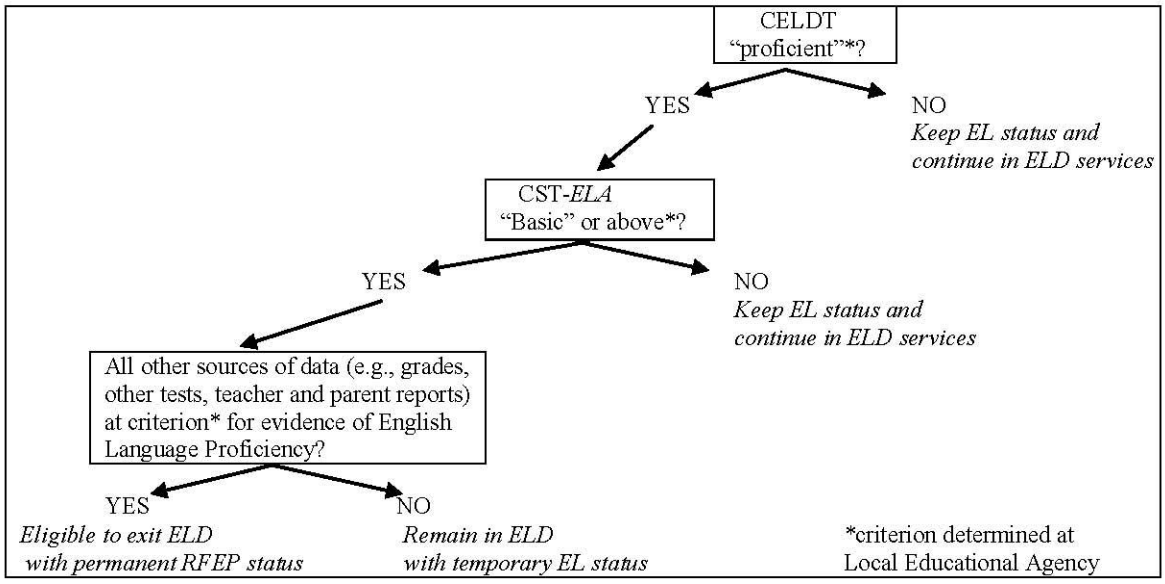


Figure 1. Aggregate Decision Rule, EL Reclassification in California Grades 2-12

In the aggregate decision, the California state education code stipulates the use of other criteria in addition to the CELDT when determining whether to reclassify a student *out* of ELD (e.g., California Educational Code 613 and 60810). This conjunctive model requires all criteria to be met before a student can exit ELD. While this system is intended to ensure that students receive the benefit of EL programming as long as needed (Linquanti, 2001), it is unclear whether adequate checks and balances exist at the state or district level to verify the precision of each performance standard and decision rule. Without adequate precision, the likelihood of misclassification is greater with multiple measures in a conjunctive rule than it would be with a single measure. This has bearing on the decision of which students remain in ELD. For, depending on the decision rule applied, a single criterion could be used to keep a student *in* ELD. In the case of California's composite and aggregate decision rule, the gatekeeper could be a CELDT subtest score, or an overall score, where a student missed a criterion cutoff by one raw score point.

To date, there is no uniformity in how states and districts report their decision rules and criteria. Few studies have investigated the impact data use in decision making for EL status changes. No known studies to date have examined ELPA evidence within the context of district policies and procedures for educational programming of EL students (EL), EL students with disabilities (EL/SWD), or ELs who are high-performing on state standards-based assessments (EL/HP) including EL students with disabilities who are also high-performing (EL/SWD/HP). Poorly conceived data use practices in EL programs can have a negative impact which is costly to districts (e.g., the price of ongoing testing for students who may not need it), schools (e.g., the misallocation of services for long-term EL students) and students and their families (e.g., the cost of missed opportunities; Gándara, Rumberger, Maxwell-Jolly, & Callahan, 2003). Thus, data use research in local control states should describe practices and outcomes at the local level.

Purpose

The purpose of the current study is to explore the decision rules within a K-12 district in order to understand possible antecedents to the long-term English learner phenomenon in California. This study explores the use of assessment data for making classification decisions about language minority students, specifically EL, EL/SWD, EL/HP, and EL/SWD/HP students. It examines local decision rules, composite and aggregate, to describe potential differences in redesignation outcomes. The study also reports actual outcomes of the rules implemented by the district in 2012-13, and then describes the students who would, as a result, remain in ELD programs for the purpose of connecting decision rule choice with data use outcomes.

Significance

This research adds to the understanding of data use in this era of accountability, the impact of models and decision rules on high-stakes decisions about individual students, and the need for school-level and grade-level sensitive policy guidance in EL assessment systems. Findings from this study have implications for policy and practice for data use in systems involving multivariate classification.

Research Questions

This dissertation study expands upon prior work by Carroll (2012) and uses a mixed methods approach to examine the impact of composite and aggregate decision rules in determining educational programming for language minority students. The research questions are:

(1) *When applied to actual student data, to what extent do composite decision rules classify students as "proficient" or "non-proficient? To what extent do composite decision rule outcomes differ among EL, EL/SWD, EL/HP, and EL/SWD/HP students?*

(2) *To what extent do aggregate decision rules classify students as eligible/ineligible for program exit? To what extent do aggregate decision rule outcomes differ among EL, EL/SWD, EL/HP, and EL/SWD/HP students?*

(3) *As a result of the current district aggregate rule, what is the composition of the group of students who would remain in ELD programs?*

(4) *What are the influences and constraints on data use routines for initial classification, reclassification and educational programming decisions? What composite and aggregate decision rules are used at the local level?*

CHAPTER TWO

REVIEW OF RELEVANT LITERATURE

This study addresses the phenomenon of long-term English learners by examining influences and constraints on multivariate data use routines affecting language minority students and the educational programming they receive. Considering the complexity and breadth of literature which has a bearing on this topic, the following review of literature seeks to address topics related to classification models and decision rules, data use in the context of English Learner programs, data use for students experiencing reading difficulties, and studies which address English learner classification including decision rules.

Classification Models and Decision Rules

A classification model is a guiding principle for combining two or more measures, subtests, or indicators to which a performance standard and decision rule for classification is applied. All decision rules used in classification of EL students can be organized under one of these four models:

- Conjunctive (all indicators at the standard)
- Compensatory (uneven indicators, some allowed below the standard)
- Mixed (two models combined)
- Complementary (one of two possible indicators at the standard)

All decision rules operating under these models are either *composite*, for combining sub-tests towards an overall classification such as reclassification as ELP "proficient," or *aggregate,* for combining separate indicators (e.g., ELPA and academic content assessments) for a multiple-

measures decision such as redesignation as Fluent/Fully English Proficient (FEP) required in some states and/or districts. Impact from decision rule choice has been reported for composites which combine multiple subtests, such as the General Education Degree test (Douglas & Mislevy, 2010), and aggregates which combine multiple indicators, such as fourth grade promotion (Chester, 2003) and eligibility for gifted programs (McBee, Peters, Waterman, 2014).

The current study describes the impact of model and decision rule choice in one district using the CELDT. I explore both the composite combination of subtests in terms of which students are classified as English Language "proficient" or "non-proficient," and the aggregate which uses multiple measures to determine which students are eligible/ineligible for redesignation as Fluent English Proficient (FEP).

Whether overt or tacit, ELPA classification models adhere to different theoretical stances on second language acquisition. It is, however, important to examine the theoretical underpinnings of an assessment as well as its intended uses (Bailey & Heritage, 2010) and to make these assumptions explicit in the test development process. Thus, a compensatory ELPA model operationalizes English language proficiency as compensatory in nature such that strong performances on one or more subtests will compensate for slightly weaker performance in other subtests. Compensatory models may minimize false negatives (proficient students mistakenly classified as non-proficient), but may allow more false positives (non-proficient students mistakenly classified as proficient). Among these false positives there could be students with an especially uneven profile of proficiency where one subtest performance is very weak and all other subtest performances are strong – seen as an especially undesirable outcome. This possibility has thus been used as a rational for avoiding compensatory models across a number of contexts (Abedi, 2004; Clauser, Clyman, Margolis & Ross, 1996; Hambleton, Jaeger, Plake, &

Mills, 2000; Hambleton & Pitoniak, 2006). Even with the possibility of false positives, there is reason to believe that extremely uneven profiles are less prevalent (Carroll & Bailey, 2012) especially when subtests are highly correlated. Since other measures could help minimize the likelihood of false positives during redesignation, compensatory models are desirable to minimize false negatives.

A mixed ELPA model operationalizes English language proficiency as compensatory in nature with the caveat that all subtest performances must meet or exceed a minimum level (e.g., California, Overall Level 4 with all subtests at Level 3 or above). Mixed models have been suggested as preferable to the compensatory model in some contexts (e.g., Clauser, Clyman, Margolis & Ross, 1996) but no known studies have examined the cost benefit of mixed ELPA models. Mixed models may less reliable due to their complexity. For example, California's mixed model relies on eight cut-scores[3] which, depending on the variance, weights, reliabilities and intercorrelations of the subtests, could result in an attenuated composite reliability (Mosier, 1943; see also Livingston & Lewis, 1995; Wang & Stanley, 1970).

A conjunctive ELPA model operationalizes English language proficiency as non-compensatory, thus only achievable when all subtest performances reach equally high levels. Conjunctive models eliminate the possibility of uneven proficiency profiles, but may create more false negatives than compensatory or mixed models as has been observed with other assessments (e.g., GED test, Douglas & Mislevy, 2010).

[3] For Grades 2-12, the eight cut-scores are: listening, speaking, reading and writing cut points for level 3 (4 cut scores) and level 4 (4 cut scores) in order to determine "proficient = OPL 4 (an unweighted sum of the scale scores of all 4 subtests) plus no subtest lower than level 3". For K-1, eight cut-scores are used plus an additional complexity of weighted components (L=.45, S=.45, R=.05, W=.05). A table of cutscores is provided in Appendix F.

Federal Guidance

Under the Elementary and Secondary Education Act (ESEA) of 1994, a compensatory approach had been applied to ELPAs, but under the reauthorization of ESEA (known as the No Child Left Behind Act, 2001) the conjunctive model was established as the preferred model (see Abedi 2004, 2007). In 2008, the United States Department of Education again allowed compensatory models as long as states could demonstrate that the model meaningfully measured proficiency[4] and also provide evidence of valid interpretations and uses of results. In local control states such as California, this means that local education agencies could hold the burden of proof for demonstrating validity of inferences from their chosen models. Should evaluations occur, the level of rigor is set by test vendors and state departments of education.

Validity and Fairness

Evidence created from assessment data involves both the true score of the construct being measured, as well as error: construct-relevant and construct-irrelevant variance in an unknown proportion (Allen & Yen, 2002). The goal of measurement professionals is to develop assessments, evidence, and classification systems which minimize construct-irrelevant variance in order to approximate the inferences of the scores for the construct (AERA, APA, & NCME, 2014) so that "the extent to which decisions based on test scores match decisions that would have been made if the scores did not contain any measurement error" (Hambleton & Novick, in Zhang, 2010, p. 126). In the EL context, this applies to the aggregate as well as the composite classifications.

[4] "A State can use a composite score so long as the State can demonstrate that the composite score meaningfully measures student progress and proficiency in each of the language domains and, overall, is a valid and reliable measure of student progress and proficiency in English, consistent with the purpose for which the assessment is used" (USDE, Federal Register, Nov 18, 2008.)

According to Kane, validity is not a property of the test, but rather a property of "the proposed interpretations and uses of the test scores" (p.3, 2013). Determining the validity of overall levels and classifications is a quest to validate interpretations, not the ELPA itself, and to do so within the context of data use. From raw scores to redesignation, there are several points where ELPA validation is needed. To determine proficient and non-proficient classifications, test consortia and individual states choose models and decision rules, set subtest weighting, set performance standards on subtests, create composites of ELP, and set an overall performance standard. To determine readiness for redesignation and exit from EL services, state policymakers choose criteria, set performance standards, and prioritize criteria in order of decision use. All of these choices allow for discrepancies between and even within states which directly affect the redesignation eligibility of individual students. Furthermore, these choices adhere to theories of language acquisition, language development, and proficiency. When applied in a multivariate assessment, classification models and decision rules embody and enforce these theories.

Closely related to validity is the notion of fairness. Substantive fairness in testing, as defined by Kane (2010), "requires that the score interpretation and any test-based decision rule be reasonable and appropriate, and in particular, that they be equally appropriate for all test takers (at least roughly)" (p. 178). If fairness cannot be assured, then the validity of score use comes into question. In this way, validity and fairness – especially for high-stakes decision making – are intertwined (Kane, 2010; see also Baker & Linn, 2004). When assessments are used to determine student readiness for certain educational programming, lack of substantive fairness can lead to differential outcomes. For example, when EL students with the same level of the underlying construct of English Language Proficiency are classified differently, this may lead

to different educational experiences. Within this scenario, if some EL students are more likely to be classified "non-proficient" than others despite comparable levels of English proficiency, we can say that the classification lacks substantive fairness for that certain group of students. In addition, the impact and consequences of score-based decisions on test taker groups need to be investigated to see if the use of the test incurs any non-comparable consequences on different test taker groups (Xi, 2008, 2010; Kane, 2010). Investigating decision rules is a necessary aspect of any investigation of fairness as it may illuminate differences in the extent to which models and rules influence student outcomes.

Data Use in the Assessment System for English Learners

Data use in schools serves two main purposes: accountability and improvement (Weiss, 2012). ELP assessment data is used to satisfy both purposes as shown in Table 2.

Table 2

Accountability and Improvement Uses of ELP Levels and Classifications

Evidence	Accountability at the State/District Level	Improvement (Instructional Programming) at the Child Level
Initial Classification	(none)	EL Initial Program Entrance/Excusal: Determination of LEP status or IFEP status
Annual Classification	AMAO 2: Report Proficiency	EL Annual Program Continuation/Excusal: Eligibility for status change to FEP
Overall Performance Level	AMAO 1: Report Progress	EL Within-Program Placement: Type, Pace, Intensity

Note. AMAO = Annual Measureable Achievement Objective.

At the state/district level, the ELPA Overall Performance Level (OPL) is used to report how many students grew, declined, or maintained one level. ELPA classifications are used for accountability purposes to report how many students achieved the ELPA criterion of "proficient". At the child level, OPL is used as evidence for educational programming for

improvement; for example, an OPL 2 can infer the student would be best served in an ELD 2 class[5]. Classifications are used as evidence to infer the need to remain in, or eligibility to exit from, EL programs. There is currently no federal accountability around the use of ELPA data for placing students in an ELD program.

Data use in schools in the current environment of accountability rely not only on assessment data, but the gathering of other evidence in a process of what ideally can be seen as data-driven, student-centered decision making. With multiple decision-makers involved, it is true that "evidence-use processes are fundamentally a social sense-making process" (Honig & Venkateswaran, 2012) and the scope of the demands on data use vary according to level of leadership and responsibility. While school staff may be called upon to make decisions using data, often there are "designated colleagues to make evidence available to them" or so-called "brokers" within the system (Honig & Venkateswaran, 2012). The constraints on data use in schools are best described as organizational and political in nature (Coburn & Turner, 2011) but when decisions also involve data and technology, additional forces can be at work (Piety, 2011).

Data Use for Decision Making about Language Minority Students

ELPAs were created to satisfy the mandate of NCLB regarding measurement of the progress and proficiency of EL students. ELPA data are used by schools to identify which language minority students need to be placed in ELD, when ELs will exit from ELD, and the educational programming ELs will receive while receiving ELD.

Studies about initial classification are few. Bailey and Kelly (2012) conducted a systematic review of identification practices and found most states use some form of a Home

[5] Although class titles may differ, most states align services with ELPA test levels. In addition, most states recommend no more than 2 ELD levels combined in one EL class setting to ensure instruction can match the level of a student's language acquisition.

Language Survey to identify students as "language minority" and needing additional screening or assessment to determine need for language support services. However, the wide variety of practices and non-systematic use of non-standardized surveys suggests that students identified for testing in one state would not be identified for testing in another, with obvious implications for initial classification. According to the National Research Council (2011), state practices also vary in terms of which screener or assessment is used. Some states use the same ELPA used for annual testing (e.g., California) while most others use a brief screener or placement test. State policies also vary with a majority of states mandating the use of a single test, either a screener (27 states), or their own ELPA (4 states). The remaining states allow districts to choose between their ELPA and a screener (2 states), or to select their own screener, often from a pre-approved list (17 states; NRC, 2011, p. 84). Standard-setting procedures applied to these assessments determine a twofold educational placement: Students who meet or exceed the standard are designated IFEP, and receive instruction with non-EL students, and students who fail to meet the standard are designated Limited English Proficient (LEP) and provided with EL services. There is a dearth of research on the efficacy of standards to determine which language minority students need EL services, although some studies suggest the potential for over-identification of already proficient students as LEP (see California Department of Education, 2011).

When students are ready to exit ELD programs, there is a determination of readiness to receive all instruction in English-language medium settings. Reclassification has been the focus of most validation studies investigating ELPAs and mostly with an eye on accountability: specifically, looking at the ability of ELPA "proficient" to predict future academic success. Such studies have examined the outcomes of reclassified students in comparison to English only students (Crane, Barrat, & Huang, 2011), in comparison to other reclassified students with higher

or lower eligibility levels for RFEP (Kim & Herman, 2010), and in comparison to students who did not reclassify (Abedi, 2008a; Grissom, 2004). Studies focused on policy have explored methods for setting reclassification standards (e.g., Robinson, 2011) and performance standards for accountability targets related to reclassification (e.g., Cook, Boals, Wilmes, & Santos, 2008).

Few studies have examined academic outcomes for students remaining in EL programs despite high levels of reclassification readiness, with one exception being the current work of Estrada (2014). In the year one findings of this reclassification study, researchers found EL students at all grades, but especially third grade, who were prevented from reclassification due to ELPA evidence despite strong performances on all other criteria. These students were subsequently kept in educational programming below their abilities, limiting their chances to access pathways, or *curricular streams* (Estrada, 2013) that would give them the best chance at college readiness.

ELPA use for students experiencing reading difficulties

Reading difficulties, whether delays or disabilities, play a major role in assessment outcomes for language minority students (Abedi & Gándara, 2006) and may impact the special education referral process itself (Klingner & Harry, 2006). School psychologists use special matrices for interpreting language minority students' levels of language proficiency and literacy development (see Kranzler, Flores, & Coady, 2010) yet research suggests that many school psychologists feel ill-equipped to interpret results (NASP, 2007; Kritikos, 2003). Thus, EL students may be at risk for misinterpretation of low reading skills as simply a language acquisition problem rather than a developmental delay or potential deficit. In the general student population most referrals to special education are for reading difficulties (approximately 80%; Fletcher, Lyon, Fuchs, &

Barnes, 2007) and some research suggests the same is true in the EL population (e.g., 95% of referrals for K-5 ELs; Del Rio, 2010). While language minority students are protected under the law against discrimination based on cultural differences[6], decisions based on assessments in English – especially those related to reading – can still pose a risk to fairness and equity (Abedi & Gándara, 2006). According to Ragan and Lesaux (2006), "many language minority learners in the primary grades may benefit more from intervention services for struggling readers than from EL programs with emphasis on oral language proficiency" (p. 19; see also Haager, 2007).

Classification accuracy on an ELPA is especially high-stakes for those EL students who are at risk for reading difficulties due to factors other than home language influences. If EL student performances are poor due to other factors, crucial educational services could be missed. Furthermore, classification accuracy is particularly high-stakes for struggling readers transitioning into secondary school settings (Kim, 2011; Walqui et al., 2010). In sixth grade and beyond, classes designed for LEP students can supplant grade-level English Language Arts instructional minutes, or even intervention reading programs, leaving students with less opportunity to engage with appropriately-leveled content and vocabulary (Estrada, 2014). EL programming is often co-scheduled with remedial classes in other content areas in such a way that students classified LEP are often clustered into low-performing, within-subject tracks for the entire instructional day (Estrada, 2014; Menken, 2008) diminishing chances of equal access to high-quality instruction (Callahan, Wilkinson, & Muller, 2010). For students who are high-achieving in math, despite experiencing difficulties in reading, this pattern of educational programming impacts high school and college readiness (Kim & Herman, 2012; Kim, 2011).

[6] When defining LD in terms of culturally and linguistically diverse students - including EL students - the "exclusionary clause" in the Individuals with Disabilities Education Act of 1997 states: "The term [LD] does not include learning problems that are primarily the result of visual, hearing, or motor disabilities or mental retardation, of emotional disturbance, or of environmental, **cultural,** or economic disadvantage." (34 §C.F.R. 300.7(c)(10)ii)

ELPA studies addressing classification

With so much depending on classifications, surprisingly few studies have investigated ELPA classifications themselves. Studies related to ELPA have investigated the test development process (Garcia, Lawton & Diniz de Figueiredo, 2010; Davidson, Kim, Lee, Li, & López, 2007), the constructs tested (Bailey, 2007) the impact of test formats and measurement models (Zhang, 2010), the impact of cut scores (Florez, 2012; Wang, Niemi & Wang, 2007), the predictive ability of classifications as readiness indicators for federal reporting of progress and proficiency (Abedi, 2008b; Robinson, 2011) and the performance of students who have been reclassified FEP (Abedi, 2008b; Kim & Herman, 2010; Ragan & Lesaux, 2006). In terms of classifications, ELPA studies have been conducted by test developers and consortia to report technical qualities for state Title III accountability reporting, but not necessarily for individual student-level decision making.

Two studies were conducted by a test developer analyzing the performance of EL and non-EL students (n=400) on the Stanford English Language Proficiency assessment (SELP, 2003). In the first study, Stephenson, Johnson, Jorgensen, and Young (2003) used Analyses of Variance (ANOVA) to compare the performances of EL and non-EL students grouped as "proficient" and "non-proficient" by the SELP on the SAT9 reading comprehension subtest. They claimed validity of SELP classifications was justified by evidence that students classified "non-proficient" scored lower on average on the SAT9 than students classified "proficient." There was no examination of individual-level performance data for anomalies such as a "non-proficient" SELP classification converging with a high SAT9 performance. Further investigation to rule out any such counterevidence will be necessary in order to support the warrant that an ELP classification inference accurately and consistently identifies ELP at the individual level.

In a related study, Stephenson, Jiao and Wall (2004) again used ANOVA to claim validity of SELP classifications based on evidence that EL students scored lower on average than non-EL students. However, discriminant analyses by group membership in the same study indicate some non-EL students were classified "non-proficient" (by grade cluster: primary, 36 percent; elementary, 28 percent; middle, 13 percent; and high, 17 percent). The authors cited this as evidence that classifications at primary and elementary levels are "not very reliable" (p. 8). Unfortunately, these findings were not accompanied with recommendations for how classifications should be interpreted when used in decision making for individual EL students. Again, there was no investigation of classification model itself or its potential contribution to lower levels of reliability and validity.

The California Department of Education (CDE, 2011) used EL and non-EL student performances to investigate validity of reading and writing subtests on the CELDT in kindergarten (non-EL, *n*=1,386; EL, *n*= 4,350) and first grade (non-EL, *n*=495; EL, *n*= 3,985). This was part of an assessment development process to include and validate inferences from reading and writing subtests newly added to the K/1 CELDT (see also García Bedolla & Rodriguez, 2011). Non-EL students scored higher on average than EL students which CDE cited as evidence that CELDT scores are valid for allocating EL support, even though those decisions depend on CELDT classifications, not scores. In addition, researchers cited group membership predictability as validity of classification accuracy as the majority of EL students fell below the "early advanced" cut point (K, 94%; 1st, 92%). However, a majority of non-EL students also fell below the "early advanced" cut point (K, 74%; 1st, 50%), a finding which was not interpreted in the study. The CDE report noted that if the CELDT were used as a stand-alone criterion "a great many more students would be receiving [English language development] services" (p. 46)

implying that the K/1 CELDT with its current decision rule can over-identify proficient students as non-proficient. No recommendations were made for interpreting these classifications even though districts do rely solely on the CELDT to determine which language minority students should receive English language services.

Although some studies have suggested methods for setting an ELP performance standard, there has only been mention of the standards' sufficiency for state Title III accountability. Validity for individual student-level decisions may be implied, but it has not been assured. Linquanti and George (2007) present California's approach to establishing annual measurement achievable objectives (AMAOs) to measure and report on progress towards meeting Title III accountability goals. Within the development of AMAO 2 used to measure student attainment of ELP (i.e., percent proficient), the authors report efforts to operationalize and test decision rules to set the state ELP performance standard. Three rules were created: two mixed and one conjunctive. Although the method and sample were unspecified, Linquanti and George reported a "relatively small difference in outcomes" between the mixed rules which were said to be "substantially more challenging to attain" than the conjunctive (p. 109). Based on this finding, researchers recommended the first mixed rule as "sufficiently rigorous" (p. 109) as an ELP performance standard for use in the development of AMAO 2. The report did not define nor interrogate the sufficiency of the ELP standard for other uses, such as determining which students should continue to receive ELD services.

Elsewhere, Cook (2009) has outlined how the state of Kentucky set its ELP performance standard for AMAO 2. For this process, ELP was defined as "the point where students' language proficiency level becomes less related to academic achievement" (p. 4). ACCESS for ELLs™ and the Kentucky Core Content Test (KCCT) tests of reading and mathematics were used to

analyze the relationship between ELP and academic achievement, respectively. Kentucky's overall ELP composite used weights provided by the WIDA consortium (listening*0.15 + speaking*0.15 + reading*0.35 + writing*0.35) creating a continuous variable along which a point could be chosen to define the ELP performance standard. To identify where the relationship between ELP and reading/math became less related, researchers first conducted a correlation then a decision consistency analysis. Six decision rules were chosen: three mixed and three conjunctive. To test the impact of each decision rule, a logistic regression was conducted to predict the likelihood of reading and math proficiency for students meeting or exceeding the ELP standard set by each rule. Results were presented in clusters for grades 3-5 and 6-8, and individually for grades 10 and 11; students in Kindergarten, grades 1 and 2, and those taking the newcomer Tier A test were excluded. Based on the discussion of these results, the expert panel chose a mixed decision rule which classified the fewest students as ELP proficient, 20 percent less than one alternative, yet offered the highest percentage predicted to be proficient in reading and math. However, in relation to impact, students deemed "non-proficient" by the chosen standard were not described, nor was their level of reading or math predicted. This omission suggests the possibility of students with actual reading and math proficiency, and the language ability to achieve that result, being classified ELP non-proficient in error as only half the data, in effect, was examined.

More recent recommendations to policymakers have acknowledged that setting ELP criterion too high can negatively impact exit decisions. Cook, Linquanti, Chinen, and Jung (2012) offer three approaches to setting an ELP performance standard empirically connected to academic content assessments. They warn that a criterion set too high "will be less predictive of content proficiency at higher ELP levels" (p. 9) and this imprecision adds greater error which

could lead to inaccurate inferences. The authors note that such imprecision might also lead to policymakers to establish higher criterion on content assessments, thus raising expectations for EL exit decisions beyond what is currently attained by mainstream non-ELs. Of concern to the current study is the impact of these methods on inferences used for decision making at the individual level. While some imprecision may be tolerable for Title III accountability at the state level, imprecision leading to the misclassification of individual students in terms of their English language services is a policy outcome that also requires a method of evaluation.

None of the studies described above included evidence to determine classification accuracy at the individual level, yet measurement professionals broadly concur that determining classification accuracy requires an analysis of decision rules, not simply cut scores (Kane, 2013). No known ELPA studies to date have investigated classification accuracy at individual grade levels since NCLB was enacted in 2001. The paucity of evidence for classification validity emphasizes the need for more carefully constructed validation studies for ELPAs (Bachman & Palmer, 2010; Forte & Willhoft, 2010; Sireci, Han & Wells, 2008). Furthermore, while ELPA evidence may reveal something unique about a student's language acquisition that is not revealed in other measures, that revelation may also indicate difficulties not related to language minority status. Thus, classifications should be judged with careful consideration of measurement limitations and potential alternate explanations.

Prior research on ELPA decision rules

Four empirical studies conducted by Carroll and Bailey (2012, 2013a, 2013b, 2014) have examined decision rules as applied to ELPA using EL and non-EL student performances. In a model comparison study, Carroll and Bailey (2012) analyzed differences between a conjunctive and a compensatory model using a subgroup of fifth grade EL student performances (n=875).

The authors examined the prevalence of uneven profiles in the compensatory model alongside the prevalence of false negatives (proficient students classified as non-proficient) by compensatory and conjunctive models. While the compensatory model produced some uneven profiles (*n*=15) it created far fewer potential false negatives (*n*=162) than the conjunctive model (potential false negatives, *n*=400). This finding suggests the need for a well-reasoned model choice that takes into account the potential effects of false-negatives and uneven profiles in decision-making about individual EL students.

Two other model comparison studies were conducted by Carroll and Bailey: one preliminary analysis of conjunctive, compensatory and mixed models on a subgroup of fifth grade EL and non-EL students (2013a) and sixth, seventh and eighth grade EL and non-EL students (2013b). In both studies, the conjunctive model currently in use by State A was found to classify the most students as non-proficient in the EL group (5^{th}, 6^{th}, 7^{th}, 8^{th},) and non-EL (5^{th}, 6^{th}, 7^{th}, 8^{th}). By contrast, compensatory and mixed models classified far fewer students as non-proficient. Additional analyses of high-performing students showed that EL and non-EL students with the highest scores on state standards-based assessments were classified non-proficient in larger numbers under the conjunctive model than the compensatory and mixed models. A fourth study (Carroll & Bailey, 2014) explored model effects on new grades (K-4) while including 5^{th} grade to explore grade cluster effects (3-5) and adding a second conjunctive model to more fully explore model and decision rule effects. As seen in prior work, results from models and decision rules differed in the extent to which students were classified ineligible for redesignation. However, the two conjunctive models differed the most: the conjunctive-1 created the largest pool of ineligible students overall (EL, *n*=1740, 59.0%) and in each grade, as high as 61.2%, while the conjunctive-2 created the smallest pool of ineligible students overall (EL, *n*=793,

26.9%) and in each grade, as low as 17.7%. These findings suggest that the decision rule, not just the model, deserve scrutiny.

While these studies have illuminated the impact of models and rules, there is one more layer of influence which is at the local school district. The missing piece in the literature so far, what this study seeks to remedy, is an exploration of the impact of local decision rule choices on educational experiences and outcomes of language minority students. Beyond an examination of student performance data writ large, the current study makes key contributions in two phases: examining student data to which both composite and aggregate rules are applied, with a description of students remaining in ELD programs as a result of the district-chosen aggregate; and examining local decision rule practices and policy implementation to describe the extent to which local classification routines determine which students receive ELD.

Theoretical Framework: Overall Approach & Rationale

Rationale for using a mixed methods approach

This dissertation takes a pragmatic, concurrent mixed methods approach to describing the use of the CELDT in determinations of placement in, placement within, and exit from EL programs. A concurrent, mixed-methods approach was employed to gain a broader perspective from different types of data (Creswell, 2008).

Prior research by the author on decision rules in EL assessment systems has taken a quantitative approach which described classifications based on the state-mandated decision rule and results from prior years gave reasonable confidence that the state rule had been followed. However, in local control states like California where state decision rules are presented as guidelines, alignment with state policy at the district level cannot be assumed. The current study

seeks to more deeply understand data use in terms of district-level decision rules which are not necessarily a matter of public record. Thus, qualitative data was needed to describe district and school level decision rules within the EL assessment system. Quantitative data was used primarily to examine the impact of decision rules on student-level outcomes, e.g., status change to RFEP. Quantitative data were also used to describe the incidence of certain trends/patterns in the data which could inform a broader discussion of the theoretical "fit" of models and decision rules in the EL assessment system. Qualitative data was used to help identify and in some cases further interpret trends/patterns in data in relation to participants' reflections on data trends. For example, participants expressed concern over the number of students who seemed "stuck" at ELD level 3 and some participants identified a pattern of performance where students surprisingly received a "2" in listening when all other subtests were at a "5". While the "stuck at 3" trend was already known from prior literature the additional knowledge of certain low-performing subtests allowed closer examination of the influence of these trends. Poor performing subtests are more injurious to test fairness under certain decision rules, so this method of concurrent/exploratory interpretation was both important and efficient to achieve research aims.

For these reasons, this study takes a concurrent mixed methods approach as displayed in Figure 2. Quantitative and qualitative data were collected and analyzed concurrently. Quantitative results and qualitative results were considered together to provide nested interpretation of quantitative results within the qualitative interpretation in order to fully contextualize interpretations and implications for individual, school, district and state levels.

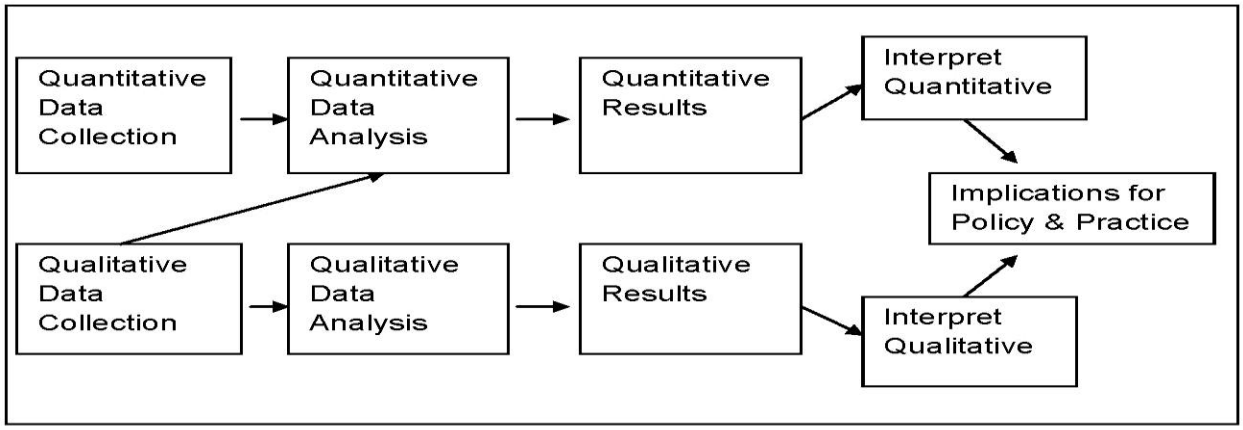


Figure 2. Procedural Diagram: Mixed Methods Concurrent Design

Quantitative Lens

Pragmatically, assessment data should produce valid and reliable inferences for every purpose and use, especially data mandated for use in making individual-level decisions. As a whole, the entire accountability and assessment system in schools deserves investigation to ensure that all the decision rules are operating as they should in the greater scheme of ensuring academic outcomes for language minority students (Bailey & Carroll, 2015). In terms of ELP assessments, models and decision rules deserve examination as part of a comprehensive process of evaluating the validity of classifications (Carroll & Bailey, 2015).

Qualitative Lens: Framework for Data Use

Coburn and Turner (2011) designed the "Framework for Data Use" for evaluating data use interventions in terms of student outcomes. Their framework considered tools to be an input and holds the processes of data use as central. Piety (2011) proposed extending the Coburn and Turner framework from a sociotechnical perspective which examines infrastructure and tools in terms of interactions that influence the processes of data use. Piety's framework keeps social processes of data use central, yet accounts for how systems have become encoded with tools and

infrastructures by overlapping policy and tools upon data use yet leaves out accountability policy. Both frameworks informed the coding and analyses of influences and constraints on data use routines, but Coburn and Turner (2011) was the main organizing idea (see Figure 3).

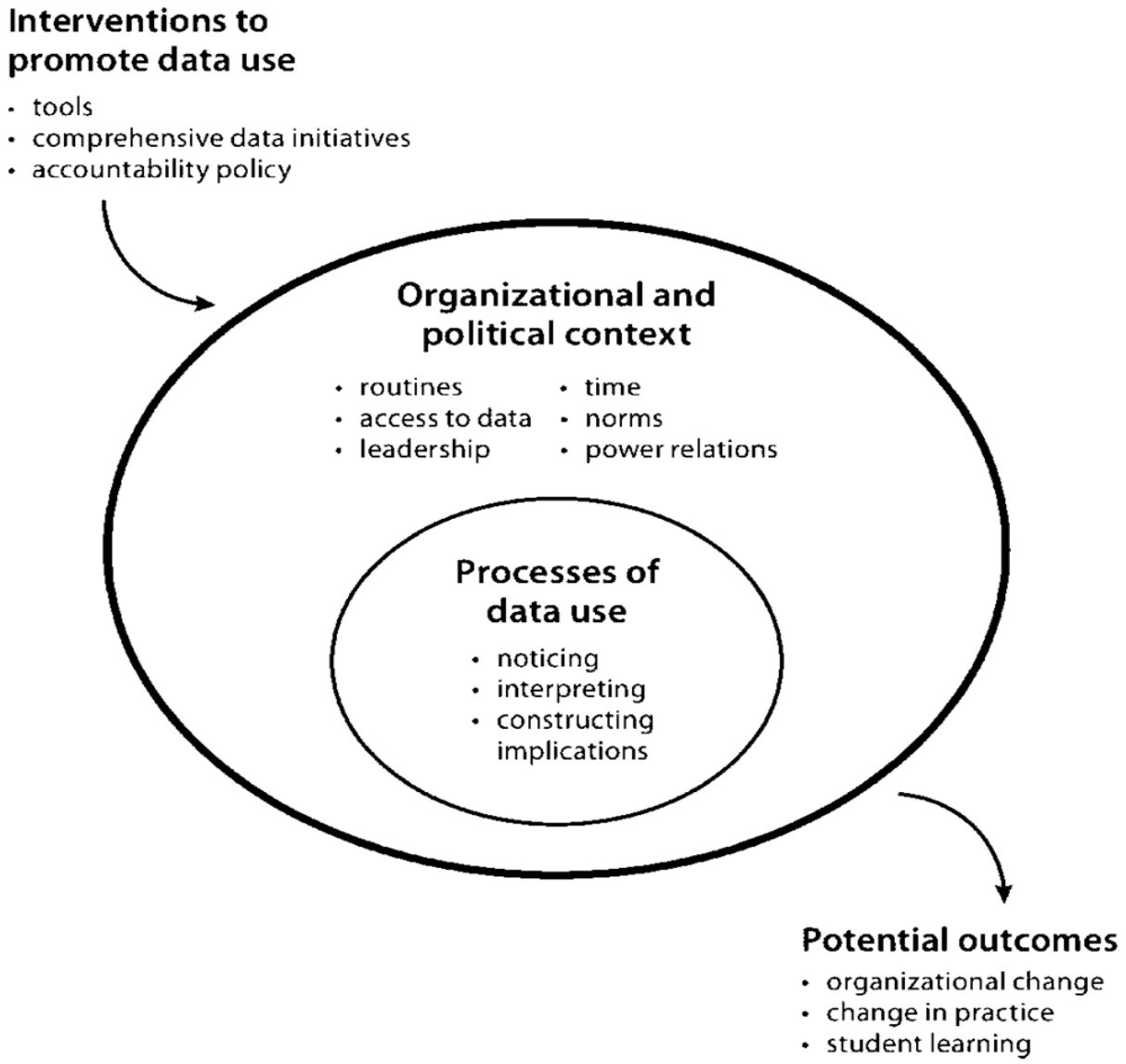


Figure 3. Framework for Data Use. From Coburn, C.E., & Turner, E.O. (2011). Research on data use: A framework and analysis. *Measurement: Interdisciplinary Research and Perspectives, 9*(4), p. 178. Used with permission of the authors.

Research Questions

High-stakes decisions made about students should be made with multiple sources of evidence deemed valid and reliable for each decision. Prior studies have not addressed the extent to which composite and aggregate decision rules impact outcomes for students, especially in terms of long-term placement in ELD programs. This study takes a descriptive, exploratory, mixed-methods approach to allow qualitative findings to enhance understanding of the quantitative results and lead to a deeper understanding of the potential impact of models and decision rules on outcomes for EL students. I use district and school personnel interviews, state and district documents, and student assessment data to answer the following:

(1) *When applied to actual student data, to what extent do composite decision rules classify students as "proficient" or "non-proficient? To what extent do composite decision rule outcomes differ among EL, EL/SWD, EL/HP, and EL/SWD/HP students?*

Based on findings from prior studies, I hypothesized that mixed models would be a more stringent composite for all student groups.

(2) *To what extent do aggregate decision rules classify students as eligible/ineligible for program exit? To what extent do aggregate decision rule outcomes differ among EL, EL/SWD, EL/HP, and EL/SWD/HP students?*

Based on findings from prior studies, I hypothesized that the more stringent aggregate models would lead to the largest percentage of students, including high-performing students, remaining in ELD programs in comparison to other decision rules.

(3) *As a result of the current district aggregate rule, what is the composition of the group of students who would remain in ELD programs?*

Because the actual aggregate is based on additional factors, I hypothesized that students who were not reclassified would be within the range of proficiency best served by the ELD program and would exclude the students with the highest performances on CELDT and on the state standards tests.

(4) *What are the influences and constraints on data use routines for initial classification, reclassification and educational programming decisions? What composite and aggregate decision rules are used at the local level?*

I hypothesized that state and district policies and procedures would influence the use of data during these classification and programming routines, as well as aspects of the infrastructure and assessments themselves. It seems that federal accountability may also be a constraint on the timing of these routines. Regarding decision rules, I hypothesized that variations could exist at the school level.

CHAPTER THREE

STUDY METHODOLOGY

This chapter presents the methods used for two types of data collection and analysis, qualitative and quantitative. Data were collected from state and district documents, 18 semi-structured interviews, and one year of student data to explore data use practices related to the CELDT assessment related to EL programs. This chapter provides a description of the method chosen, the overview of study procedures, and the sample.

Mixed methods

Mixed methods approaches to research combine quantitative and qualitative approaches to data collection and analysis when the use of either approach alone would fall short of providing the best understanding of research problems (Creswell & Plano Clark, 2011). Mixed methods research provides more evidence for studying a problem, and encourages a broader use of multiple paradigms which can extend beyond the typical belief system aligned with one particular method. This study takes an exploratory mixed methods approach because both qualitative and quantitative data are together needed to adequately describe data use and decision rules in an English learner assessment system within a local control state. Taking a pragmatic viewpoint which seeks to find practical solutions to phenomena from multiple perspectives, this study also utilizes a semi-convergent design where data are collected and interpreted separately, then analyzed in a way which allows results to inform the other. For example, interview data can illuminate which decision rules participants have used, but an analysis of student data will reveal the collective results of decision rules which have been applied in a given year. Additionally, an analysis of student data may reveal a large number of students remaining in ELD programs in

certain grade levels which is a phenomenon that could be informed by interrogating localized reclassification procedures at the school level resulting from leeway provided within a local control state, but that may only be cataloged informally as part of district documents.

In an effort to balance the feasibility of conducting the study while also providing a relatively comprehensive description of the role played by decision rules, this study was conducted in a mid-sized district using one year of student assessment data (2012-2013). Gathering data directly from a district had several advantages. Instead of relying on publically available state-level data which does not provide classification and reclassification dates at the individual level which was a limitation mentioned in prior studies (Kim & Herman, 2010), district data were able to provide this information. In addition, district data allowed for the investigation of academic achievement and other characteristics of individual students remaining in ELD programs after reclassification decisions have been made, including students with disabilities, which were limitations observed in other studies (e.g., Kim, 2011; Hill et al., 2014b, Burke, 2014). Partnerships between researchers and school districts are known to have additional mutual benefits (see Wolf, Herman, Bachman et al., 2008) opening up the opportunity for practical application of findings and for extending the current study into an ongoing research-practice partnership. The individual student was the unit of analysis in keeping with the recommendation of a child-study approach for investigations involving the assessment of EL students (Mahoney & MacSwan, 2005, 2009). Pertinent to both questions is the rare opportunity to analyze individual-level impact of decision rules as applied to CELDT data. Since its inception in 2001, the CELDT has been used while validity evidence was still being collected, as has been the case for many English language proficiency assessments (see Rabinowitz & Sato, 2006; Wolf, Farnsworth, & Herman, 2008). Analyzing the 2012-13 year of data allows a chance

to analyze the impact of decision rules in the final year of the California Standards Test (CST) in which allows for comparisons which could prove useful for future policy and test development.

This study took a mixed method approach to data collection and analysis. Combining qualitative analyses of district procedures with quantitative analyses of procedural outcomes allowed for careful description, hypothesis generation and triangulation of findings (Creswell, 2008). Conducting the analyses with one midsized school district allowed for an in-depth investigation with a population of language minority students of a manageable enough size that most programming decisions about these students have been made by, or with the direct knowledge of, personnel in the district office. With school district personnel providing a centralized approach to EL assessment and programming, I was able to limit threats to hypothesis generation that have arisen in comparison studies due to within-district differences at large school districts, or between-district differences. The single-year design allowed for analysis of effects of certain static, grade-specific and grade-cluster specific features of the CELDT while avoiding pitfalls in longitudinal analysis which may include years where there were major changes in the CELDT (e.g., split of the K-2 test into K-1 and 2 in 2009).

Because quantitative data in this study are measurements of student performances within the state of California that are largely de-contextualized and limited by the scope of the test, there are limits to how these data can generalize to other states. Because qualitative data in this study are an intensive description of a single phenomenon as it has taken place in one California district, there are limits to generalizability to other districts and states. However, regardless of the degree of generalizability, this study's findings can provide generalizable insights to be drawn upon by other districts and states following federal mandates regarding the education of

language minority students. Educators can compare the results of this study to their own contexts to see if inferences can be drawn that illuminate similar challenges or phenomena.

Overview of Study Procedures

The study design and procedures were submitted to and approved by the Institutional Review Boards at three institutions: the University of California Los Angeles, California State University Los Angeles, and the school district. Data collection activities were organized directly between the researcher and the district's assistant superintendent.

Information Collection Plan

Research activities spanned 10 months between April 2013 and January 2014. A calendar of research activities is provided in Table 3.

Table 3

Schedule of Research Activities

Visit	Date	Activity
1	April 2013	Met with assistant superintendant to discuss possible participation of district in study
2	November 2013	Met with assistant superintendant to discuss details of IRB and schedule future visits
3	December 2013	Distributed flyers, visited leadership meeting, scheduled and conducted interviews, scheduled future interviews
N/A	December 2013	Scheduled and conducted phone interviews as needed. Revisited analytic plan and checked progress toward data saturation and diversity of interviews. By email, discussed plans with assistant superintendant to clarify adjustments to the remaining schedule
4	January 2014	Distributed flyers, scheduled and conducted interviews, completed recruitment and data collection activities

The data collection plan included multiple sources of interview data to triangulate and confirm findings. Furthermore, I used regular member checks in discussions with participants who were familiar with the EL Master Plan and reclassification processes in order to tentatively confirm initial interpretations of the data and seek understanding related to possible disconfirmation of interpretations. After the third district visit, all data collection practices and tools were re-evaluated to ensure the timeline and activities would be adequate for reaching data saturation by January. All data collected were cataloged, coded, duplicated and then stored in two separate locked cabinets to ensure a proper audit trail procedures. In addition, field notes were written in a multiple-column format to separate descriptions from interpretations and self-reflection as an interviewer and participant observer.

During the data collection period, field notes and interviews were organized and analyzed on a weekly basis to ensure analyses were happening as close as possible to the interview.

Setting

The "District" is a K-12 school district in a midsized suburban area of California. According to publically available data provided by the California Department of Education, in 2013 the district had six elementary schools (K-5), one middle school (6-8), two high schools (9-12), and one continuation high school. Total student enrollment in 2012-13 was 5,855 with approximately 1167 students, or 29 percent, designated as English Learners. The ethnicity of students in the 2012-13 enrolled population included White - not Hispanic (47.56%), Hispanic (39.29%), Asian (3.85%), African American (2.06%), Filipino (1.3 %), American Indian/Alaskan Native (1.04%), Pacific Islander (0.59%), multiple ethnicity (4.15%), and not reported (0.10%).

Data Sources

Data Sources: Qualitative. Data gathered from the district regarding data use routines included the following:

English Language Learner Master Plan. This document outlines district goals and procedures regarding assessment, classification, reclassification and educational programming for language minority students placed in English Learner programs.

Classification and Reclassification Decision Documents. These documents provide information about the criteria for classification as IFEP or LEP and reclassification as FEP, and include score reports provided by the CELDT vendor as well as those created by district personnel.

District Documents ("other"). A request will be made for documents which outline district goals and procedures regarding assessment and educational programming and individualized educational plans for students eligible for specialized educational services (e.g., gifted and talented programming; intervention programming in reading; special education services related to reading or language) in order to document the use of CELDT in determining *curricular streams* (Estrada, 2013) for language minority students at all ability levels.

Assessment Technical Manuals: CELDT and CST All relevant technical manuals for the CELDT and CST were acquired from the California Department of Education and publically available websites in order to analyze features of assessment evidence.

Semi-structured Interviews of District Personnel: Classification and Reclassification Procedures. Interviews were conducted of all district employed personnel

knowledgeable about the use of the CELDT in making decisions about EL students who had volunteered for the study. The recruitment form and interview protocol are provided in appendices B, C and D. These were updated to reflect understandings gained from initial conversations with the school district and document analysis in order to promote iterative triangulation of findings. The first part of the interview focused on typical classification procedures by grade and by year, if applicable. The second part of the interview focused on exceptional procedures by grade level or by individual-level vignette to illustrate the circumstances under which exceptions are made. The third part of the interview focused on educational programming which may use CELDT data to determine placements for language minority students in ELD as well as ELD/Special Education classes. Anonymity of students discussed in individual-level examples was assured according to data security procedures agreed upon by district personnel.

***Data Sources: Quantitative*.** The district- and student-level performance data used for analysis were collected from the following sources:

California Department of Education [CDE] DataQuest. This website provides publically available data for every school and district in the state. It is the source of district-level CELDT data, including performance levels by grade for initial and annual assessments, which was used for preliminary analyses.

Data Provided by District. The district voluntarily provided student performance data through their data management company. The following data were collected: (1) district-level data (e.g., enrollment by grade); (2) district-level reclassification rates (by grade and year); and (3) student-level data (all personal identifiers removed), including

demographic data, performance data on CELDT, CST, and the California High School Exit Exam (CAHSEE), as well as the date of reclassification.

***Measures: Quantitative*.** Measures of student performance include the following:

California English Language Proficiency Test (CELDT). This is a standardized, criterion-referenced test used to assess English language skills. The CELDT is administered to every student classified as "language minority" in order to provide levels for initial classification (either as "Initial Fluent English Proficient" or "Limited English Proficient") and then annually to every student classified as "Limited English Proficient"/"English Learner" until reclassified as "Fluent English Proficient". The CELDT is administered by grade span (K-1, 2, 3-5, 6-8, and 9-12) and assesses four modalities: listening, speaking, reading and writing. The District administers the annual test in the July 1 to October 31 assessment window and provides initial testing of new students within 30 days of enrollment. CELDT evidence used for reclassification, as recommended by the state, includes performance levels for each subtest (five levels: Beginning, Early Intermediate, Intermediate, Early Advanced, Advanced), an overall performance level (average of all four subtests, weighted equally grades 2-12), and an overall classification (Proficient or Non-Proficient, calculated by a mixed compensatory-conjunctive decision rule: Proficient = OPL of Early Advanced with no subtest level below Intermediate). CELDT evidence used for initial classification, as recommended by the state, is the overall performance level (IFEP= OPL Advanced). A table of CELDT 2012-13 raw and scale scores for level cutoffs by grade is provided in Appendix F.

California Standardized Test (CST). This is a standardized criterion-referenced test which assesses academic achievement in English language arts, mathematics and science

according to California state standards. The CST is administered each spring to all students in grades 2-11. CST evidence used for EL reclassification typically includes a certain English Language Arts (ELA) scale score (from range of 150 to 600) and performance level (from five levels: far below basic, below basic, basic, proficient, and advanced). District CST-ELA criterion is "at or above mid-Basic (≥325)."

California High School Exit Exam (CAHSEE). The CAHSEE is a requirement for high school graduation in California and is a standardized criterion-referenced test which assesses academic achievement in reading, writing, and math. Students may begin taking the test in their tenth grade year and the test must be passed before a student can receive a high school diploma (except for students with certain disabilities). Each CAHSEE section renders a scale score and a status: Passed (P), Not Passed (N), Modification Used (I) in which case no score results in Passed, Not Attempted (X or Z), Tested Before (T), and Previously Satisfied Requirement (R). The Overall Status indicates whether or not the student has passed the CAHSEE in the current or any prior administrations of the test.

Analytic Procedures: Quantitative

Student performance data were made available by the district for their entire population of students who had taken the CELDT in 2012-13. This included students who were classified IFEP, EL (LEP), and RFEP during that school year.

Procedures

Data were gathered from the California Department of Education *DataQuest* website for descriptive purposes and downloaded in PDF format. Data were gathered from the district through their data management vendor and provided in Excel spreadsheet format. Data were

cleaned, matched by unique study-specific identification number provided by the data vendor, and imported into SPSS version 22. A codebook was created for all variables. Additional variables were computed and added to the codebook as needed.

Data Storage

All data were kept in a protected filing system and organized to ensure anonymity. Specifically, data were stored in a password-protected computer and backed up daily on an external storage drive that was kept in a locked filing cabinet.

Research Questions

With data from 2012-13, these questions were asked:

(1) When applied to actual student data, to what extent do composite decision rules classify students as "proficient" or "non-proficient? To what extent do composite decision rule outcomes differ among EL, EL/SWD, EL/HP, and EL/SWD/HP students?
(2) To what extent do aggregate decision rules classify students as eligible/ineligible for program exit? To what extent do aggregate decision rule outcomes differ among EL, EL/SWD, EL/HP, and EL/SWD/HP students?
(3) As a result of the current district aggregate rule, what is the composition of the group of students who would remain in ELD programs?

Sample

This study used data from 2012-13. The students who were retained in the sample had CELDT and CST-ELA and CST-Math scores in 2012-13. Students remaining in the group were in grades 2-11 (n=667) and were split into two groups. Students with disabilities were extracted

from the main EL group to create an EL/SWD group (*n*=78), with the remainder staying in a group labeled EL (no SWD) (*n*=581). The students in the EL/SWD group had CELDT, CST-ELA and CST-Math scores in 2012, and one or more of the following: *special education eligible, primary disability code, student has an IEP, student has a 504 plan*. Two subgroups were created of high-performing students: EL/HP (*n*=107) and EL/SWD/HP (*n*=8). High-performing students had scored in the top two quintiles ("proficient" and "advanced") on the CST-ELA and CST-Math. Descriptive data were compiled for each group by grade level.

Table 4

Inclusion criteria for groupings

Group	Inclusion criteria
EL	EL student, not IFEP/RFEP, not SWD, took CELDT and CST
EL/SWD	EL student, not IFEP/RFEP, took CELDT and CST, with Special Education designation according to one of the following variables: 504 plan, has IEP, receiving special education, primary disability
EL/HP	EL student, not IFEP/RFEP, not SWD, took CELDT and CST, and scored "proficient" or "advanced" on CST-ELA and -Math
EL/SWD/HP	EL student, not IFEP/RFEP, took CELDT and CST, with Special Education designation according to one of the following variables: 504 plan, has IEP, receiving special education, primary disability; and scored "proficient" or "advanced" on CST-ELA and -Math

Methodology

Based on information from state and district documents, and data from interviews conducted with school and district personnel, six composite rules were created to reflect rules that are being used or have possibly been used in this district or in the state. All variables were created as dichotomous variables so that the mean would equal the percent proficient according to the rule.

Table 5

Composite Rules

Type	Rule	Origin
Compensatory	Overall Performance ≥ Level 4	State/district documents
Conjunctive-1	All subtests ≥ Level 4	Interviews
Conjunctive-2	All subtests ≥ Level 3	Interviews
Mixed-1	Overall Performance ≥ Level 4 *and* all subtests ≥ Level 3	State recommended
Mixed-2	Overall Performance ≥ Level 4 *and* all subtests ≥ Level 4	Interviews
Mixed-3	Overall Performance ≥ Level 4 *and* all subtests ≥ Level 3 *and* ≥3 subtests ≥ Level 4	Interviews

In addition, two aggregate rules were created to reflect the state recommended rule and the current district rule. Note that this is a truncated aggregate that only uses CELDT and CST-ELA; the full aggregate recommended by the state also uses measures of classroom performance, teacher recommendation, and parental consent. All variable were created as dichotomous so that the mean equals the percent eligible for redesignation by each rule.

Table 6

Aggregate Rules

Type	Rule	Origin
Conjunctive	CELDT "proficient" *and* CST-ELA ≥ 300 (Basic)	State recommended
Conjunctive	CELDT "proficient" *and* CST-ELA ≥ 325 (mid-Basic)	District recommended

Procedures

To answer the question *When applied to actual student data, to what extent do decision rule outcomes differ among ELL and for EL/SWD students?*, composite then aggregate rules were applied to student data. Results were reported overall and by grade. It was hypothesized that differences between rules would show that conjunctive and mixed were more stringent than compensatory rules, but the outcomes of rule differences for EL/SWD students was difficult to hypothesize due to lack of prior literature regarding this population of English learners and the variety of primary handicapping conditions present among these students. For students with learning disabilities, for example, it might be easier to be classified proficient by a compensatory rule which only requires an OPL at a certain criteria. Since the EL/SWD students in this study had not received accommodations on either the CELDT or CST assessments, however, a broader range of outcomes and differences was likely from the presence of more uncertainty in the data.

To answer the question *When applied to actual student data, to what extent do decision rule outcomes differ among high-performing students: EL/HP and EL/SWD/HP?*, composite and aggregate decision rules were applied to data. Results were reported overall and by grade. It was hypothesized that some decision rules would classify 100% of these high-performing students as "proficient" on the composite and "eligible for reclassification" on the aggregate.

To answer the question *As a result of the current district aggregate rule, what is the composition of the group of students who would remain in ELD programs?*, descriptive statistics were compiled and reported. It was hypothesized that these students could have a broad range of scores on the CELDT but likely lower scores on CST-ELA which typically co-occur with ineligibility for redesignation by definition.

Methods: Qualitative

Overview of Procedures

I collected state documents, district documents, and assessment technical manuals and conducted a cursory analysis of policy and procedures related to classification, reclassification and educational programming. When I reached saturation of my own understanding of these topics from available documents, that is, I was not finding new or disconfirming information (Brantlinger, Jimenez, Klingner, Pugach, & Richardson, 2005), I conducted the interviews and proceeded until saturation had been reached. Credibility of data and findings was established through a number of strategies, including triangulation, member checks, and rich descriptions. I transcribed and analyzed the interviews to confirm, disconfirm, complicate, and expand the findings from the document analysis. I continuously reviewed all documents to triangulate findings, asking for clarification from the California Department of Education and the district office when applicable.

Access

This research took place one of the interested districts which met the minimum requirements to be considered for this study, that is, it had a minimum of 25% EL population and included more than one school at each level of elementary, middle and high school. This district was chosen due to its interest in participating, and to the availability of key personnel (e.g., Title III coordinator) during the data collection window. I knew one employee, the assistant superintendant, and on an initial trip to the district, I approached her/him to see if s/he would be willing and interested in participating in a discussion on this topic as part of my doctoral research. Because of the current work being done in the district related to the R-FEP process, especially in relation to raising reading performances, s/he was enthusiastic about the project and

offered her/his support. S/he was as a subject of research and also my sponsor, granting access to district personnel familiar with the CELDT as used in decision making.

Participants

The participants in this study included district personnel from the district office, school site principals and vice principals, teachers, and coordinators. Because this district is not large, some participants had more than one district role (see Table 7). Schools differed in the number and percentage of EL students they served in 2012-13, with elementary school EL populations ranging from 11% to 64%. Participant recruitment sought volunteers from all corners of the district to try to ensure representation from schools with large and small EL populations.

Table 7

Participant Roles in the District

1.	Assistant Superintendant, Educational Services; Elementary Principal (K-1)
2.	District Office Data Clerk
3.	District Translator (Spanish language)
4.	Special Education Director
5.	Elementary Principal (K-5) of school with large EL population
6.	Elementary Principal (K-5) of school with small EL population
7.	Middle School Principal (6-8)
8.	Middle School Vice Principal (6-8)
9.	Middle School Principal (6-7)
10.	High School (9-12) Principal; Continuation School (10-12) Principal
11.	High School (9-12) Vice Principal
12.	Kindergarten Teacher
13.	ELD Teacher, Middle School (6-7)
14.	ELD Teacher, Middle School (6-8)
15.	ELD Teacher, High School (9-12); District Title III Coordinator
16.	Teacher of Special Education (Severely Handicapped, K-12); Preschool Teacher

Sampling Procedures: Recruitment for interviews

Due to the limited time frame of this study, it was not possible to interview all school personnel who were familiar with CELDT use. For this reason, purposive sampling was employed as the most effective way to ensure a breadth of data was gathered. Sampling was purposive insofar as it began with those who had responded to the recruitment flyer, and proceeded from there to include personnel recommended by participants. Recruitment flyers were distributed in person to the district who made them available to all district employees. Flyers were also provided at district leadership meetings, including one that I attended to speak about the study and answer questions.

When people who were interested in the study contacted me by email, I replied to their email with more information about the study. If they agreed to be in the study, we scheduled an interview. From all the people who emailed me directly (*n*=3), 100% followed through with an interview. Second, I reached out to all the people in the district who I thought may have contact with EL students or have used CELDT data for an individual student level decision. I wrote each of them a personal email inviting them to join the study (*n*=9). From all the people who responded to my email (*n*=4), 100% followed through with an interview.

Third, I was invited by the Assistant Superintendant to share information about my study at a principals' meeting. I brought flyers and business cards and made a brief announcement about the study. Several people indicated their interest. To maintain confidentiality, I did not make contacts within the meeting, but rather reached out by email after the meeting to each one in attendance. Some of these people were ones I had emailed before. From all the people who responded to this email (*n*=5), 100% followed through with an interview.

Finally, I also encouraged recommendations from interviewees to recruit participants. At the end of each interview, I asked participants if they would pass along my information (flyer, business card) to anyone else they thought might have contact with the CELDT and possibly be interested in the study. Several participants recommended people I should contact, and also said they would pass along the information. From this activity, I was put in touch with (*n=3)* more people and 100% followed through with an interview. Altogether, I reached out to 20 people. 15 responded and all 15 followed through with an interview for an 85% response rate and 100% follow through rate. During one interview, a participant invited another person to join in, making for a grand total of 16 participants.

Efforts were made to ensure that among the participants in the study, there were those connected to elementary, middle and high school levels, and that specific district positions directly related to English learners and Special Education were recruited. Once participants from all levels of the district, and from both administrative and teaching positions within those levels, were recruited, interviews proceeded and additional interested participants were added. This allowed for a saturation of data on the questions of interest in the study, namely the use of assessment data for classification, reclassification and educational programming. During the data collection period, I continued to accept new participants into the study. It was during the final round of interviews that I began to notice how the same information was being given by participants in similar roles (e.g., middle school principals, EL teachers). This indicated to me that saturation had been reached insofar as "recent interviews discern(ed) the same information given by earlier respondents" (p. 198, Brantlinger et al., 2005). I completed the interviews that had already been scheduled and stopped recruitment. Data collection was then concluded.

Semi-Structured Interviews

Interviews were scheduled to be held in a central location, a private office at the district office with its own private entrance, to ensure confidentiality. All participants were given this option. Some participants preferred to meet in their own office at their school site. Some participants could only meet on days I was not in the district and thus agreed to a phone interview. Altogether, 16 participants were interviewed. 18 interviews were conducted which included one participant being interviewed 3 times, one participant being interviewed twice, and two participants being interviewed at the same time. Interviews ranged in length from 18 to 62 minutes. The average time spent in the interview was 35 minutes.

Interview Procedures

Interviews were conducted of selected district-level personnel who volunteered for the study and were knowledgeable about the use of the CELDT in decision making for English learners. The first part of the interview focused on typical classification procedures, reclassification procedures, and educational programming for language minority students who had taken the CELDT. The second part of the interview focused on exceptional procedures by grade level or by individual-level vignette to illustrate the circumstances under which exceptions are made. Anonymity of students discussed in individual-level examples was assured according to data security procedures agreed upon by district personnel.

Data Storage

All data were kept in a protected filing system and organized to ensure anonymity (Creswell, 2008). Specifically, data were stored in a password-protected computer and backed up daily on an external storage drive that was kept in a locked cabinet. Names of participants were masked and coded for anonymity. Codes were separated from names and kept secure in a separate file.

Credibility of Findings

In my role as researcher and interviewer, it was important for me to acknowledge and clarify my own intrinsic biases, such as my personal background and motives for conducting the study (Yin, 2011). From my education and experience working in schools, I had my own biases and beliefs about the effective use of assessment data in serving students in the language minority population of students, as well as students with disabilities. However, I did not express personal beliefs or judgments during recruitment activities and interviews so that I could remain open to opinions different from my own, and information that could disconfirm my prior assumptions. I was careful to actively seek out alternate explanations for phenomena and disconfirming evidence so that the descriptions I would make from these data would not simply validate my own preconceived opinions or biases.

Analytic Procedures: Qualitative

Following the suggestions of Corbin and Strauss (2007) and Creswell (2008) in developing a grounded theory and using a constant-comparative method, analysis proceeded iteratively. The process is described as follows (see Appendix H for codebook):

Organization and preparation of data for analysis: Raw audio data were transcribed and assigned a two-digit code number not attached to the participant information. After transcriptions were completed, they were checked thoroughly for accuracy while listening to the audio data two more times.

Reviewing all data: The final transcripts were read closely twice, clustered by participant role. The first reading provided a general sense of the data and gave the opportunity to reflect on the overall meaning (Creswell, 2008). During this time, transcripts were annotated with my general thoughts and impressions of the data at this stage. Theses coder's notes were

procedural, e.g., "check to see if other EL teachers have discussed this"; analytical, e.g., "note different CELDT composite rule"; and theoretical, e.g., "note the different uses of 'we' and 'them' in reference to teachers, district leaders, and students." These notes functioned as a way to keep track of initial impressions and served as an advanced organizer for potential codes. After this, coding proceeded in three stages.

A Priori Coding*:* Applying a grounded theory (Merriam, 2002) the coding process began with a deductive approach of coding based on a priori codes related to the theoretical framework (Coburn & Turner, 2011), the research questions, and the interview questions. Data were coded to identify the data use routines of *initial classification*, *reclassification*, and *educational programming*. Other *a priori* codes included *special education, long-term EL*, and *tools*.

Axial Coding: A second round of coding was also primarily deductive and included additional, axial codes to represent different aspects of the Coburn & Turner (2011) and Piety (2011) data use frameworks. These included codes such as *timing*, *access*, *motivation*, *data use*, and *inference.* After the axial coding was complete, themes and descriptions began to be developed in the form of a narrative.

Selective Coding: A third round of coding took an inductive approach, seeking to add selective codes which emerged from the axial coding and the subsequent building of themes and descriptions. These codes were meant to more fully describe phenomena and contributors to the system which had emerged as themes in the data. These included codes such as *stuck at 3*, and *parents*. When this stage of coding was completed, I built the descriptive narrative of all themes.

CHAPTER FOUR

QUANTITATIVE DATA PRESENTATION AND ANALYSIS

The aim of this study is to describe the potential role of decision rules, composite and aggregate, in the LTEL phenomenon. To this end, quantitative data were gathered for analysis to shed light on the extent of difference between decision rules for EL students in two groups: EL, EL/SWD; and two sub-groups: EL/HP, EL/SWD/HP. In this chapter, I will provide an analysis of actual student performance data to which six composite and two aggregate decision models were applied. Then, I will describe the results of the district's actual aggregate decision outcomes in terms of who was reclassified and who remained in EL programs.

Descriptive Data

Data were compiled to describe the test performances of students by group, provided in Table 9. Students in all groups showed the full range of performance, from level 1 through 5, in both CELDT and CST-ELA and CST-Math. For students with disabilities, descriptive data regarding primary disability is provided in Table 8. Most students receiving services for specific language impairment were in elementary (grades 2-5) while students receiving services for learning disability were in all grades (2-11) as is typical in special education.

Table 8

Descriptive Data: Students with Disabilities

		Primary Disability						
	Has IEP	220	240	260	270	280	290	999 & 0
		HH	SLI	ED	OI	OHI	SLD	N/A
EL/SWD (*n*=78)	36% (29)	1% (1)	27% (21)	1% (1)	1% (1)	3% (2)	35% (27)	31% (24)
EL/HP/SWD (*n*=8)	25% (2)	0	64% (5)	0	0	0	12% (1)	25% (2)

Note: No students were on 504 plans. HH=Hard of Hearing; SLI=Speech or Language Impairment; ED=Emotional Disturbance; OI=Orthopedic Impairment; OHI=Other Health Impairment; SLD=Specific Learning Disability

Table 9

Descriptive Data by Group: Mean Levels and Scale Scores, CELDT and CST

	CELDT				
	Listening	Speaking	Reading	Writing	Overall
EL	3.47	3.76	2.73	3.26	3.28
EL/SWD	3.17	3.69	2.50	3.04	2.99
EL/HP	4.32	4.07	3.34	3.64	3.93
EL/HP/SWD	4.00	3.50	3.13	3.75	3.63

	CST			
	ELA		Math	
	Level	Scale	Level	Scale
EL	2.53	306	2.74	314
EL/SWD	2.09	286	2.29	289
EL/HP	4.32	389	4.43	416
EL/HP/SWD	4.25	370	4.13	381

Note: CELDT and CST have 5 levels. EL (*n*=474), EL/SWD (*n*=78), EL/SWD (*n*=107), EL/HP/SWD (*n*=8)

Composite Rules: EL

Six rules were applied to CELDT results for the EL group (*n*=474). The results are provided in Table 10. The state rule, Mix-1, resulted in 58% non-proficient and 42% proficient. The other rules produced a range of 42-86% non-proficient and 14-58% proficient. This difference means as few as 199 (42%) and as many as 407 (86%) students would continue receiving EL services.

Table 10

Percentage Proficient by Composite Rules: EL

	COMP	CONJ-1	CONJ-2	**MIX-1**	MIX-2	MIX-3
Non-Proficient	57	86	42	**58**	86	72
Proficient	43	14	58	**42**	14	28

n=474

Composite Rules: EL/SWD

Six rules were applied to CELDT results for the EL/SWD group (*n*=78). The results are provided in Table 11. The state rule, Mix-1, resulted in 72% non-proficient and 28% proficient.

The other rules, from highest to lowest, resulted in 55-88% non-proficient and 12-45% proficient. This difference means as few as 43 (55%) and as many as 69 (88%) students would continue receiving EL services.

Table 11

Percentage Proficient by Composite Rules: EL/SWD

	COMP	CONJ-1	CONJ-2	**MIX-1**	MIX-2	MIX-3
Non-Proficient	68	88	55	**72**	88	74
Proficient	32	12	45	**28**	12	26

n=78

Composite Rules - Subgroup: EL/HP

Six rules were applied to CELDT results for the EL/HP subgroup (*n*=107) who were extracted from the EL (no SWD) group. The results are provided in Table 10. The state rule, Mix-1, resulted in 67% non-proficient and 33% proficient. The other rules, from highest to lowest, resulted in 18-72% non-proficient and 28-72% proficient. This difference means that as few as 19 (18%) and as many as 77 (72%) high-performing students would remain in ELD.

Table 10

Percentage Proficient by Composite Rules: EL/HP

	COMP	CONJ-1	CONJ-2	**MIX-1**	MIX-2	MIX-3
Non-Proficient	29	72	18	**33**	72	46
Proficient	71	28	82	**67**	28	54

n=107

Composite Rules - Subgroup: EL/SWD/HP

Six rules were applied to CELDT results for the EL/HP group (*n*=8) who were extracted from the EL/SWD group. The results are provided in Table 13 but listed as frequencies instead of

percentages due to the small size of the group. The state rule, Mix-1, resulted in 3 non-proficient and 5 proficient. The other rules resulted in between 2 and 8 students non-proficient and from none to 6 students proficient. This difference means as few as 2 and as many as 8 high-performing students with disabilities would continue receiving EL services.

Table 13

Number Proficient by Composite Rules: EL/SWD/HP

	COMP	CONJ-1	CONJ-2	**MIX-1**	MIX-2	MIX-3
Non-Proficient	2	8	2	**3**	8	6
Proficient	6	0	6	**5**	0	2

n=8

Aggregate Rules: *CST-ELA Criteria*

There are two CST criteria for reclassification that have been used in this district: The state rule of "CST-ELA at Basic or above (≥300)," and the district's recent rule of "CST-ELA mid-Basic or above (≥325)." The results for all groups are provided in Table 14. In the EL group, 57% (*n*=272) of students met the state criterion and 33% (*n*=156) met the district criterion. In the EL/SWD group, 36% (*n*=28) met the state criterion and 23% (*n*=18) met the district criterion. All students in the high-performing group met both criteria by definition.

Table 14

Percentage of Students Meeting CST-ELA Criteria

	State ≥300 (Basic)	District ≥325 (mid-Basic)
EL	**57**	33
EL/SWD	**36**	23
EL/HP	**100**	100
EL/SWD/HP	**100**	100

Aggregate Rules: EL

Two aggregate criteria were applied to the results of the six composite rules for the EL group (*n*=474). The results are provided in Table 15. The state aggregate, Mix-1 plus CST-ELA ≥ 300, resulted in 71% remaining classified as LEP and 29% eligible for RFEP. The other rules resulted in 62-94% remaining classified as LEP and 6-38% eligible for RFEP. This difference means as few as 292 (62%) and as many as 447 (94%) students would remain in EL services. The actual number of EL students remaining classified as LEP was 401 (85%) while 73 (15%) were reclassified.

Table 15

Percentage of Students Eligible for RFEP: EL

	COMP		CONJ-1		CONJ-2		**MIX-1**		MIX-2		MIX-3		ACTUAL
	300	325	300	325	300	325	**300**	325	300	325	300	325	
Ineligible	70	80	91	94	62	76	**71**	81	91	94	81	87	85
Eligible	30	20	9	6	38	24	**29**	19	9	6	19	13	15

n=474

Aggregate Rules: EL/SWD

Two aggregate criteria were applied to the results of the six composite rules for the EL/SWD group (*n*=78). The results are provided in Table 16. The state aggregate would result in 86% remaining classified as LEP and 14% eligible for RFEP. The other rules would result in 76-95% remaining classified as LEP and 5-24% eligible for RFEP. This difference means as few as 59 (76%) and as many as 74 (95%) students would remain in EL services. The actual number of students who remained classified as LEP was 64 (82%) while 14 (18%) were reclassified.

Table 16

Percentage of Students Eligible for Reclassification: EL/SWD

	COMP		CONJ-1		CONJ-2		**MIX-1**		MIX-2		MIX-3		ACTUAL
	300	325	300	325	300	325	**300**	325	300	325	300	325	
Ineligible	85	90	92	95	76	86	**86**	91	92	95	87	91	82
Eligible	15	10	8	5	24	14	**14**	9	8	5	13	9	18

n=78

Aggregate Rules - Subgroup: EL/HP

Two aggregate criteria were applied to the results of the six composite rules for the subgroup of EL/HP (*n*=107). By definition, 100% of these students are eligible for reclassification according to CST-ELA criteria. Therefore, if less than 100% are found eligible for reclassification, we know that the CELDT composite rule has controlled the outcome. The results are provided in Table 17, and are by definition the same as the results from the composite rules due to the fact that these students have already met both CST-ELA criteria. The state aggregate would result in 33% remaining classified as LEP and 67% eligible for RFEP. The other rules would result in 18-72% remaining classified as LEP and 28-82% eligible for RFEP. This difference means as few as 19 (18%) and as many as 77 (71%) of these high-performing students could be staying in ELD. The actual number of students who remained classified as LEP was 69 (65%) while 38 (35%) were reclassified. Notably, all 70 students remaining in ELD were in grade levels 2-5.

To foreshadow the findings presented in the next chapter, according to district documents and interviews there are differences in the district's reclassification rules and procedures which could explain these results. Participants reported between-site and within-site (between-grade) differences in decision rules. For example, some elementary schools in the district do not start reclassifying until 3rd grade, and even 4th grade, even if students have met the CST-ELA criteria.

Coupled with the fact that most of the ELs in this district are in elementary, it is not surprising to discover that the majority of students remaining classified as LEP at the end of 2012-13 were in elementary. From the whole sample (*n*=667), there were 450 students in grades 2-5. In those grades, only 71 students (16%) were reclassified, and no students in 2nd grade were reclassified, even if they had met the CST-ELA criteria. In contrast, of the 124 students in 6th through 8th grade, 41 students (33%) were reclassified with the largest percentage at 7th grade (43%). This suggests the enactment of site-level decision rules, not necessarily a lack of adherence to state or district rules; this is an important distinction related to policies affecting high-performing ELs and other sub-groups.

Table 17

Percentage of Students Eligible for Reclassification: EL/HP

	COMP		CONJ-1		CONJ-2		**MIX-1**		MIX-2		MIX-3		ACTUAL
	300	325	300	325	300	325	**300**	325	300	325	300	325	
Ineligible	29		72		18		**33**		72		46		65
Eligible	71		28		82		**67**		28		54		35

n=107

Aggregate Rules - Subgroup: EL/SWD/HP

Two aggregate criteria were applied to the results of the six composite rules for the subgroup of EL/HP (*n*=8) who all are eligible for reclassification according to the state and the district CST-ELA criteria. The results are provided in Table 18. The state aggregate resulted in 3 remaining classified as LEP and 5 eligible for RFEP. The other rules resulted in between 2 and 8 remaining classified as LEP and between zero and 6 eligible for RFEP. The actual number of students who

remained classified as LEP was 3 while 5 were reclassified by district rules. Notably, all 3 students remaining classified as LEP were in grade levels 2-4.

Table 18

Number of Students Eligible for Reclassification: EL/SWD/HP

	COMP		CONJ-1		CONJ-2		**MIX-1**		MIX-2		MIX-3		ACTUAL
	300	325	300	325	300	325	**300**	325	300	325	300	325	
Ineligible	2		8		2		**3**		8		6		5
Eligible	6		0		6		**5**		0		2		3

n=8

Differences between rules: McNemar's test

To investigate the significance of differences between rules, nonparametric analyses were conducted using McNemar's test. Tables are provided in Appendix G. First, all composites were compared with the state composite (Mix-1) by group. No tests were conducted for the EL/SWD/HP group (*n*=8) due to insufficient group size. For the EL, EL/SWD and EL/HP groups, all compensatory and mixed composites were significantly different from the state ($p \leq .001$), but the conjunctive model was not significantly different for two groups, EL/SWD (*n*=78, $p \leq .250$) and EL/HP (*n*=107, $p \leq .125$). This suggests that the compensatory criterion (overall ≥ Level 4), which is in common between the conjunctive rule and the state rule, is a stronger predictor of categorical outcome than the less stringent conjunctive criterion (all subtests ≥ Level 3). This may suggest the rarity of achieving the conjunctive criterion without also meeting the compensatory criterion. These analyses were conducted again to explore differences by gender. For the female group (*n*=335), all of the composites were significantly different from the state ($p \leq .004$ to $p \leq .001$). For the male group (*n*=332), the conjunctive composite was not significantly different from the state ($p \leq .063$) but all the other composites were ($p \leq .001$).

Next, all aggregates were compared with the aggregate from the state (Mix-1 & CST-ELA ≥ 300) and district (Mix-1 & CST-ELA ≥ 325). For the EL group, all aggregates were significantly different from the state ($p \leq .031$ to $p \leq .001$). For the EL/SWD group, only the least stringent conjunctive-2 aggregate was significantly different from the state ($p \leq .008$) but not the district ($p \leq .125$). This may suggest that for this population, the more stringent CELDT composites provide similar information as the CST-ELA at the state criterion; additionally, the more stringent CST-ELA district criterion may render the CELDT information redundant. More analyses would be necessary to quantify this relationship, which is an important topic for future research in light of current conversations about a national reclassification criterion and whether it should include evidence from ELA tests.

Decision Rules – Students remaining in ELD programs in 2013-14

To describe which students will remain classified "LEP" and in EL services, descriptive statistics were compiled. Of the 667 students in this study, 539 (81%) were not reclassified in 2012-13. This group was comprised of 263 male (49%) and 243 female students (51%). The grade level distribution showed how the size of the cohort diminished as the grades advanced. The largest portion of students remaining in EL were in elementary (percentages are of the whole retained group): Grade 2 (27%, *n*=145), Grade 3 (18%, *n*=99), Grade 4 (14%, *n*=75), Grade 5 (11%, *n*=60). A diminishing population was seen in middle and high school: Grade 6 (7%, *n*=37), Grade 7 (5%, *n*=25), Grade 8 (4%, *n*=21), Grade 9 (6%, *n*=33), Grade 10 (5%, *n*=28), Grade 11 (3%, *n*=16).

There were some students listed as migrants (4%, *n*=21). None of the students remaining in EL programs had been designated as gifted, although this was reported by participants to be due to the lack of gifted programs and thus the lack of testing to determine designations. The

most frequently recorded primary language was Spanish (87%, *n*=473), with a broad list of other languages including: other non-English language (2%, *n*=11), Arabic (2%, *n*=10), Filipino (2%, *n*=9), Vietnamese (1%, *n*=5) and other languages with fewer than 5 speakers in each, listed here in alphabetical order: Burmese, French, German, Hindi, Ilocano, Indonesian, Khmer(Cambodian), Korean, Mandarin, Portuguese, Punjabi, Russian, and Thai.

On the CST-ELA, the average achievement was at the "Basic" level with a mean scale score of 311 and a full range of performances from 202 (Far Below Basic) to 527 (Advanced). On the CST-Math, the average achievement was at the "Basic" level with a mean scale score of 321 and a full range of performances from 177 (Far Below Basic) to 491 (Proficient). On the CELDT, students' performances spanned the full range of Level 1 (Beginner) to Level 5 (Advanced) overall and on every subtest. The mean performance level overall was 3.13. The mean performance level for the subtests suggest that the students' greatest area of need is in reading (2.55), followed by writing (3.06), listening (3.41) and speaking (3.67). Notably, even for the students who were reclassified, the lowest mean performance level was for the reading subtest (3.91).

Some students remaining in EL had already passed the CAHSEE-ELA (*n*=10) and the CAHSEE-Math (*n*=14). Further analysis of test performances indicated that of the 10 students who passed the CAHSEE-ELA, none had high performances on the CST-ELA: Far Below Basic (*n*=1), Below Basic (*n*=6), and Basic (*n*=3). In contrast, their overall performances on the CELDT were high: Level 5 (*n*=3), Level 4 (*n*=4), and Level 3 (*n*=3). This finding suggests that students who fail to meet the CST-ELA criterion for reclassification may already have passed the CAHSEE, indicating a possible misalignment of proficiency evidence for educational programming, EL program exit, and high school exit decisions.

Overall, these finding suggest that students remaining in ELD programs in this district in 2013-14 will have a broad range of educational needs. From absolute beginners to students who have already passed the CAHSEE, and in small numbers in middle and high school grades, the educational needs of this group are likely to vary considerably. This range and distribution of students could be perpetuating the difficulty of providing adequate and effective language services to all students, a notion that is explored in the next chapter and in the discussion.

CHAPTER FIVE

QUALITATIVE DATA PRESENTATION AND ANALYSIS

In this chapter I will use evidence from state and local documents, and interviews with school district personnel to describe the context and practices of data use for initial classification, reclassification, and educational programming. The description will be organized according to the framework set forth by Coburn & Turner (2011). I will begin by describing some of the influences on data use in terms of the ***organizational and political context*** and the ***infrastructure and tools context.*** Then I will describe data use practices of *noticing, interpretation*, and *constructing implications* for two key groups of students: EL students with disabilities (EL/SWD) and students who spend extended time at ELD Level 3.

Organizational and Political Context

Coburn and Turner (2011) suggest six dimensions of influence within the ***organizational and political context****(s)*: *data use routines, access to data, leadership, time, norms,* and *relations of power and authority*. *Data use routines*, defined as the recurrent, patterned interactions which guide the use of data for a particular purpose (Coburn & Turner, 2011), were an organizing principal for data collection and analysis. *Data use routines* of initial classification, reclassification, and educational programming were central to this study. Interviews and documents yielded information on the other five dimensions. The description and analysis that follows explore dimensions which exert influence on the *data use routines*.

Access: Timing of the CELDT

All the data use routines were influenced by the timing of CELDT administration, the timing of results, and the timing of state accountability reports (see summary in Table 19).

Table 19

The Seasons of Annual Testing and Data Use for EL Status Decisions

Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul
Annual CELDT testing			Waiting for official CELDT results		RFEP decisions			CST Testing		Waiting for CST results	
Data Use for ELD: CELDT prior (12+ mo. old)			Data Use for ELD: CELDT prior (15+ mo. old); OR CELDT current, unofficial (3+ mo. old)		Data Use for ELD: CELDT current, official (5+ mo. old)					[school out of session]	
Data use for RFEP decisions: CELDT prior (12+ mo. old) CST current (3+ mo. old)					Data use for RFEP decisions: CELDT current (6+ mo. old) CST prior (10+ mo. old)						

Timing of CELDT Testing: The timing of CELDT testing is determined by state policy. For initial classification and placement in ELD, students must be tested within 30 days of enrollment and this occurs between July 1 and October 31. The rest of the "initials" happen at the same time as annual testing.

The annual testing in the district happens from approximately August 20 to October 31. District leadership decides how to allocate the limited resources of internal and externally-hired staff to complete the testing activities. The annual CELDT testing happens from approximately August 20 to October 31. Even under the best of circumstances, participants shared that the testing is time-consuming, labor-intensive, and can also be disruptive to the provision of ELD: *"Our EL specialists pull all the kids and give the CELDT. It takes way too long to get those kids classified to get ELD started. …We wanted it done and over with because we wanted those support people back in our classrooms for the support."* Participants also reported that students new to the district, and new to English, are affected the most by limited access to EL specialists in their classes during the testing period: *"I have kids that have been here since the end of August, it's now the end of October, and they're drowning. Six weeks has passed and they now*

have gone into a pattern of non-production, or despair, or like 'Where the hell am I?' The newcomers and 1s and 2s would get a little triage but the 3s are on their own."

Timing of CELDT results: In accordance with state policy and procedures, CELDT raw scores are recorded at the district, converted to preliminary scale scores, then sent to CELDT vendor and data management company (henceforth, CELDT vendor). The state is the entity which negotiates with the CELDT vendor to establish timelines for completing the work needed to produce the "official" CELDT results and which are sent to districts.

In order to determine what ELD services (e.g., which ELD level class) students need, district personnel rely on raw or preliminary scale scores. These are used to make preliminary determinations of services and programs. If preliminary scores indicate a student is ready for a new classification, either IFEP or RFEP, no status change or service change would happen until official scores are delivered. District and school site leadership can influence the extent to which these scores are used through local data use procedures until the "official" scores arrive. Teachers expressed some frustration about having to wait for "official" scores from the state, yet some also reported that CELDT scores were not of much use when they arrived: *"The CELDT scores results are given so late in the year. They would be much less ineffective if we had them, like if the kids could test at the end of the school year, or, like right even before the school started so that the scores were available. Because we're just getting scores, you know, official scores don't even come out until the beginning of the year, January or whatever, so by then you've already had the kids for a whole semester."* This seems to indicate that students may stay in the same ELD level class as the year prior until "official" scores are delivered. This delay in matching students to services can influence how students are able to progress in ELD, leaving teachers with potentially less influence over student outcomes. Alternatively, the frustration

could be due to an unsettled feeling of knowing test results have yet to be delivered.

Because district leadership allocate additional resources to hire enough extra staff to complete the testing according to state policy guidelines, the delay of the results was frustrating: *"We hire [people] to CELDT our school because we have 30 days to do the initials and we'd better get it done, you know? I mean there's that piece that nobody really owns, 'So, when do I get my results?' 'I don't know, January, maybe, I don't know.'"* This frustration could be related to the fact that CELDT results are used for different purposes and have different seasons of urgency at different levels of this system. The following example illustrates this difference.

Data Use: Different Seasons of Urgency

There are two main purposes for data use in the EL system: accountability use, and student use for educational programming. For both uses, data needs to be entered into database systems in a timely fashion. The timing of the data entry, however, seems more closely related to accountability purposes than students' educational programming. In the new California Longitudinal Pupil Assessment Data System (CALPADS), the district is no longer able to leave a student's EL status blank until "official" scores arrive. So, for students who had taken the initial CELDT and had no prior EL status on file, the district created a "W" status meaning "initial wait to classify" as a placeholder. These students are considered EL and given ELD services until the official scores can verify those who are IFEP: *"We had to wait for the state report to come back before we can change them to an "I" [IFEP] because of how CALPADS calls the student information. We have to wait until we get the final data from CELDT to say officially, okay, this kid's an IFEP. We had originally been (able to say) they're an IFEP, or they're an EL and we'd put them in a program. But now, we can't put them in a program until we know officially. You get the raw scores but we're saying, alright, we're going to wait because*

we don't have the official word from the state. When that data gets pulled for CALPADS, which is now of course the vehicle that's measuring the demographics, we get errors if we don't have accurate data. So we put in Ws on all the kids until we knew for sure, because we didn't want to inaccurately represent." This suggests the urgency of accountability reporting must be satisfied, but the timely availability of "official" scores for students' educational programming can wait.

Participants concerned about the preliminary scores being "inaccurate" may be related to how the state explains the delay in reporting official results: *"The state said,* 'O*h, there was an error in some sort of calculation with CELDT and you're not going to get the scores 'til later.' And so, then, what do we do?"* The wait time for results to determine initial classifications, as well as annual reclassifications, can delay the accurate educational programming of individual students by up to 6 months. Around January, the state sends the official results accompanied by an errata sheet with revised cut-points for each grade. These official scores are also used by the state for federal accountability reports and to complete the technical manuals which are made available at end of the school year. While at the district level there is some urgency for timely results which could lend to more accurate inferences for educational programming, the power to change the timeline rests with the state or another entity. District personnel defer to the state and accept the delay: *"I mean, we have the raw data but we still always wait for validation from the state because, who wants to undo it if it's wrong?... We tend to not enter it into our system until we know, which delays the grouping and services. But of course, if you're assessing your kids from August 20th until October 31st, how do you start the year grouping and supporting them?"* When official scores are received, district personnel are then tasked to complete all updates to EL status and all reclassifications in time for the "Annual R-30 Student Nation of Origin Report" due March 1st. The timing of results in January allows this process to be conducted smoothly in

time for state and federal accountability. Even though scores may no longer match students' current language abilities, and thus may prove to be a weaker inference for decision making use, still the "official" CELDT results supersede all other measures in terms of deciding which students continue in ELD. Similarly, for students whose status changes to IFEP or RFEP, educational programming changes are only official with "official" scores.

In summary, decision-makers prefer to have the most recent CELDT data in time for the main data use activity: putting students in programs and, if in ELD, creating instructional groupings. Yet, the timing of results seems to favor purposes at the state level which seem to operate within different seasons of urgency. The timing of results is due to state policy and leadership, but could also be influenced by leadership related to the infrastructure and tools, namely the CELDT vendor and CALPADS.

Relations of Power and Authority[7]: Determination of Decision Rules

The State Board of Education (SBE) provides suggestions for districts to follow in terms of the composite CELDT proficiency determination and the aggregate reclassification decision (*EC* Section 313). As a local control state, districts can establish their own reclassification policies and procedures. The routine of reclassification according to state guidelines starts with CELDT and CST-ELA, followed by teacher evaluation, other measures if used, and parent consultation. Local districts have control over the criterion points on the CELDT and CST which can lead to differences in outcomes (Hill et al., 2014). Districts are expected to choose criteria and use an aggregation process that is most effective for their own purposes. The guidelines of the state and by the district in this study are presented in Table 20.

[7] The leadership of test vendors and data management companies could also influence the determination decision rules by the data support they provide, or even the cost and availability of certain grain sizes of data to create decision rules. Although not a focus of the current study, the power/authority of these entities is worth examining, perhaps using a sociotechnical framework (Piety, 2011).

Table 20

Reclassification Criteria: State and District

	State guidelines	District guidelines		
	2011-2013	2011-12	2012-13	2013-14
Performance in basic skills	CST-ELA ≥300 (Basic)	CST-ELA ≥300 (Basic)	CST-ELA ≥325 (mid-Basic)	(no CST available)
Assessment in English proficiency: CELDT	Overall ≥ Level 4 Subtests ≥Level 3	Overall ≥ Level 4 Subtests ≥Level 3	Overall ≥ Level 4 Subtests ≥Level 4	Overall ≥ Level 4 Subtests ≥Level 4
Teacher evaluation of student academic performance	Teacher certification that the student meets the district's academic performance indicators	Teacher certification that the student meets the district's academic performance indicators	Teacher certification that the student meets the district's academic performance indicators	Teacher certification that the student meets the district's academic performance indicators
Parent or guardian opinion and consultation	District provides notice to parents/guardians of their right to participate in the reclassification process and encourages them to attend a face-to-face meeting	Parents contacted regarding student's eligibility to reclassify with a face-to-face meeting option; letter sent by mail which must be returned with a signature	Parents contacted regarding student's eligibility to reclassify with a face-to-face meeting option; letter sent by mail which must be returned with a signature	Parents contacted regarding student's eligibility to reclassify with a face-to-face meeting option; letter sent by mail which must be returned with a signature
Additional criteria chosen by local educational agency (optional)		Elementary – varies; Secondary – minimum C average/GPA; achievement in reading intervention program Special Ed - varies	Elementary – varies; Secondary – achievement in reading intervention program; no GPA/grade requirement Special Ed- varies	Elementary – varies; Secondary – achievement in reading intervention program; no GPA/grade requirement Special Ed- varies

Note: "District" refers to the district in this study; table format adapted from Hill, Betts, Chavez, Zau, & Bachofer (2014), page 7, Table 2.

The district has used the authority vested by the state to determine its own composite and aggregate decision rules. This is done for local reasons and in response to the changeover of assessment systems in the state which affects test availability. Not only at the district level, but also at the school site level there can be differences in how composite and aggregate decision rules are determined and applied. During the interviews, participants mentioned more than one composite rule for the CELDT. Some of the rules mentioned were only mentioned once which could indicate lapses in recall or the sharing of misinformation. However, all rules are plausible and, for the purposes of this study, useful for describing the extent to which student outcomes can differ under alternate rules. The composite rules included the following, in order of stringency: (1) OPL 4 or 5, all subtests at 4 or 5; (2) OPL 4 or 5, one subtest at 3 or above, all other subtests at 4 or above; (3) OPL 4 or 5, reading at 3 or above, all other subtests at 4 or above; and (4) OPL 3 or higher, "no gaping hole" in subtests.

In addition to locally determined decision rules, participants at the elementary level noted that reclassification didn't often occur in the earlier grades. This was confirmed by site principals whose data use routines for reclassification did not start reclassifications until 3rd grade at one site, until 4th grade at another. At the secondary level, reclassification is available at all grades but seems to be emphasized at transition points of 8th and 10th grade. Thus, within the district authority, school-site authority can also influence reclassification in terms of rule choice and grade level. Differences in how site leadership in elementary, middle, and high school considered decision rules could be attributed to many factors, including the difficulty in keeping up with changes in state and local decision rules over time, different levels of knowledge about rules, different levels of experience enacting reclassification, different levels of adherence to district's suggested decision rules, or other

factors. For example, some participants described and explained rules by talking about the impact on school site scheduling and resources for hiring. For secondary schools, the number of students remaining in ELD classes affects district resources for keeping EL specialists and, in some cases, ELD teachers on staff. There was no direct link made between the decision rules used and the number of students needed to keep the program running; rather, participants talked about reclassification being the goal. However, the lack of resources was a consistent theme. Site and district leaders spoke of managing ELD with the resources and staffing they had, despite the funding cuts, the size of their ELD programs, and the number of students in each ELD level.

According to several administrators, schools with fewer EL students struggle to create ELD classes that are appropriately leveled in a master schedule with additional constraints due to blocked classes. One participant shared: *"It's really hard because we're blocked. We do try to level them, just even as a general policy, we try to go in our classes, medium to high, and medium to low. Just because of, I mean, a teacher with low to high in one class is just, it's nuts. And so the highs lose out, everybody loses out. And so we really try to do that and to level it as much as possible. Sometimes we're successful, sometimes we're not."*

For placement of students in the rest of their instructional day, there are also challenges due to block scheduling when ELD is the elective class: *"When we're building class lists, we look at CELDT scores and try to group kids with similar CELDT scores into the same classes, so that we can, so the range isn't quite so big. There are other constraints on the master schedule that don't always make that possible. You know, we get a new student who's just here from Mexico, CELDT level 1, doesn't know any English, and we can't put her where the other ones are because the teachers will go into overages. Something like that can happen. So then we just*

put her where we can put her. So we do use them to place kids and we do try to group them according to their CELDT score but, sometimes that doesn't work."

What participants have recounted serves to illustrate the constraints placed upon data use routines, including staffing resources, the master schedule, and the number of students at each of the five ELD levels. When under constraints, those in leadership use their authority to create conditions for optimal outcomes and these changes can be studied through the examination of artifacts (Piety, 2011). In this case, one of the artifacts is the creation of newly defined ELD classes, such as an ELD 2/3 combination class, or the splitting of ELD 3 into "low 3" and "high 3" classes. These artifacts of the constraint suggest that decision makers are noticing students' actual needs and the limitations of the resources available, and using whatever authority they have to produce the desired outcome.

Relations of Authority and Power

According to Coburn and Turner (2011), relations of power and authority "play a role in how data use processes unfold" and are "a near omnipresent characteristic of the context of public schools" (p. 184). When examining data use routines in the context of a local control state, the dynamics between federal, state, district, and school-site authority varied by routine.

Initial Classification

In contrast to the reclassification routine, districts are not given the authority to define initial classification decision rules. Initial classification is determined based on the results of the Home Language Survey with, if needed, the results of the initial assessment on the CELDT to which the state's composite rule is applied. There are no other assessments, classroom performance data, teacher consultation or parent consultation which are used: there is no

aggregate classification, just the composite rule. However, according to the district's EL Master Plan, parents are allowed to request a different educational setting for their student: "at any time during the school year, you may have your child moved into the English Language Mainstream Program" (p. 9, District EL Master Plan). This option was verified by interviews where participants confirmed that although parents had this right, it was seldom exercised. This illustrates how authority over educational programming decisions can be shared among parents and districts.

There were several examples presented during the interviews which illustrate how data use routines and inferences drawn from data are influenced by power and authority relations, but also how power and authority relations can be influenced by data use routines.

Parent Involvement in the Reclassification Routine

According to the state, parents are given the chance to provide their opinion and be consulted in the process of reclassification. Districts are encouraged to notify parents of their rights and invite them to participate in the reclassification process. In this district, parents' are invited to participate in the reclassification process either in person or by phone. Sometimes there is an individual student study team (SST) or a group SST, but often parents simply receive a phone call with a request for a signature of approval on the RFEP form. One participant involved with this process shared a typical notification call: *'The reason I'm calling is your child was recommended to be reclassified. He meets the requirements. The teacher has recommended it. They approved it. I am going to send you a form, if you agree, and it says where to sign and to return, and I'm sending you a copy for yourself, so you keep that, sign it, return and send the other one. Congratulate your kid and, he will be monitored for 3 years. So if you see your child still struggling, coming home and telling you, 'I don't know. This is hard,' tell them to come back*

and ask for help. This is the administration number, counselors, vice principals, strategies how to help your child, and my phone number."

The District English Learner Advisory Committee (DELAC) is a strategic force in adding the parents' voice to the process of reclassification. According to one participant, parents of elementary students who want their kids to be exited from ELD may be asked to consider the implications first:*"Our DELAC, 'I don't want my kid… I want my kid in the regular' and it's like this moral dilemma, okay, 'We understand that, we want to honor what you want for your kid, recognize there may be long term implications,' because if they're not explicitly taught certain aspects of the English language they're probably not going to get that."*

For parents of secondary students, the district seems more likely to agree with the concern as students who reclassify are then able to take the electives they want and the extra support classes they may need: *"From our DELAC perspective, parents are saying, 'Look, we don't want to hold them there forever. It ruins their options for electives when they've been held in ELD, because they have to have the extra ELD period.' Families are saying, 'But I want my child to have support in math. He's never speaking Spanish at home, and he has to be in ELD in middle school, but he really needs the support in algebra and he can't get that period because he's in ELD, so, help us let him out.'"* These examples illustrate how bringing in the parents can open up a two-way data use conversation in the reclassification process.

School Site Authority: Who is allowed to "opt out" of ELD?

When it comes to "opting out" of ELD, participants reported that parents had the right, although it was a rare occurrence. Some expressed concern over educational progress and believed a student would receive more benefit from staying in ELD: *"He has been marginally successful in all his other classes with the support, so without the support, where does that leave him?"* Others

discussed specific situations where parents made a convincing case for a certain elective: "*If they feel strongly about it, and they bring their concern that they want them to play an instrument, yeah. It's always band. I can't think of one instance where it's any other elective that they could possibly want."*

In Special Education, the question of parental input in opting out of services seemed to be twofold. On the one hand, participants were quick to say that parents could decline the ELD services. On the other hand, participants were unclear if certain primary handicapping conditions could or should be treated differently in terms of ELD services: *"I'm assuming that they can, given the way the laws work that if parents don't want a service, they can decline it. We're not going to make them take it. One of the things that we've looked at is, and that is a question for us, is if a child has autism, and they're an English language learner, can we, as a team, decide to not offer the EL instruction if we, including the parents, believe that that service is not gonna make a difference for that child. That the primary handicapping condition is what actually is impacting that child's language, not the fact that their parents may speak a different language but also speak English in the home."*

These scenarios suggest that the reasoning can vary when deciding whether or not a parent's request to opt their child out of ELD should be supported or encouraged. Such variation suggests a potential push-and-pull between the authority of parents and the authority of school personnel. These examples also suggest that decision makers are concerned about the broader scope of educational outcomes and opportunities for each child, not just those related to ELD.

Infrastructure and Tools

Although Coburn and Turner (2011) include tools in their influences on data use, Piety (2011) places infrastructure and tools at a more prominent level of influence over data use which

includes an overlap with organizational and political context. There were many examples in these data to illustrate the influence of infrastructure and tools on data routines, yet at the heart was the language used to discuss EL students and this language was branded with terminology taken from those systems. In this section, I will illustrate how categories, classifications and logics of action inhabit decision makers' interactions and conversations about data use routines.

Systems of Meaning: categories, classifications, and logics of action

Classification systems are embedded in how data are collected and presented to decision makers (Coburn & Turner, 2011; Piety, 2011) and the EL assessment system has multiple layers of these classifications, categories, and logics of action (Bailey & Carroll, 2015). Data use routines around educational programming provide the opportunity to observe how classification systems are addressed in decision makers' interactions. It is not unusual in California districts for EL students to be referred to by their CELDT level, e.g., "He's a CELDT 3", or the ELD class level to which they've been assigned, e.g., "He's an ELD 3." This kind of "classificatory talk" has been found outside of ELD as well, and is known to carry implications for how shared categories can frame interpretations of data and instructional responses (Little, 2012). There were several instances during the interviews where OPL and ELD categories seemed to both influence and constrain discussions of instructional improvement for EL students.

Participants used classificatory talk when speaking of educational programming and funding constraints. When considering what services can be offered, the numbers of students at each CELDT level determine where the focus will be. Instructionally, providing services to students at the beginning levels is more effective when students can be assigned to a class designated for beginners. This is rarely an option at secondary due to the small population of beginners and constraints on funding for classes with small enrollment: "*We just don't have the*

resources to really deal with Level 1s very easily, and that's been cut in the last couple of years anyway. We focus our resources on Level 3. Level 3s and 4s. A lot of our kids get stuck in Level 3. …It'd be nice if they actually went to the class they needed to go to, but like I said, sometimes the master schedule prohibits them to go. So we might actually have a 2 in the 3/4 because it's not offered." The use of CELDT OPL to describe students and services does not seem to constrain the placement of students with low levels of proficiency into beginning levels of ELD. In this example, the constraint was reported as the master schedule which by extension could be interpreted as the constraint of funding and consideration of the educational needs of the rest of the student population.

In some instances, participants explained how CELDT OPL categories needed disambiguation yet struggled to find other meaningful expressions. In the following example, the participant used another classificatory term, "newcomer," to clarify: *"We have EL assistants who work with the newcomers. And the newcomers -- the 1s and 2s sound almost like they'd all be newcomers, but there's the child who just got here who has a 1, and there's the child who is very language deprived who just got, has a 1. So the newcomer gets seen more than one time a day. They get seen two or three times a day. They get support pull-out twice and then get push-in to help with the class work. But the other group, they come and they join the newcomers and that's really about language, vocabulary building, sentence frames, you know, that kind of thing, more structure, and really getting them going because [for] a lot of the kids who are still 1s by the time they're in second grade, it's just vocabulary, language poor."* By disambiguating, the participant showed how services did not have to be constrained by OPL terms while at the same time showing an advanced understanding of the different reasons why students could be classified CELDT OPL 1 or 2.

In terms of other educational services provided, inferences about classifications seemed to be influenced by general beliefs about students' access to educational opportunities. In one example, an exception was made when parents requested that their child be allowed to exit ELD in order to take band: *"They don't, they can't do band. Well, I say can't, you know, we basically say, 'Nope, you're going to take EL.' But sometimes parents, and we respect that, sometimes they go, 'No, you know, grandpa played the trumpet. We really want so-and-so to learn the trumpet.' And so, we'll have a discussion with parents, say, 'Okay, fine.' If they were a 1 or 2, we'd really, really balk at it, but a lot of our guys are 3s, so, okay fine. If it's important to the family, we'll have them sign off saying, 'We recommend this [remaining in ELD], [but] we understand and honor the fact that they want to be in band.' I do support EL kids, you know, all kids taking the arts. And it's an unfortunate drawback that it's one or the other. Especially at middle school."*

In other specialized programs such as specialized college readiness programs (AVID) or gifted programs (GATE), the CELDT OPL did not appear not preclude participation: *"I think we also recognize the limits of the test. So, if we see a student that we think has potential, we move them in to the class."* There also seemed to be a system-wide approach to providing these options: *"For sure every single child is contacted. For example just yesterday AVID came. I had core, AVID kids came to talk to all the 8th graders, so, because my EL kids didn't have an elective, where they went in, they just came in to the EL class and talked."* School funding, however, can limit these opportunities for EL students at all levels: *"Well, we don't have a GATE program here per se, but I will tell you we certainly have gifted children. We have many children who are bilingual who are GATE."*

These examples seem to confirm the embedded nature of classifications and categories when speaking of EL students with implications for noticing, interpreting, and constructing implications for actions that could support or restrain academic progress. The prevalence of this type of classificatory talk suggests that the precision and accuracy of overall performance levels may be more impactful in the daily experiences of students, beyond just the determination of IFEP, EL and RFEP status and services.

Data Use: Noticing, Interpreting, and Constructing Implications

In the following section, I will describe data use routines within the context of two groups of students: students with disabilities (SWDs) and students who remain in ELD 3 longer than expected. These descriptions are intended to explore the way participants notice, interpret, and construct implications about CELDT and other data in terms of these students in EL programs.

Special Population: Students with Disabilities

In terms of data use routines for students with disabilities (SWD), it was not easy to get a complete picture from documents and interviews. While the SELPA manual is meticulous in providing guidance on most topics, there is no mention of EL testing in general, or the CELDT specifically. Participants admitted to having limited knowledge about CELDT accommodations or alternate tests: *"Nope. Don't know it exists, if it does. I'm assuming, I mean, we have nonverbal assessments that we can do with children. But I don't know of one for the CELDT. Truth be told, I don't really know what that assessment consists of."* In Special Education, the IEP is expected to trump the EL plan as the priority is federal compliance, but the testing mandates are not clear: *"Do we always have to give a kid the CELDT even if we know the kid's got language processing disorders? You know, do we have to do that? The feds say yes, but, the IEP still is the child's individual education plan, and you can determine what tests you give and*

don't give, and so, can you waive the CELDT?" This has direct bearing on the validity of the test scores for use in initial classification, subsequent educational programming, and eventual reclassification. In fact, several participants raised concerns about the inability of the CELDT to differentiate between students with language needs and learning disabilities.

Data Use Routine: Initial Classification

In terms of initial classification, there was a more urgent concern about identifying which language minority students may need services for Speech/Language Impairment (SLI) or learning disabilities (LD). The district uses bilingual specialists to identify these needs, at least for Spanish speakers, to be sure that they are not *"inaccurately identifying a child with a Speech and Language need or with some kind of Learning Disability or any other kind of educational need if in fact there's a language problem."* There was some awareness of the reverse, that is, language minority students who need additional in-class support, but not special education, and the possibility that language minority students could be overidentified as needing special education. Some participants mentioned overidentification in terms of the philosophy held by the prior administration and the mechanisms for identification that have since been replaced. However, there was no mention of the possibility of language minority students who may have been classified as ELs instead of students with a learning disability or speech/language impairment. The focus was, primarily, on making sure language needs don't inhibit proper designation of special education as had reportedly been the case in the past: *"It's really important to sort out first to address the child's language needs without Special Ed, as we're trying to do in this district through our school-wide interventions, before you say, 'Oh, okay, something's wrong with the kid, let's make him Special Ed, oops, it's because (he's) never heard English at home.' [Those are] big systems problems."*

In summary, documents provided by state agencies seem to lack coordinated guidance on issues related to students with disabilities who may or may not need EL services, and vice versa. It is unclear how the lack of guidance is influencing the district's ability to carry out assessment and special education services for language minority students who were identified as IFEP and not as EL. Without accommodations, and with a test that may classify even native speakers as EL (CDE, 2010), there seems to be little chance that a student classified as IFEP will be a student needing extra support services for a language disability or a speech and language impairment, among other possibilities. In fact, district data suggest that it is extremely rare to find language minority students with disabilities who are also identified as IFEP[8]. Taken together, this suggests that language minority students with disabilities would be identified as EL before being diagnosed with a disability, making it more likely that students will be dual identified as EL/SWD rather than a single identifier as SWD. This has implications for how data is used to allocate educational programming, and to determine which students qualify for reclassification.

Data Use Routine: Educational Programming

In the context of providing special education services, participants shared that it was hard to know from the CELDT levels and scores which students only needed EL services, which students only needed special education, and which students needed need both. Bilingual assessors are used to determine, for example, whether a language processing disorder is found in Spanish and in English. When EL students are unable to express themselves in English but when tested in Spanish have no difficulty communicating, then *"there's a different kind of intervention that is required."* Participants noted that different procedures may have been used in the past, with different outcomes: "*I don't know that the practices and policies and procedures were*

[8] In 2012-13, no IFEP students were receiving special education services. Data from prior years are similar.

implemented with the same value system that we may have. So there very well could be kids in Special Ed who are English language learners that, if they were going through an IEP today, might not be."

For English learner students with disabilities (EL/SWDs), participants were careful to mention that the EL designation was just a marker of EL status and CELDT data was only occasionally considered as helpful to the IEP process: *"Sometimes in an IEP when you want to look at progress on goals and objectives, if a student is making progress towards or is not making progress and the student is an English language learner, we may refer back to the CELDT scores just for some clarification to help us solve that puzzle about, 'Why is this child still not making progress?' Or, 'Gosh, this child has made so much progress, met these goals and objectives, and is completely at grade level. What does the CELDT say? What? This kid is still an English language learner? I don't think so.' So it could go either way. But it's more of a reference point. ...The IEP super-cedes really everything. And an IEP team super-cedes everything. So, if an IEP team says, 'This child has met all the goals and objectives and is no longer eligible for Special Education and is functioning at grade level' it's absolutely irrelevant what the CELDT may say about the child's language needs in terms of Special Ed."*

It is notable in this exchange that EL reclassification is not mentioned, nor is "functioning at grade level" interpreted in terms of still being an English learner or not. With this in mind, it may not be surprising that when it comes to Special Education programming, the CELDT is only used to identify if a child is EL, nothing more: *"In an IEP the CELDT is simply, tells us whether the child is EL or not. And that's all we do with the CELDT. Are we confirming that the kid is EL? So we don't drill down at all. It's just not the practice that we ever felt we needed in terms of Special Ed eligibility which is so microscopically, legislated and formulated that, we pretty*

much know. So the CELDT doesn't give us anything that would assist, because with the IEP, as you know, we're looking at goals and objectives, we're looking at an instructional program, and so we get that really much better from an individualized Special Ed assessment. [The CELDT] doesn't seem necessary because we have to do a Special Ed assessment anyway, so we're going to rely on our Special Ed assessment rather than the CELDT."

Another participant shared a situation where Special Education services took precedence over ELD on the master schedule. If it's a choice between EL and Special Ed, then *"they're in Special Ed…(and) we don't pull them out of any of that to go into an EL class,"* so ELD is provided under the auspices of the resource room or special day class. When EL students in special education don't make educational progress, however, it is difficult to know what the cause is. As one participant noted: *"The piece I'm really trying to push is that we've got to take our kids as far and as fast as we can. Because the thing that happens, when I see the 3s, you know, in the past they've been pulled out and pulled out and pulled out, and so… when they come to us [in middle school] and they're still failing, is it a Special Ed issue? Is there a processing issue? Is there a [language] piece?"*

In summary, there was an emphasis on allowing students to respond to intervention and extra support before proceeding with Special Education testing: *"We're very, very good in this district about what can we do. …**We don't rush to assessment or identification**."* However, there was no mention of students with disabilities classified as EL who may not need to be in an ELD program.

Data Use Routine: Reclassification

The participants who were most closely affiliated with Special Education had little direct experience with reclassification. One participant stated that "certain qualifications" had to be

met. In regards to EL/SWD students, it was unclear to participants whether or not the students are able to reclassify or, if so, what the procedure was. The special education manual provided by the state, and the website county office that provides special education guidance (SELPA) had some information about serving EL/SWD students, but no mention of reclassification.

In the prior year, the school district had initiated a new procedure whereby students with a learning disability could be redesignated out of ELD if it could be proven that their inability to meet reclassification criteria was due to their learning disability. When discussing this possibility, one participant and I examined recent CELDT scores for one EL/SWD student in RSP which were: *OPL=4, Listening=4, Reading=3 (missing Level 4 by one raw score point), Writing=5, and Speaking=5*. The student's CST-ELA scale score was 275 (Below Basic). This student's IEP indicated a visual processing problem and provided educational goals for both visual and auditory. To verify that the missed reclassification criteria were due to a learning disability, all the student's teachers must approve that *"this is not a language problem, this is a learning problem"* in conjunction with the IEP. Even with the new guidelines, reclassification for EL/SWD students in RSP was said to be infrequent *"because the resource teacher isn't looking for [it]."* This seems to indicate that EL/SWD students might not be reclassified, even when there is a way for them to be reclassified, suggesting that EL/SWD students could eventually be considered long-term ELs.

Special Population: Students "Stuck at 3"

According to participants, students with CELDT OPL 3 are not at all the same. Some OPL 3 students moved quickly up through the levels, while others seem to get "stuck" in ELD 3. These so-called "stuck ELs" were said to need something different than the other OPL 3 students. In this way, an OPL 3 on the CELDT did not seem to be explanatory enough to indicate which ELD

services would be most appropriate and effective. One participant shared: *"In the absence of other formative tools, how am I going to tell someone who's got an EL 3, who's been an EL 3 for 5 years, there's the federal mandate so they give that kid a program, so I'm required to provide that child daily ELD, but that daily ELD may be in a group of 3s that have got all sorts of different needs and levels, and so it may not be the best thing for that kid. Or, the 4 and 5 who, yeah, you can read and you're compliant so you end up in a class where, you know, you're less compliant and you get to stay at 3. We default on the CELDT scores to decide where kids go. None of us override the CELDT and say, 'Oh no, let's just not provide a 1 or a 2 support'. You know, it's like, we have to, and so you put them in those boxes, and we don't have a good tool to get them out of those boxes other than the CELDT, which is once annually. …We need to get to where we're really doing something with the long-term EL kids and saying, 'You didn't grow, and we're going to do this for you.' I don't think we really know what to do for the kids, other than in reading, you know? …The CELDT doesn't help us pick what to do."* Schools have different ways of handling the different programming for OPL 3 students. At some sites, *"anyone who scores 1 or 2 on the CELDT is put into our immersion. 3s, if it's a high 3, they go into the mainstream English, low 3, they're going into the first program, 4s and 5s are in the mainstream"* but how the low/high distinction is made depends on the site: *"I'm not sure how each school does it. Some person, people, will just look and say, oh they scored, they scored a 3, they're gonna be placed in our program 3. Some will look at it and actually do the actual division, and say, okay, were they high, were they low, and then decide which program to put them in. [At the elementary level] there's not a big difference between both the programs because they all have teachers that have their CLADs, the SDAIE."*

There was a perception among some participants that students who get stuck at 3 may be students with learning problems. *"Another place we use the CELDT scores is in IEPs. It's always part of an IEP if a student is an English language learner. Sometimes it's very confusing as to whether a student has a learning problem or a language problem. And, we look at that to see if, you know, sometimes they get locked at Level 3 because of their learning problems. So their CELDT information becomes part of the picture."*

At minimum, most participants shared the belief that the CELDT did not give them all the information they needed to know what to do next with the "stuck" ELs, but did not openly question the veracity of the "non-proficient" classification given by the CELDT. In other words, participants didn't doubt that the students who remained in ELD needed to be there, but seemed at a loss for what to provide the students who didn't seem to benefit from staying. The way participants interpreted the notion of a "stuck" ELs seemed to indicate that they had noticed that information was still missing, that the CELDT was somehow not reporting the data that may help them unlock the secret to the "stuck" or "stalling" ELs: *"I wish there was a way, if we're in a perfect world, I wish there was a way to somehow tell if the reason that the student is struggling is because of an actual language barrier or a learning barrier. I think that sometimes we get caught up that it's the language, but realistically, students … have been in American schools since kindergarten and they're stalling out at 3s and 4s – there's gotta be a reason for that. They're getting the same background, the same building blocks as everybody else, but they're stalling, and so I don't think that the CELDT really is able to show that. I think the CELDT works better for students that are new to English. You can see their progress and how it says, you know, they go up one [level] a year. I mean you can see that, but the kids that started school here, I don't think it's as helpful."*

Many participants shared that sentiment, with one noting "*a lot of students that were born across the street just kinda stay at intermediate for three, four years."* Another participant shared: "*Generally, kids who are the long term 3s, end up being kids that don't show a lot of proficiency on the CELDT – I mean on the CST either, and so, they're held in because they're not academically proficient and they're not linguistically proficient, but they also may not be true English Language Learners. So you have that group of kids that's just kind of stuttering in our system, where they've been here since K, they've been in ELD their whole life, and they are still considered an English language learner although they never had a second language. English has been their only language."*

Inferences about Role of CELDT

Along these lines, other participants commented that there were questions on the CELDT that native English speakers would be unlikely to answer correctly, leading them to be concerned:*"I mean sometimes kids, for example, you can't get a 5, at least it used to be, you can't get a 5 on the CELDT if you do the their-they're-there wrong, or the you're-your. Those types of things? Well, native English speakers do that too. Do you know what I'm saying? That's just such a common error in middle school. So I just wonder sometimes if kids sometimes just get held back because they're just making common mistakes that should be dealt with, that would be dealt with in by any class?"* Participants also mentioned specific vocabulary words that seemed to be difficult for their students, like *microscope*, *currency*, and *portrait*: *"So, I'm sometimes like, wow, that was hard. …Sometimes the vocabulary that's thrown out at them, just like any standardized testing, it's possible that a student that has never had a science class, probably doesn't know what a microscope is. You would hope by 8th grade they would have used a microscope, but really in elementary school they wouldn't have used it, so to have that on the 6th*

grade CELDT as a vocabulary word is kind of odd." As sixth grade students take a 6th, 7th, 8th grade cluster test, so it would not be unusual for higher-level academic words to be present. However, this seems to indicate a potential problem or interaction effect with the cluster format and the academic words chosen for the CELDT test.

Participants mentioned that secondary students experienced testing fatigue with the CELDT and the CST, which then became a sort of "EL fatigue" from being in the program too long. Some connected that extended time in the program with students' expressions of lack of confidence and motivation: *"If they have great grades, can't pass the CELDT; if they have good CELDT, can't get the grades. I think it, it affects them. And there is a sense of, how do you get them to believe that they can? When you have a student that's been in the system for ten years, and hasn't been reclassified, they don't have the confidence that it's ever going to happen."*

Inferences about Student Motivation

One participant shared how discouraged students had become around the time of reclassification decisions and testing: *"I had so many of them that were just going to quit. And, at that point, in fact, they'd been here since second grade, and one of them said, 'You just need to stop worrying about it. This is just how it's gonna be.' …'I can try, but I'm gonna fail' is what he said."* A few participants were of the opinion that unmotivated students were affected by negative circumstances in the home such as possible lack of parental support, lack of home literacy, and other factors related to what one participant called "culture." One of these participants expressed frustration at a lack of compliance among some students during the testing: *"The other thing is they absolutely refuse to take the CELDT seriously. We had a student earlier in the year that was just angry, that was just overall angry, and he just refused to write anything on the CELDT that even pertained to the test. He wrote, just nothing that was -. He's just an angry young man."*

In contrast, motivational issues were also addressed in terms of what a school site could provide. It was in the context of a discussion about a college- partnership program offered to EL students that one participant noted how immigration status could impact motivation for college readiness. When students' parents are unable to provide support towards college, students are left to depend on their grades and other school achievements, making it even more crucial that EL students stay motivated to persevere in school. In the words of this participant: *"I'd have students [say], 'I'm a, you know, my parents are undocumented. There's no way I'm gonna get to college.' So, I mean, there's a lot of things in their life that, yes, they would love to do that, but they need to have more internal stamina and motivation because, you know, they're school dependent. They're trying their best, but when they go home, there's a lot of life that they have to deal with."*

Inferences about Student Educational Needs

Participants also noted that the educational programming provided to EL students could have played a role in the current phenomenon of long-term ELs. Some noted that there could be deficits due to lack of clear instruction in vocabulary and syntax. Others mentioned a lack of exposure to native speakers of English. This was due in part to an old policy that had led to middle and high school students taking ELD in place of their English/Language Arts for regular English credit. While this increased the chance that students could graduate, the policy reduced the chance of being in classes with native English speakers, and access to the classes needed to meet college readiness requirements. Differences of opinion on this policy seem to reflect differences of opinion on likely pathways for language minority students, even among the students themselves: *"The kids didn't want out. Of ELD. 'Cause it was their only English class, and then we changed it… I told them, 'It is based on your state test scores, that's the only thing*

holding most of you back.' And they were like, 'Yeah, but, we don't wanna go.' And then when it got changed [to an elective], I don't remember the exact numbers, but it was pretty high, to where they got their state test scores back and we were redesignating like crazy."

At the elementary level, there was also a type of unofficial policy at some sites regarding goals and expectations for EL progress: *"There for a while, in '03-'04, the elementaries were getting kids, their goal was to get kids to 3. And then they'd come to the middle school. But what happened is they didn't get access to the enrichment [elective] classes. Because, 3s, then we had to put them in intervention, so it came out of their enrichment. ...But then in talking with the principals, they went, 'Oh man!' So they started really ramping up what they did to get kids to 4s and 5s and have them redesignated by the time they came to middle school."* This was confirmed by another participant who shared: *"We really have changed our focus to truly try to have kids reclassified and redesignated by the time they leave elementary school, because of the sort of fear that they may never be redesignated if they don't do it then. Just because students change, the difficulty of the test changes, the buy-in of the students taking the test changes, and you find that it's far too many students are just never reclassified, and that can be a really dangerous slope for the students, because then they're not allowed to take, you know, more rigorous academic classes often, or more other kind of electives, and it just seems to, I don't know, they become a little bit, disenfranchised."*

Participants expressed concern for student stuck in ELD in relation to school completion. According to the analyses conducted by the district, some long-term EL students drop out, while others end up in an alternative school setting: *"We look at our credit recovery, so many of our kids are disproportionately EL, our dropouts disproportionately ELs. ...you have to look at all the other kids who are those casualties, who by 10th grade, 11th grade are in credit recovery,*

because that EL model didn't work for them so their de facto EL model becomes credit recovery." It was explained that students fall behind in their high school credits from taking ELD classes which supplant their electives. While these ELD classes are essential points of access for some students, without which students would not be able to successfully complete coursework in English, it was apparent that for other students the ELD requirement is a barrier.

By the time students reach high school, these pathway decisions should be theirs to make. However, there is a direct impact of CELDT data on students' access to earning college readiness credits. While the reclassification criteria are a necessary check-and-balance to ensure students' English is sufficient for success in an all-English educational setting, it seems that it is not necessarily the English ability, but rather the scheduling of ELD, which can prevent students from having access to courses needed for college readiness. Thus, any misclassification on the CELDT could ultimately impact college readiness. As one participant summarized: "*That ELD period, if you're a 3 who's been a 3 forever, deprives you of access to other content. And that's based on what your CELDT level is."*

Data Use Routine: Inferences about Reclassification (for all students)

Regarding reclassification for all students, the goals at elementary differ from goals at middle and high. Participants at the elementary level talked about not reclassifying before 4th grade. It seemed as though 4th and 5th were the prime grades for reclassification because students *"get to that point, they've been here, they're doing well academically."*

Many participants stated that the goal of the district was to reclassify as soon as possible. Some said this was because *"the test gets harder"* while others cited missed opportunities for students staying in EL programs too long. Staying in too long was often defined as not reclassifying before middle school, or high school. Reclassifying by high school was commonly

stated as an imperative. However, the consensus seemed to be that the students at grades 6-12 were expected to get themselves reclassified. Several participants made this point clear. In the words of one: *"By the time they're in high school, they wanna go, they wanna be reclassified, but they don't try. They don't realize that they have the power to get themselves out of there. …I get lots of calls, 'Well my kid doesn't like that class. He doesn't wanna be there.' I said, 'You're right. And he shouldn't be there,* ***but he needs to get out of it.*** *You just can't call me and say 'I want him to be pulled out of that class.' You know, there's certain requirements. …He's gotta score, he's gotta take that test seriously.'"*

Some participants pointed out that the use of grades and test scores to determine educational programming was unique to EL students: *"It's an interesting reclassification system. I mean,* ***a student can hold themselves back*** *from being reclassified just by not getting their grades, and not participating in class work too, whereas their counterpart who is English only can not be getting grades but not -. But you could also say they're getting a support class that the other student isn't getting."*

Some participants, even teachers, saw that their role was to encourage students to take the tests more seriously in order to qualify for reclassification. Even though the test scores were not universally viewed as usefully descriptive, the onus for reclassification was placed squarely on the student: *"Yeah, sometimes I think the CELDT doesn't give me as much information as the ELA portion of the CST because sometimes students are not as motivated in the CELDT as they might be with the CST, for some reason. Sometimes they have to have a pep talk, to say, 'If you wanna not be an ELD student, you know,* ***you're given the test now, and this is a ticket to move out****, and you really don't want to be in high school as an ELD student. You might wanna take auto shop, you might wanna take art, you might wanna take an elective and this is your way out.*

So, you know, I think that you should try your best because you might wanna do some real fun things that are your interests.'" In conversations like this, it was notable that there seemed no question of students' readiness to reclassify. The focus was placed on the students' motivation and subsequent ability to reach the standardized testing criteria. It wasn't always just the CELDT that students weren't able to pass, or the California high school exit exam (CAHSEE), but sometimes just the CST-ELA at the criteria level set by the district: *"A lot of them graduate without being redesignated. ... I would say that most of them are passing the CAHSEE and their CELDT but not the state test score. There's been nothing linked to their state test score, they don't have to, there's no grade, there's no, you know, anything. I do feel like in the last two years,* ***finally, students have caught on that they need that to get redesignated,*** *each state test score test gets harder."* For the students mentioned here, it was in fact easier to graduate from high school than to get redesignated.

Perhaps this is one reason why participants used metaphors of containment when speaking about ELD and EL status, especially for secondary students. Participants spoke about reclassification as the chance to *"get them out"* or to not *"hold them in."* A few participants described reclassification as a marker of readiness for a new learning environment: *"In middle school, you don't want to send them up to high school without any support, if they need it. So 7th grade, it's more like, 'Well, maybe we should hold on, if possible, 'cause they're on the cusp.' Whereas 8th grade we may be like, 'Well, are they ready to handle an elective? Are they ready to lose their EL class?'"* Even when acknowledging student readiness, participants were quick to re-state the necessity of meeting test criteria: *"There are some kids that, yeah, they're ready, and so maybe you can go move them up. But usually the numbers have to be all in line for it to work."* When participants spoke about reclassification at the elementary level, many suggested it

was the school's job to get kids reclassified: *"Our plan is to **move the children out** by the time they're in, about 5th grade so that they leave elementary school reclassified."* So even while reclassification is a goal, and it should indicate readiness for the next educational setting, the participants' comments seem to suggest that in secondary school, the students carried the responsibility. It was up to the students to be sufficiently motivated to meet the criteria.

Another aspect of readiness to reclassification was the reality of what was expected when entering mainstream classes. Most teachers in California now have a teaching credential that includes cross-cultural, language acquisition and development certification and some training in sheltered instructional techniques. However, the language in the teachers' collective bargaining agreement suggested that teachers were still allowed to choose whether or not they want EL students in their class. It states:

> The district will solicit volunteer teachers for placement of full inclusion and LEP/NEP students. If no teacher volunteers are forthcoming, placement of the students will be done in the usual manner.[9]

Participants shared about teachers who had used the language of the policy when questioning why EL students were in their classrooms, e.g., *'I didn't volunteer to have these kids in my U.S. History Class,'* and *'Please advise me about what [the] English Language Learner placement policy is. I don't understand why I have these kids in my class.'* These sentiments may be expressed infrequently, yet they suggest that EL students could be the object of teacher resentment in classes outside of ELD. The extent of possible effects of this policy on students' motivation to be reclassified is a topic for future study.

[9] From "Agreement Between XYZ Educator's Association and XYZ School District, July 1, 2011 through June 30, 2014: The Teacher's Collective Bargaining Agreement. Article 18 SB1870 AB3075 Special Education, Section on 'Full Inclusion/LEP/NEP students', p. 114.

CHAPTER SIX

DISCUSSION AND CONCLUSION

This chapter provides a summary of the key findings from quantitative and qualitative analyses, and discusses the results of the study in terms of implications for policy and practice. The significance of the study and its limitations are discussed in terms of directions for future research.

Summary of Findings

This study sought to describe and explore the influences and constraints on data use routines in local control state in terms of initial, reclassification, and educational programming through the analysis of state and local documents, interviews with district personnel, and analysis of student performance data. Influences and constraints on data use were described and analyzed.

Composite and Aggregate Decision Rules

Findings from this study suggest that the choice of decision rules, composite and aggregate, can have an impact on which students remain in ELD and which students are reclassified with implications for serving "stuck at 3" and long-term EL students. The finding that the district aggregate outcomes were not an exact match for any of the decision rules suggests that additional criteria may have also affected reclassification decisions, such as other measures, teacher recommendation, and parent consultation. However, this cannot be confirmed without further study. In addition, findings suggest that beyond the decision rules there are other school-based procedures which constrain reclassification. The interaction between composite, aggregate, and local routines and procedures suggests that broad analyses of reclassification trends could make false assumptions about who could be in the long-term English learner group. For example, in elementary schools that don't reclassify until fifth grade, students who have been an EL since

kindergarten will have already met one criterion for the long-term EL designation (6 or more years as an EL) without yet having the chance to reclassify. Likely, students in that year where reclassification finally happens will be students who could have been reclassified sooner and have higher CELDT scores than newcomers. California has chosen to exclude high-performing long-term ELs from the long-term EL subgroup. There is no designation for EL students remaining in EL program yet have already met CST-ELA criteria, or even the CAHSEE. This population of high-performing EL and EL/SWD students remaining in ELD programs depends on CELDT decision rules to be accurate, precise and fair.

While the main goal of decision makers is to reclassify and exit students, these decisions simultaneously decide who will remain in EL programs. This means that while reclassification decisions may consider multiple sources of data, the retention decision can rest on only one. In this district, retention in K-1 is decided by a policy of not reclassifying at these grades. In grades 2-12, retention was unlikely to rest on parent opinion and teacher recommendation as these were expressed to be confirmational in nature. Rather, students continuing to receive EL programming for an additional year were likely to receive this placement based on CELDT or CST-ELA evidence. While CST-ELA evidence is based on one cut score, CELDT evidence can be based as many as nine cut scores: level 3 cutscore in 4 subtests, level 4 cutscore in 4 subtests, plus one cutscore for the OPL. When decision rules have multiple cut scores, reliability and precision decrease exponentially. Districts should use caution with considering mixed decision rules.

Student performance data from state assessments were gathered for analysis to shed light on the extent of difference between decision rules for EL in two groups: EL, EL/SWD; and two sub-groups: EL/HP, EL/SWD/HP. The results of applying six composite and two aggregate decision models indicate that differences do exist between composite rules, between and within

models, and between and within grade levels. Analysis of the actual district aggregate rule indicates that what was indicated in district personnel interviews – that some elementary schools wait until 3rd or 4th grade to reclassify – was verified through the description of grade levels of students remaining in ELD programs. Descriptive data about students who weren't reclassified seem to indicate a wide range of abilities, including students capable enough to pass the CAHSEE. This illustrates how aggregate rules, and even more so the composite rules, can influence which students will be participating in ELD, and at the same time potentially reduce the district resources needed to run a fully-staffed ELD program.

Timing of assessment and results

State policy influences when "official" CELDT results are delivered which can constrain the timely placement of students into programs and services. As decision makers wait for "official" scores, they defer to state timelines. This impacts how districts view the "unofficial" scores, and how they use "unofficial" data in the meantime. This can also affect what time of year students are reclassified, for example in February, even if the timing is less optimal for students' educational programming. These different seasons of urgency are ever-changing with new tests, new delivery systems, and new timing of accountability reports.

Determination of decision rules

In California, districts may exert local control and determine composite and aggregate rules for reclassification and educational programming. However, districts defer to the state for initial classification decision rules. Participants shared several different composite rules that had been or were currently used in the district, including site-specific guidelines about which grades typically are allowed to reclassify. Some decision rules facilitate reclassification as soon as possible, while other rules and school-site policies are chosen to promote extra time in ELD

before students are allowed to reclassify. This illustrates how the authority to enact decision rules comes with the power to define "readiness" for program exit and entrance into mainstream educational setting. This "readiness" may also contain the notion of the district's readiness to lose Title III funding, with all the resource and staffing implications. Thus, the authority to set composite and aggregate decision rules is the authority that ultimately prevails in deciding who is initially classified EL or IFEP (state rule), and when students in EL programs remain EL (district rule) or are reclassified as RFEP (district rule).

When districts use a classifying decision rule that can leave students with CELDT OPL 4 and 5 in ELD, it means that the definition of "readiness" for exit is limited to certain students with OPL 4 and 5. In this district, some EL students in secondary with OPL 4 and 5 are receiving all instruction in mainstream classes with support from bilingual aides. One reason for this is the lack of funding to provide a separate 4/5 ELD class, and creating a 3/4/5 class would step outside of state guidelines which require no more than two CELDT levels in one classroom. This illustrates some of the challenges confronting smaller districts who have EL students in all five CELDT levels. If students are at OPL 4 and 5 and need ELD, yet no ELD class is available that meets their needs, the benefit to staying in ELD is unclear and presents a challenge for districts aiming to limit the number of students slipping into long-term EL status. This also illustrates the extent to which classificatory decision rules are closely knit with educational programming and suggests that evaluations of decision rule effectiveness must consider educational programming and placement outcomes. How those in authority decide on these rules is a question worth asking, although beyond the scope of the current study.

The impact of reclassification on the staffing of ELD programs is also a topic in need of research in the current policy context. In the case of initial classification, the district defers to

state policy, and this is also a topic in need of future study. Since the CDE study (2011) found the possibility of misclassification during initial testing, the impact of state chosen decision rules is worth investigating in terms of the influence on students' access to appropriate programming.

Relations of Authority and Power

According to state education code, parents are granted a voice in data use routines. It seems that DELAC can provide a venue for that conversation with the district which can have an influence on local decision rules for RFEP and educational programming. More could be known about parents' involvement in these decisions, and future research could explore the dynamics of DELAC participation in terms of student outcomes towards reclassification, successful high school completion, and college readiness.

Infrastructure and Tools: Classificatory Talk

There were many examples in these data where infrastructure and tools had an influence on data routines. At the heart of tools such as the CELDT are the classifications and categories which can emerge in the use of "classificatory talk" (Little, 2012). Classificatory talk illustrates what participants are noticing when in the process of data use decisions. Examples seemed to indicate that the noticing of available resources, by way of system constraints, was paired with the noticing of students' linguistic needs. Decision makers used CELDT overall levels to describe students, which were sometimes aligned with ELD programming. However, constraints on resources did seem to impact the ability of participants to act on their knowledge of language proficiency levels in scheduling educational programming. In addition, there were examples of other programs – like band, AVID, and GATE – where CELDT OPL was not considered a constraint.

Data Use for Special Populations: EL/SWD

Documents provided by state agencies seem to lack coordinated guidance on issues related to students with disabilities who may or may not need EL services, and vice versa. It is unclear how the lack of guidance is influencing the district's ability to carry out assessment and special education services for language minority students who were identified as IFEP and not as EL. Without accommodations, and with a test that may classify even native speakers as EL (CDE, 2011), there seems to be little chance that a student classified as IFEP will be a student needing extra support services for a language disability or a speech and language impairment. In fact, district data suggest that it is extremely rare to find language minority students with disabilities who are also identified as IFEP[10]. Taken together, this suggests that language minority students with disabilities would be identified as EL before being diagnosed with a disability, making it more likely that language minority students will be dual identified as EL/SWD rather than a single identifier as SWD. This has implications for how data is used to allocate educational programming and how data is used to determine which students qualify for reclassification.

There was an emphasis on allowing students to respond to intervention and extra support before proceeding with Special Education testing. However, there was no mention of students with disabilities classified as EL who may not need EL services. This suggests a lack of knowledge or understanding of the risks of misclassification. Findings also seemed to indicate that EL/SWD students might not be reclassified, even when there is a way for them to be reclassified, suggesting that EL/SWD students could become long-term ELs.

[10] In 2012-13, no IFEP students were receiving special education services in this district. Data from prior years were reportedly similar.

Data Use for Special Populations: Students "Stuck at 3"

The students who are "stuck at 3" are a heterogeneous group according to participant accounts. There seems to be a focus on figuring out what these students need educationally, what will unlock their motivation to exit the EL program, and concern about the implications of some students having unidentified learning disabilities. The term "stuck at 3" seems to encompass two groups: (1) students who remain in an ELD 3 class, and (2) students at CELDT OPL 3, 4, and 5 who haven't yet met the reclassification criteria. In this way, "stuck at 3" is a broad category which could limit decision makers' noticing of potential solutions to challenges faced in programming and progress towards proficiency.

Infrastructure and Tools: CALPADS

Taking a sociotechnical perspective, infrastructure and data systems are likely to evolve over time as part of an organizational context and can be altered: "As states enhance their infrastructures and as districts connect to them with their own local technologies, the specific data available to practitioners working at different grades and with different subjects will change in ways that mirror and are related to the organizational and political contexts" (p. 219, Piety, 2011). This was observed in how the introduction of CALPADs forced a new EL status "Wait to Classify" to emerge and be used by the district in order to meet accountability obligations while also serving student needs. Similarly, the reports created by decision makers when operating between old and new reporting systems could be considered a type of bridging system to continue the work of school during the transition. These "designed artifacts" are the "products of choices made under constraints" (Simon, 1996, in Piety, 2011) and are rich in potential for future research on how decision makers negotiate the boundaries of infrastructure systems. In the upcoming transition to new ELP and state standards-based assessments in California, decisions

about EL students will continue to be made, acted upon, and reported in the context of a new infrastructure. It seems likely that such "designed artifacts" will be a part of that transition.

LIMITATIONS AND FUTURE RESEARCH

This study examined data from a mid-sized school district in California with a predominantly Spanish-speaking language minority student population. Potential differences in the fairness of decision rules for different language groups were not able to be tested. Since no data was available to describe the socio-economic status (SES) of the students whose data were used for this study, differences in fairness based on SES were also not able to be determined. Future research could replicate the analyses from this study in districts with more diversity in languages and SES to answer such questions. Unmeasured factors, such as absenteeism, student motivation, and/or access to high-quality instruction, may have affected scores and thus efforts should be made to control for these factors in future studies. Future studies might also include a more comprehensive investigation of decision rules considering the effects of each step in the test development process from students' raw scores to final classifications. Combined with an analysis of the likelihood of misclassification at the entrance to EL programs, would help evaluate factors which impact decision rule performance and thus decision rule choice. Finally, future research might consider the impact of the ordering of reclassification criteria and the mechanisms which may or may not be in place to challenge ELP assessment evidence in terms of delays to reclassification, among other aspects of decision making (see also Honig & Coburn, 2008; Moss, 2007; Spillane, 2012). Mixed methods approaches are recommended to address these issues (Bailey & Carroll, 2015) as well as research-practice partnerships which could provide mutual benefit in real time to school districts, states, and researchers.

IMPLICATIONS FOR POLICY AND PRACTICE

The hypotheses generated from this mixed methods study are intended to stimulate discussion about how measurement professionals, policy makers, test vendors, and state- district- and school-level educators interpret the risk and cost of evidence used for high-stakes decisions. What follows here are topics to begin that conversation.

Decision Rules Effects on Programming and Long-Term ELs

Because the antecedents to the long-term EL phenomenon are complex and interrelated, this study used a mixed methods approach to examine decision rule outcomes in light of educational programming. Participant interviews, in conjunction with findings from quantitative data, indicated that schools struggle to provide all levels of ELD. Students with language proficiency at CELDT OPL 3, 4 and 5 can be in a "stuck" situation where their skills are not sufficient for exit, yet the ELD programming available at some schools may no longer be an appropriate match for their linguistic abilities and needs. Understanding more about this could lead to important conclusions about how to best serve EL students while they are in ELD programs. The policy and practice implications are twofold.

First, schools, districts and states that employ decision rules that can leave EL students at the top levels of the CELDT in ELD classes should be called upon to justify those decision rules in light of the ELD programs they can offer. Second, test vendors and policy makers should provide the guidance, tools and expertise to enable an evaluation of the match of decision rule and program. Providing validation documentation for decision rules as part of technical documentation of assessments is unlikely to be as effective as creating dynamic systems which

integrate real-time student data and current educational placement options when considering decision rule effectiveness.

Furthermore, more research is needed on not just the annual assessments but also the screening or initial assessments. The number of students entering EL programs is directly related to the decision rule applied to the initial assessment. These rules and outcomes also deserve evaluation, especially for students with disabilities who may be at the greatest risk for misclassification in this system.

Reporting Standards for Validity of Composite and Aggregate Rules

Currently, there are no guidelines in the *Standards* (2014) for evaluating multivariate decision rules. In the context of the new English language proficiency assessments currently being developed, a coordinated effort is needed between researchers, test vendors, and psychometricians on standards, methods, and reporting of multivariate validation, not just for technical audiences but also for data users at states, districts, and schools.

Funding and Staffing of ELD Programs

California has enacted a new Local Control Funding Formula (LCFF) which may inadvertently provide disincentives to reclassifying students. As schools reclassify students, funding is lost which further limits the range of services schools can offer. For example, in the District in this study, the funding allocated for the district Title III coordinator was equal to a 0.20 FTE in 2012-13, and thus the position was held by a teacher with a 0.80 FTE teaching position. Due to the dwindling number of EL students remaining in ELD in 2013-14, the allocation for the Title III coordinator was reduced to 0.00 FTE, and thus subsumed into another person's job duties. These resource challenges are worth investigation. Future research should consider the value of

conducting such research at small- to mid-sized districts in order to fully appreciate the impact of changing reclassification rates and resources. Policymakers may wish to consider developing a calculation of the minimum funding needed to staff and run an ELD program, for example, in small to mid-sized districts which may have a few EL students at each grade level and at each ELD level. Research on size of program, staffing, and time-to-reclassification could provide useful baseline information for policy development.

CONCLUDING REMARKS

This study was designed to describe the influences and constraints on the data use routines of classification, reclassification, and educational programming for language minority students. The investigation and the results it has produced have provided a more detailed understanding of potential antecedents to the long-term EL phenomena, including a description of effects of decision rules on outcomes for rarely studied EL groups: high-performing students and students with disabilities. The insights gained from this study could prove useful in future efforts to serve EL students, and to study effects of data use routines and classificatory decision rules on all students.

LIST OF APPENDICES

Appendix A: Abbreviations

Appendix B: Study Invitation

Appendix C: Interview Protocol

Appendix D: Study Information Sheet

Appendix E: Students Meeting CELDT Criteria in Study District, 2012-2013

Appendix F: CELDT Raw and Scale Scores for Level Cutoffs for 2012-13, by Grade

Appendix G: McNemar's Test Results

Appendix H: Codebook

APPENDIX A

ABBREVIATIONS

AMAO	Annual Measurable Achievement Objective
CAHSEE	California High School Exit Examination
CALPADS	California Longitudinal Pupil Assessment Data System
CDE	California Department of Education
CELDT	California English Language Development Test
CST	California Standards Test
DELAC	District English Learner Advisory Committee
EL	English Learner
ELA	English language arts
ELD	English language development
ELPA	English language proficiency assessment
FEP	Fluent English proficient
GPA	Grade point average
HP	High performing
IFEP	Initially Fluent English proficient
LCFF	Local Control Funding Formula
LEA	Local Educational Agency
LEP	Limited English Proficient
Non-EL	Native English speaker
OPL	Overall proficiency level
RFEP	Reclassified Fluent English proficient
SBE	State Board of Education
SWD	Student with a disability

APPENDIX B

STUDY INVITATION

An Invitation to Participate

You are invited to participate in a research study conducted by UCLA that is investigating the role of the CELDT (California English Language Development Test) in the classification, reclassification and educational programming decisions for students in grades K-12.

Who can participate in the research study?

"District" employees who have direct knowledge of district procedures for EL students

What is the study for?

The purpose of this study is to learn more about how data from the California English Language Development Test (CELDT) is used to make decisions about EL students regarding their classification, reclassification and educational programming.

Do I have to participate in the study?

No. Participation is entirely voluntary.

If I participate in the study, what will I be asked to do, and how long will it take?

You will participate in a one-on-one interview where you will be asked some questions about the procedures for CELDT data use in your district. These questions will help the researcher better understand how CELDT data is used to classify, reclassify and provide educational programming for language minority students in K-12. Your interview will be recorded for note-taking purposes and used only by the researcher and will take no more than 30 minutes. All responses will be kept strictly confidential and, if reported in the research study, will be kept anonymous.

What are the benefits of participating in the study?

- → You will have a chance to help researchers better understand the district perspective on the CELDT with the hope that better understanding will lead to more effective test development for and state policy regarding the education of language minority students.
- → You will be entered in a drawing for a chance to win a thank you prize: a $25 Amazon gift certificate. Up to eight participants will win.

How do I volunteer?

To volunteer, please contact Patty Carroll by voicemail (xxx-xxx-xxxx) or by email at: patriciacarroll@ucla.edu and mention "CELDT study" in your message. You will receive a response with details about setting up an appointment for the interview.

APPENDIX C

INTERVIEW PROTOCOL

Interviewer will begin with the following:

- Thank you for volunteering to participate in this research study.
- The purpose of this study is to understand the role that CELDT test data plays in the classification, reclassification and educational programming decisions for language minority students at grades K-12 in California.
- The purpose of these interviews is to help the researcher better understand normal practices as well as the exceptions that can arise in the real world of district data use.
- This interview will be recorded, with your permission, for transcription purposes only.
- This interview is strictly confidential: you don't have to answer any question(s) you don't want to and the interview can stop at any time you wish. Your responses will be made anonymous.
- In order to maintain confidentiality of students, please do your best to avoid using student names during this interview. Any accidental use of student names will be made anonymous in the transcript.
- This interview is simply for the purposes of gathering procedural information and your remarks will not be judged. Feel free to share your opinions and understandings from your own experience.
- Do you have any questions before we start?

Questions:

1) Please describe your understanding of how the CELDT is used in classification decisions K-12 in this district. What are the normal practices? Are there any exceptions?

2) Please describe your understanding of how the CELDT is used in reclassification decisions K-12 in this district. What are the normal practices? Are there any exceptions?

3) Please describe your understanding of how the CELDT is used for educational programming decisions K-12, for example for placement in mainstream classes, in special education classes, in remedial reading programs, and in gifted and talented programs. What are the normal practices? Are there any exceptions?

4) Is there anything else you'd like me to know in relation to anything we've discussed?

5) Would you be willing to participate in a small focus-group discussion at the end of this study? (I will be presenting my findings and asking for feedback to ensure accuracy.) If so, what is the best way to stay in touch with you?

APPENDIX D

UNIVERSITY OF CALIFORNIA LOS ANGELES
STUDY INFORMATION SHEET

THE ROLE OF THE CELDT IN CLASSIFICATION, RECLASSIFICATION AND EDUCATIONAL PROGRAMMING DECISIONS FOR STUDENTS K-12 IN ONE CALIFORNIA SCHOOL DISTRICT

Patty Carroll, MA-TESOL, MA-Ed., and faculty sponsor Alison Bailey, Ed.D, Ed.M, from the Graduate School of Education and Information Studies at the University of California, Los Angeles (UCLA) are conducting a research study.

You were selected as a possible participant in this study because of your knowledge of "DISTRICT" practices in using the CELDT to classify, reclassify or determine educational programming for students in "DISTRICT." Your participation in this research study is voluntary.

Why is this study being done?
The purpose of this study is to learn more about how data from the California English Language Development Test (CELDT) is used to make decisions about EL students regarding their classification, reclassification and educational programming. The main research questions are:

1) *What are the procedures for using CELDT evidence in classification, reclassification and educational programming decisions for language minority students in a K-12 school district?*
2) *In instances where CELDT evidence prevents reclassification K-12, what (if any) is the incidence of measurement error?*

To answer these questions, the researcher will conduct interviews and analyze technical manuals and CELDT data. Together, these data will "tell the story" of how the CELDT may affect students in K-12. This "story of the CELDT" stands to inform future test development, assessment system validation, and English learner policy.

What will happen if I take part in this research study?
If you volunteer to participate in this study, the researcher will ask you to do the following:

- Participate in a 30-minute one-on-one interview where you will be asked questions about the procedures for CELDT use in "DISTRICT", such as:
    - How is the CELDT used for classification, K-12?
    - How is the CELDT used for reclassification, K-12?
    - How is the CELDT used for educational programming, K-12, for all types of English Learner students – including gifted, special education, and students in remedial reading programs?
- Location for the interviews will be the conference room in the "DISTRICT" business office unless otherwise arranged.

How long will I be in the research study?
Participation will take a total of about 30 minutes for the one-on-one interview. These interviews will be held at a time convenient for the participants.

Are there any potential risks or discomforts that I can expect from this study?
There are no anticipated risks or discomforts.

Are there any potential benefits if I participate?

- You may benefit from the study as you will have a chance to help researchers better understand the district perspective on the CELDT.
- The results of the research may lead to more effective test development for and state policy regarding the education of language minority students.

Will I be paid for participating?
You will receive the opportunity to enter your name in a drawing for a chance to win a thank you prize: a $25 Amazon gift certificate. Up to 8 participants will win. Your entry ticket into the drawing will be an envelope which, if you are one of the winners, will be returned to you with the gift certificate inside.

Will information about me and my participation be kept confidential?

Any information that is obtained in connection with this study and that can identify you will remain confidential. It will be disclosed only with your permission or as required by law. Confidentiality will be maintained by means of encrypting any files containing personal data, keeping all data password protected on a secure server at UCLA. Only Patty Carroll and Alison Bailey will have access to the data. All participants who participate in the focus group discussion at the end of the study will be asked to keep what is said during the focus group discussion between the participants only. However, complete confidentiality cannot be guaranteed.

What are my rights if I take part in this study?

- You can choose whether or not you want to be in this study, and you may withdraw your consent and discontinue participation at any time.

- Whatever decision you make, there will be no penalty to you, and no loss of benefits to which you were otherwise entitled.

- You may refuse to answer any questions that you do not want to answer and still remain in the study.

Who can I contact if I have questions about this study?

- **The research team:**

  If you have any questions, comments or concerns about the research, you can talk to the one of the researchers. Please contact:

  Patty Carroll phone: xxx-xxx-xxxx email: patriciacarroll@ucla.edu
  Alison Bailey phone: xxx-xxx-xxxx email: abailey@gseis.ucla.edu

- **UCLA Office of the Human Research Protection Program (OHRPP):**

  If you have questions about your rights while taking part in this study, or you have concerns or suggestions and you want to talk to someone other than the researchers about the study, please call the OHRPP at (xxx) xxx-xxxx or write to:

  UCLA Office of the Human Research Protection Program
  11000 Kinross Avenue, Suite 211, Box 951694
  Los Angeles, CA 90095-1694

APPENDIX E

Students Meeting CELDT Criteria in Study District, 2012-2013

Table E1

INITIAL TESTING: Students Meeting CELDT Criterion, 2012-13

	K	1	2	3	4	5	6	7	8	9	10	11	12	All
% Students	4	17	18	22	11	31	33	89	50	22	30	***	***	14
# Students	6	4	3	5	1	4	3	8	2	2	3	***	***	42
# Tested	164	23	17	23	9	13	9	9	4	9	10	3	1	294

***total number tested ≤ 3

Table E2

ANNUAL TESTING: Students Meeting CELDT Criterion, 2012-13

	K	1	2	3	4	5	6	7	8	9	10	11	12	All
% Students	0	36	16	36	42	47	37	65	49	69	39	81	48	41
# Students	0	44	21	43	43	44	22	36	19	33	14	21	13	353
# Tested	6	121	133	121	103	93	59	55	39	48	36	26	27	867

Table E3

ANNUAL TESTING: Students with Disabilities Meeting CELDT Criterion, 2012-13

	K	1	2	3	4	5	6	7	8	9	10	11	12	All
% Students	0	8	0	8	14	10	21	38	25	33	8	57	14	19
# Students	0	1	0	1	2	2	4	8	4	5	1	4	1	33
# Tested	0	12	15	12	14	21	19	21	16	15	13	7	7	172

APPENDIX F

Table E4

CELDT Raw and Scale Scores for Level Cutoffs for 2012-13, by Grade

	Listening						Speaking						Reading						Writing					
	L[a]	L2	L3	L4	L5	H[b]	L	L2	L3	L4	L5	H	L	L2	L3	L4	L5	H	L	L2	L3	L4	L5	H
KR[c]	0	5	8	12	16	20	0	5	11	20	26	29	0	6	12	19	23	24	0	8	14	19	23	28
KS[d]	220	364	411	460	513	570	140	362	409	463	523	630	220	287	322	385	481	570	220	343	373	399	427	600
1R	0	5	8	12	16	20	0	5	11	20	26	29	0	17	20	23	24	24	0	18	21	24	27	28
1S	220	364	411	460	513	570	140	362	409	463	523	630	220	362	399	481	570	570	220	393	411	438	507	600
2R	0	6	10	14	17	20	0	6	13	21	26	29	0	12	19	28	32	35	0	11	20	27	30	35
2S	220	382	437	485	531	570	140	371	421	470	523	630	280	428	473	527	564	650	220	425	472	523	581	690
3R	0	9	11	14	16	20	0	5	11	19	25	29	0	11	15	24	29	35	0	12	18	26	29	35
3S	220	405	449	513	559	640	200	396	438	484	540	720	280	454	488	546	585	700	220	441	480	542	575	740
4R	0	9	12	15	17	20	0	6	14	21	26	29	0	14	16	26	31	35	0	14	20	27	30	35
4S	220	405	470	535	585	640	200	405	455	498	557	720	280	481	495	560	608	700	220	455	494	552	589	740
5R	0	10	13	16	18	20	0	7	15	23	26	29	0	14	18	27	31	35	0	14	21	27	30	35
5S	220	427	492	559	619	640	200	413	460	515	557	720	280	481	508	568	608	700	220	455	501	552	589	740
6R	0	8	11	15	17	20	0	7	13	20	24	29	0	11	14	20	26	35	0	13	20	26	29	35
6S	230	419	493	591	656	715	225	420	468	523	568	720	320	489	522	570	614	750	220	460	508	560	595	780
7R	0	8	12	15	17	20	0	8	15	21	25	29	0	11	15	21	26	35	0	14	20	26	30	35
7S	230	419	517	591	656	715	225	429	483	532	584	720	320	489	531	577	614	750	220	466	508	560	609	780
8R	0	9	12	16	18	20	0	8	15	22	26	29	0	12	17	23	28	35	0	14	21	26	30	35
8S	230	445	517	621	704	715	225	429	483	543	604	720	320	502	548	592	631	750	220	466	515	560	609	780
9R	0	9	12	15	17	20	0	5	11	19	25	29	0	12	17	23	28	35	0	15	20	25	29	35
9S	230	447	537	641	724	725	235	433	487	551	626	740	320	509	563	612	656	770	220	472	514	561	612	810
10R	0	9	12	15	17	20	0	5	12	20	25	29	0	13	18	25	29	35	0	15	21	26	29	35
10S	230	447	537	641	724	725	235	433	495	560	626	740	320	521	571	628	666	770	220	472	523	573	612	810
11R	0	9	12	15	17	20	0	5	12	20	25	29	0	13	18	25	29	35	0	15	21	26	29	35
11S	230	447	537	641	724	725	235	433	495	560	626	740	320	521	571	628	666	770	220	472	523	573	612	810
12R	0	9	12	15	17	20	0	5	12	20	25	29	0	13	18	25	29	35	0	15	21	26	29	35
12S	230	447	537	641	724	725	235	433	495	560	626	740	320	521	571	628	666	770	220	472	523	573	612	810

Note: [a]Lowest Obtainable Scale Score, [b]Highest Obtainable Scale Score, [c]Raw Score, [d]Scale Score; from CELDT Guide 2012-13, California Department of Education

APPENDIX G

McNemar's Test Results

Table G1

McNemar's Test: Composites Compared to State (Mix-1), by Group

Group		Comp	Conj-1	Conj-2	Mix-2	Mix-3
EL	N	474	474	474	474	474
	Chi-Square[c]		130.008	74.013	130.008	62.132
	Asymp. Sig.		.000	.000	.000	.000
	Exact Sig. (2-tailed)	.031[b]	.000	.000	.000	.000
	Exact Sig. (1-tailed)	.016	.000	.000	.000	.000
EL/SWD	N	78	78	78	78	78
	Exact Sig. (2-tailed)	.250[b]	.000[b]	.000[b]	.000[b]	.500[b]
	Exact Sig. (1-tailed)	.125	.000	.000	.000	.250
EL/HP	N	107	107	107	107	107
	Chi-Square[c]		40.024		40.024	
	Asymp. Sig.		.000		.000	
	Exact Sig. (2-tailed)	.125[b]	.000	.000[b]	.000	.000[b]
	Exact Sig. (1-tailed)	.063	.000	.000	.000	.000
EL/SWD/HP	N	8	8	8	8	8
	Exact Sig. (2-tailed)	1.000[b]	.063[b]	1.000[b]	.063[b]	.250[b]
	Exact Sig. (1-tailed)	.500	.031	.500	.031	.125

Note: McNemar Test; b. Binomial distribution used; c. Continuity Corrected.

Table G2

McNemar's Test: Composites Compared to State (Mix-1), by Gender

Gender		Comp	Conj-1	Conj-2	Mix-2	Mix-3
Female	N	335	335	335	335	335
	Chi-Square[c]		104.009	48.020	104.009	43.184
	Asymp. Sig.		.000	.000	.000	.000
	Exact Sig. (2-tailed)	.004[b]	.000	.000	.000	.000
	Exact Sig. (1-tailed)	.002	.000	.000	.000	.000
Male	N	332	332	332	332	332
	Chi-Square[c]		84.012	54.018	84.012	36.026
	Asymp. Sig.		.000	.000	.000	.000
	Exact Sig. (2-tailed)	.063[b]	.000	.000	.000	.000
	Exact Sig. (1-tailed)	.031	.000	.000	.000	.000

Note: McNemar Test; b. Binomial distribution used; c. Continuity Corrected.

Table G3

McNemar's Test: Aggregates Compared to State (Mix-1 & CST-ELA ≥300), by Group

Group		Comp	Conj-1	Conj-2	Mix-2	Mix-3
EL	N	474	474	474	474	474
	Chi-Square[c]		91.011	43.022	91.011	43.022
	Asymp. Sig.		.000	.000	.000	.000
	Exact Sig. (2-tailed)	.031[b]	.000	.000	.000	.000
	Exact Sig. (1-tailed)	.016	.000	.000	.000	.000
EL/SWD	N	78	78	78	78	78
	Exact Sig. (2-tailed)	1.000[b]	.063[b]	.008[b]	.063[b]	1.000[b]
	Exact Sig. (1-tailed)	.500	.031	.004	.031	.500
EL/HP	N	107	107	107	107	107
	Chi-Square[c]		40.024		40.024	
	Asymp. Sig.		.000		.000	
	Exact Sig. (2-tailed)	.125[b]	.000	.000[b]	.000	.000[b]
	Exact Sig. (1-tailed)	.063	.000	.000	.000	.000
EL/SWD/HP	N	8	8	8	8	8
	Exact Sig. (2-tailed)	1.000[b]	.063[b]	1.000[b]	.063[b]	.250[b]
	Exact Sig. (1-tailed)	.500	.031	.500	.031	.125

Note: McNemar Test; b. Binomial distribution used; c. Continuity Corrected.

Table G4

McNemar's Test: Aggregates Compared to District (Mix-1 & CST-ELA≥325), by Group

Group		Conj-1	Comp-1	Comp-2	Mix-2	Mix-3
EL	N	474	474	474	474	474
	Chi-Square[c]		60.016	24.038	60.016	27.034
	Asymp. Sig.		.000	.000	.000	.000
	Exact Sig. (2-tailed)	.125[b]	.000	.000	.000	.000
	Exact Sig. (1-tailed)	.063	.000	.000	.000	.000
EL/SWD	N	78	78	78	78	78
	Exact Sig. (2-tailed)	1.000[b]	.250[b]	.125[b]	.250[b]	1.000[b]
	Exact Sig. (1-tailed)	.500	.125	.063	.125	1.000
EL/HP	N	107	107	107	107	107
	Chi-Square[c]		40.024		40.024	
	Asymp. Sig.		.000		.000	
	Exact Sig. (2-tailed)	.125[b]	.000	.000[b]	.000	.000[b]
	Exact Sig. (1-tailed)	.063	.000	.000	.000	.000
EL/SWD/HP	N	8	8	8	8	8
	Exact Sig. (2-tailed)	1.000[b]	.063[b]	1.000[b]	.063[b]	.250[b]
	Exact Sig. (1-tailed)	.500	.031	.500	.031	.125

Note: McNemar Test; b. Binomial distribution used; c. Continuity Corrected.

APPENDIX H

CODEBOOK

PRIMARY CODES	DEFINITION
ALIT	Assessment Literacy – comments about personal knowledge of test/tool features
INITIAL	Initial Classification
LTEL	Long Term EL – comments about LTELs or students "Stuck at 3"
OPT	Opting in or out of EL programs
EDPROG	Educational Programming
QUOTE	Quote – direct quote that could be used as illustrative of a point
RFEP	Redesignation/Reclassification
SPED	Special Education
TOOL	Tools – comments about the CELDT itself, knowledge/experience with it
TREND	Trends – comments about trends in data, data use or student test performances
VIG	Vignette – story told that is illustrative of a point
WHATIF	What if English Only students took this test?-type of comments

SECONDARY CODES	DEFINITION
CHANGE	Changes – referring to changes in personnel, funding, rules, criteria, avail of CST, etc.
CONTAIN	Metaphor of Containment – referring to EL system/classes in containment terms
MOTIV	Motivation – to what extent student motivation is a factor for EL/RFEP/LTEL…
PHIL_EL	Philosophy of English learners – who they are, what they should be
PHIL_ELD	Philosophy of English Language Development classes
PHIL_INTV	Philosophy of Intervention – reading intervention, math intervention
PHIL_LTEL	Philosophy of Long Term English Learners – who they are, what they should be
PHIL_RFEP	Philosophy of Reclassification – when it should happen, when/how it is beneficial
PHIL_SLA	Philosophy of Second Language Acquisition – how it happens
PHIL_SPED	Philosophy of Special Education
PHIL_TOOL	Philosophy of Tool – what tests (CELDT/CST) are, what they should do
RESOURCE	Resources – referring to assets available for the EL system, typically aides and books
TERMS	Terms used for ELs – e.g., 1s, 2s, a stuck, a low, …
TROUBLE	Getting in trouble – referring to laws, regulations, power relationship with feds

REFERENCES

Abedi, J. (2004). The No Child Left Behind Act and English language learners: Assessment and accountability issues. *Educational Researcher, 33*(1), 4-14.

Abedi, J. (Ed.) (2007). *English language proficiency assessment in the nation: Current status and future practice.* Davis, CA: UC Davis School of Education.

Abedi, J. (2008a). Classification system for English language learners: Issues and recommendations. *Educational Measurement: Issues and Practice, 27*(3), 17-31.

Abedi, J. (2008b). Measuring students' level of English Proficiency: Educational significance and assessment requirements. *Educational Assessment, 13*(2/3), 193-214.

Abedi, J., & Gándara, P. (2006). Performance of English language learners as a subgroup in large-scale assessment: Interaction of research and policy. *Educational Measurement: Issues and Practice, 25*(4), 36-46.

Allen, M.J., & Yen, W.M. (1992). *Introduction to measurement theory*. Monterey, CA: Brooks/Cole.

American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME] (2014). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

Bachman, L.F. & Palmer, A. (2010). *Language assessment in practice.* New York: Oxford University Press.

Bailey, A.L. (2007). *The language demands of school: Putting academic English to the test.* New Haven, CT: Yale University Press.

Bailey, A.L., & Carroll, P.E. (2015). The assessment of English language learners in the era of new academic content standards. *Review of Research in Education, 39,* 253-297.DOI: 10.3102/0091732X14556074

Bailey, A.L., & Heritage, M. (2010). *English language proficiency assessment foundations: External judgments of adequacy.* Enhanced Assessment Grant US DOE *Evaluating the Validity of English Assessments (EVEA)* Project deliverable. Retrieved from http://www.eveaproject.com

Bailey, A.L., & Kelly, K.R. (2012). Home language survey practices in the initial identification of English learners in the United States. *Educational Policy*, 1-35. Published online before print May 18, 2012, DOI: 10.1177/0895904811432137.

Baker, E., & Linn, R. L. (2004). Validity issues for accountability systems. In S. H. Fuhrman & R. F. Elmore (Eds.), *Redesigning accountability systems for education* (pp. 47–72). New York: Teachers College Press.

Brantlinger, E., & Jimenez, R., Klingner, J., Pugach, M., & Richardson, V. (2005). Qualitative studies in special education. *Exceptional Children, 71*(2), 195-207.

Burke, A.M. (2014). *Making the cut: Indiana school accountability and English learner test performance.* (Doctoral dissertation). Available from Purdue University.

California Department of Education, 2012. "California English Language Development Test (CELDT), 2012-2013," CELDT Information Guide. Available at www.cde.ca.gov/ta/tg/el/documents/celdtinfoguide1213.pdf/.

California Department of Education. (2011). *A comparison study of kindergarten and grade 1 English-fluent students and English learners on the 2010-11 edition of the CELDT* (October 25, 2011).

California Department of Education. (2012). *California Standards Tests technical report.* Sacramento, CA: Author. Retrieved on 5/1/13 from http://www.cde.ca.gov

California Department of Education. (2013) *Dataquest.* Retrieved on 5/1/13 from http://data1.cde.ca.gov/dataquest/

California Education Code Section 313. Available at www.leginfo.ca.gov/cgi-bin/displaycode?section=edc&group=00001-01000&file=313-313.5.

Callahan, R., Wilkinson, L., & Muller, C. (2010). Academic achievement and course taking among language minority youth in U.S. schools: Effects of ESL placement. *Educational evaluation and policy analysis, 32*(1), 84-117.

Carroll, P.E. (2012). *Examining the validity of classifications from an English language proficiency assessment for English language learners and native English speakers in fifth grade* (Master's thesis). Available from ProQuest Dissertations and Theses database. (UMI No. 1516569)

Carroll, P.E., & Bailey, A.L. (2012, November). *Examining conjunctive and compensatory classification models for an English language proficiency assessment: A descriptive study of English learners and native English speakers.* Presented at the conference to honor Ronald Hambleton, University of Massachusetts: Amherst, MA.

Carroll, P.E., & Bailey, A.L. (2013a, April). *Combining multiple indicators in classifications of English language proficiency: A descriptive study.* Presented at the National Council on Measurement in Education Annual Meeting, San Francisco, CA.

Carroll, P.E., & Bailey, A.L. (2013b, April). *Language-as-Resource in English learner assessment systems: Evaluating the fit of classification models.* Presented at the American Educational Research Association Annual Meeting, San Francisco, CA.

Carroll, P.E., & Bailey, A.L. (2014, April). *Classification models and English learner redesignation: High-performing students left behind?* Paper presented at the National Council on Measurement in Education Annual Meeting, Philadelphia, PA.

Carroll, P.E., & Bailey, A.L. (2015). Do decision rules matter? A descriptive study of English language proficiency assessment classifications for English-language learners and native English speakers in fifth grade. *Language Testing*. DOI: 10.1177/0265532215576380

Chester, M.D. (2003). Multiple measures and high-stakes decisions: A framework for combining measures. *Educational Measurement: Issues in Practice, 22*(2), 32-41.

Clauser, B.E., Clyman, S.G., Margolis, M.J. & Ross, L.P. (1996). Are fully compensatory models appropriate for setting standards on performance assessments of clinical skills? *Academic Medicine, 71*(1:supplement), S90-S92.

Coburn, C.E., & Talbert, J.E. (2006). Conceptions of evidence use in school districts: Mapping the terrain. *American Journal of Education, 112*(4), 469-495.

Coburn, C.E., & Turner, E.O. (2011). Research on data use: A framework and analysis. *Measurement: Interdisciplinary Research and Perspectives, 9*(4), 173-206.

Coburn, C.E., & Turner, E.O. (2012). The practice of data use: An introduction. *American Journal of Education, 111*(2), 99-111.

Cook, H.G., Boals, T., Wilmes, C., & Santos, M. (2008). *Issues in the development of annual measurable achievement objectives for WIDA consortium states* (WCER Working Paper No. 2008-2). Madison: University of Wisconsin-Madison, Wisconsin Center for Education Research. Retrieved from http://www.wcer.wisc.edu/publications.

Corbin, J., & Strauss, A. (2007). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (3rd ed.). Thousand Oaks, CA: SAGE.

Crane, E.W., Barrat, V.X., and Huang, M. (2011). *The relationship between English proficiency and content knowledge for English language learner students in grades 10 and 11 in Utah.* (Issues & Answers Report, REL 2011 – No. 110). Washington, DC: U.S. Department of Education, Institute of Educational Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory West. Retrieved from http://ies.ed.gov/ncee/edlabs.

Cook, H.G. (2009). *Annual Measurable Achievement Objective #2 (AMAO 2): A Report on a Meeting Comparing the Kentucky Core Content Test (KCCT) and ACCESS for ELLs™ to Establish an AMAO 2 Exit Criterion.* Kentucky Department of Education.

Cook, H.G., Linquanti, R., Chinen, M., Jung, H. (2012). *National Evaluation of Title III Implementation Supplemental Report – Exploring approaches to setting English language proficiency performance criteria and monitoring English learner progress*. Washington, DC: U.S. Department of Education, Office of Planning, Evaluation, and Policy Development, Policy and Program Studies Service. Retrieved from http://www.ed.gov.

Creswell, J.W. (2008). *Research design: Qualitative, quantitative and mixed methods approaches.* Thousand Oaks, CA: SAGE.

Creswell, J.W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research* (2nd ed.). Thousand Oaks, CA: SAGE.

Davidson, F., Kim, J.T., Lee, H., Li, J., & López, A.A. (2007). English language testing: Evidence from the evolution of test specifications. In A.L. Bailey (Ed.), *The language demands of school: Putting academic English to the test* (pp. 157-170). New Haven, CT: Yale University Press.

Del Rio, V. (2010). *English-learner representation in special education: Impact of prereferral interventions and assessment practices* (Doctoral dissertation, University of Southern California). Available from ProQuest Dissertations/Theses database. (UMI No. 3261849).

Douglas, K.M., & Mislevy, R.L. (2010). Estimating classification accuracy for complex decision rules based on multiple scores. *Journal of Education and Behavioral Statistics, 35*(3), 280-306.

Elementary and Secondary Education Act of 1965, Pub. L. No. 89-10, 79 Stat.27 (1965).

Estrada, P. (2013). *English learner curricular streams in four middle schools: Triage in the trenches.* Paper presented at the American Educational Research Association, Vancouver, Canada, April 2012. (Revised).

Estrada, P. (2014). English learner *Curricular Streams* in four middle schools: Triage in the trenches. *Urban Review, 46, 535-573.*

Federal Register, 73 (202), 61828-61844. October 17, 2008.

Fletcher, J. M., Lyon, G. R., Fuchs, L. S., & Barnes, M. A. (2007). *Learning disabilities: From identification to intervention*. Guilford Publications.

Florez, I. R. (2012). Examining the validity of the Arizona English Language Learners Assessment cut scores. *Language Policy, 11(1),* 33-45.

Forte, E., & Willhoft, J. (June, 2010). Evaluating the validity of English language proficiency assessments: A consortium approach. Presented at the CCSSO National Conference on Student Assessment, Detroit, MI. Retrieved from: http://www.eveaproject.com.

Gándara, P., Rumberger, R., Maxwell-Jolly, J. & Callahan, R. (2003, October 7). English Learners in California Schools: Unequal resources, unequal outcomes. *Education Policy Analysis Archives, 11*(36). Retrieved on October 23, 2011 from http://epaa.asu.edu/epaa.

García Bedolla, L., & Rodriguez, R. (2011). Classifying California's English learners: Is the CELDT too blunt an instrument? *Policy Reports and Research Briefs, Center for Latino Policy Research, Institute for the Study of Societal Issues,* UC Berkeley.

Garcia, E. E., Lawton, K. & Diniz de Figueiredo, E. H. (2010). *Assessment of young English language learners in Arizona: Questioning the validity of the state measure of English proficiency*. Los Angeles: Civil Rights Project, University of California. Retrieved from: http://civilrightsproject.ucla.edu/research.

Grissom, J.B. (2004, July 30). Reclassification of English Learners. *Education Policy Analysis Archives, 12*(36). Retrieved on 2/22/13 from http://epaa.asu.edu/epaa/v12n36/.

Haager, D. (2007). Promises and cautions regarding using response to intervention with English language learners. *Learning Disability Quarterly, 30,* 213-218.

Hakuta, K. (2012). Understanding Language Conference: Stanford University. Available online at: ell.stanford.edu/papers

Hambleton, R.K., & Pitoniak, M.J. (2006). Setting performance standards. In Brennan, R. (Ed.) *Educational Measurement, Fourth Edition* (pp. 433-470). Portsmouth, NH: Praeger.

Hambleton, R.K., Jaeger, R.M., Plake, B.S., & Mills, C. (2000). Setting performance standards on complex educational assessments. *Applied Psychological Measurement, 24*(4), 355-366.

Hill, L.E., Betts, J.R., Chavez, B., Zau, a.C., and Bachofer, K.V. (2014 May). *Pathways to Fluency: Examining the link between language reclassification policies and student success.* San Francisco: Public Policy Institute of California.

Hill, L.E., Weston, M., and Hayes, J. (2014). *Reclassification of English Learner Students in California.* San Francisco: Public Policy Institute of California.

Honig, M.I. & Coburn, C.E. (2008). Evidence-based decision making in school district central offices: Toward a policy and research agenda. *Educational Policy, 22*(4), 578-608.

Honig, M.I., & Venkateswaran, N. (2012). School-central office relationships in evidence use: Understanding evidence as a systems problem. *American Journal of Education, 118*(2), 199-222.

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

Kane, M. (2010). Validity and fairness. *Language Testing*, *27*(2), 177-182.

Kane, M. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement, 50*(1), 1-73.

Kim, J. & Herman, J. L. (2010). *When to exit ELL students: monitoring success and failure in mainstream classrooms after ELLs' reclassification*. (CRESST Report 779). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Kim, J. (2011). *Relationships among and between ELL status, demographic characteristics, enrollment history, and school persistence* (CRESST Report 810). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Kim, J., & Herman, J.L. (2012). Understanding patterns and precursors of ELL success subsequent to reclassification. (CRESST Report 818). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Klingner, J.K., & Harry, B. (2006). The special education referral and decision-making process for English Language Learners: Child study team meetings and staffing. *Teachers College Record, 108*, 2247-2281.

Kranzler, J.H., Flores, C. G., & Coady, M. (2010). Examination of the cross-battery approach for the cognitive assessment of children and youth from diverse linguistic and cultural backgrounds. *School Psychology Review, 39*, 431-446.

Kritikos, E.P. (2003). Speech-language pathologists' beliefs about language assessment of bilingual/bicultural individuals. *American Journal of Speech-Language Pathology, 12*, 73-91.

Linquanti, R. (2001). *The reclassification dilemma: Challenges and choices in fostering meaningful accountability for English learners* (UC Language Minority Research Institute Policy Report 2001-01). Davis, CA: University of California, Davis.

Linquanti, R. & Cook, H.G. (2013). *Toward a "common definition of English learner": Guidance for states and state assessment consortia in defining and addressing policy and technical issues and options.* Washington, DC: Council of Chief State School Officers.

Linquanti, R., & George, C. (2007). Establishing and Utilizing an NCLB Title III Accountability System: California's Approach and Findings to Date. In J. Abedi (Ed.) *English language proficiency assessment in the nation: Current status and future practice* (pp. 105-118). Davis, CA: UC Davis School of Education.

Linquanti, R., & Hakuta, K. (2012). *How next-generation standards and assessments can foster success for California's English learners.* PACE Policy Brief 12-1, School of Education, Stanford University. Retrieved on 3/13/13 from http://www.edpolicyinca.org.

Little, J.W. (2012). Understanding data use practice among teachers: The contribution of micro-process studies. *American Journal of Education, 118*(2), 143-166.

Livingston, S.A. & Lewis, C. (1995). Estimating the consistency and accuracy of classifications based on test scores. *Journal of Educational Measurement*, 32(2), 179-197.

Mahoney, K. S., & MacSwan, J. (2005). Reexamining identification and reclassification of English language learners: A critical discussion of select state practices. *Bilingual Research Journal*, 29 (1), 31-42.

Mahoney, K.S., Haladyna, T., & MacSwan, J. (2009). *The need for multiple measures in reclassification decisions: A validity study of the Stanford English Language Proficiency Test*. In T.G. Wiley, J.S. Lee, & R.W. Rumberger (Eds.) *The education of language minority immigrants in the United States* (pp. 240-262). Tonawanda, NY: Multilingual Matters.

McBee, M.T., Peters, S.J., & Waterman, C. (2014). Combining scores in multiple-criteria assessment systems: The impact of combination rule. *Gifted Child Quarterly,58*(1),69-89.

Menken, K. (2008). English learners left behind: Standardized testing as language policy (Vol. 65). Cleveden, UK: Multilingual Matters.

Merriam, S.B. (2002). *Qualitative research in practice: Examples for discussion and analysis.* San Francisco: Jossey-Bass Publishers.

Mosier, C.I. (1943). On the reliability of a weighted composite. *Psychometrika, 8*(3), 161-168.

Moss, P. (2007). *Evidence and decision making*. Malden, MA: Blackwell.

National Association of School Psychologists ([NASP] 2007). *Identification of students with specific learning disabilities* (Position Statement). Bethesda, MD: NASP.

National Research Council. (2011). *Allocating federal funds for state programs for English language learners.* Panel to Review Alternative Data Sources for the Limited-English Proficiency Allocation Formula under Title III, Part A, Elementary and Secondary Education Act. Committee on National Statistics and Board on Testing and Assessment. Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

No Child Left Behind Act of 2001 ([NCLB], 2001). Conference report to accompany H.R., 1, Rep. No. 107-334, House of Representatives, 107th Congress, 1st Session, December 13. Pub. L. No. 107-110, 115 Stat. 1425.

Piety, P. (2011). Educational data use: A sociotechnical process. *Measurement: Interdisciplinary Research & Perspective*, *9*(4), 217-221.

Rabinowitz, S., Sato, E. (2006). Technical adequacy of assessments from alternate student populations: Technical review of high-stakes assessment for English language learners. Presented at the annual meeting of the National Council on Measurement in Education, San Francisco, CA.

Ragan, A., & Lesaux, N. (2006). Federal, state, and district level English language learner program entry and exit requirements: Effects on the education of language minority learners. *Education Policy Analysis Archives*, *14*(20). Retrieved on October 6, 2011 from http://epaa.asu.edu/epaa/v14n20/.

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

Simon, H. (1996). *The science of the artificial.* Cambridge, MA: MIT Pres.

Sireci, S., Han, K. T., & Wells, C. S. (2008). Methods for evaluating the validity of test scores for English language learners. *Educational Assessment,* 13(3), 108-131.

Slama, R. B. (2014). Investigating Whether and When English Learners Are Reclassified Into Mainstream Classrooms in the United States A Discrete-Time Survival Analysis. *American Educational Research Journal*, *51*(2), 220-252.

Spillane, J.P. (2012). Data in practice: Conceptualizing the data-based decision-making phenomena. *American Journal of Education, 118*(2), 113-141.

Stephenson, A., Jiao, H., & Wall, N. (2004). *A performance comparison of native and non-native speakers of English on an English language proficiency test.* San Antonio, TX: Harcourt Assessment, Inc. Retrieved from http://www.pearsonassessments.com.

Stephenson, A., Johnson, D.F., Jorgensen, M.A. & Young, M.J. (2003). *Assessing English language proficiency: Using valid results to optimize instruction.* San Antonio, TX: Harcourt Assessment, Inc. Retrieved from http://www.pearsonassessments.com.

The Stanford English Language Proficiency Test ([SELP] 2003). San Antonio, TX: Harcourt Assessment, Inc.

U.S. Department of Education, National Center for Educational Statistics. (2012). *The Condition of Education 2011 (NCES 2011-045),* Indicator 6. Retrieved on 5/28/13 from: http://nces.ed.gov/fastfacts/display.asp?id=96

U.S. Department of Education, Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service (2012). *National evaluation of Title III implementation supplemental report – Exploring approaches to setting English language proficiency performance criteria and monitoring English learner programs*, Washington, DC, 2012

U.S. Department of Education, Office of Special Education and Rehabilitation Services. (2002). *A new era: Revitalizing special education for children and their families by the President's commission on Excellence in Special Education.* Washington, DC: ED Pubs. Available at http://www.ed.gov/publs/edpubs.html.

Walqui, A., Koelsch, N., Hamburger, L., et al. (2010). *What are we doing to middle school English Learners? Findings and recommendations for change from a study of California EL programs* (Narrative Summary). San Francisco: WestEd.

Wang, J., Niemi, D., & Wang, H. (2007). *Impact of different performance assessment cut scores on student promotion.* CSE (CRESST) Report 719, 2007.

Wang, M.W., & Stanley, J.C. (1970). Differential weighting: A review of methods and empirical studies. *Review of Educational Research*, 40, 663-705.

Weiss, J.A. (2012). Data for improvement, data for accountability. *Teachers College Record 114*(11), 1-7.

WIDA Consortium. (2009). *ACCESS for ELLs®.* Madison, WI: Board of Regents of the University of Wisconsin System.

Wolf, M. K., Farnsworth, T., & Herman, J. (2008). Validity issues in assessing English language learners' language proficiency. *Educational Assessment, 13*(3), 80-107.

Wolf, M. K., Herman, J. L., Dietel, R. (2010). *Improving the validity of English language learner assessment systems* (CRESST Policy Brief No. 10 – Full Report). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Wolf, M.K., Herman, J.L., Bachman, L.F., Bailey, A.L. & Griffin, N. (2008). *Recommendations for assessing English language learners: English language proficiency measures and*

accommodation uses, recommendation report (part 3 of 3). CRESST Report 737. Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Wolf, M.K., Kao, J., Herman, J. Bachman, L.F., Bailey, A., Bachman, P.L., Farnsworth, T., & Chang, S.M. (2008). *Issues in assessing English language learners: English language proficiency measures and accommodation uses – Literature Review (part 1 of 3).* CRESST Report 731. Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Xi, X. (2008). Methods of test validation. In *Encyclopedia of Language and Education* (pp. 2316-2335). Springer, US.

Xi, X. (2010). How do we go about investigating test fairness? *Language Testing, 27*(2), 147-170.

Yin, R.K. (2011). *Qualitative research from start to finish.* New York, NY: The Guilford Press.

Zhang, B. (2010). Assessing the accuracy and consistency of language proficiency classification under competing measurement models. *Language Testing, 27*(1), 119-140.